Austronesian Undressed

How and why languages become isolating

*edited by David Gil
and Antoinette Schapper*

John Benjamins Publishing Company

Austronesian Undressed

Typological Studies in Language (TSL)

A companion series to the journal Studies in Language. Volumes in this series are functionally and typologically oriented, covering specific topics in language by collecting together data from a wide variety of languages and language typologies.

For an overview of all books published in this series, please see
benjamins.com/catalog/tsl

Volume 129

Austronesian Undressed. How and why languages become isolating
Edited by David Gil and Antoinette Schapper

Austronesian Undressed

How and why languages become isolating

Edited by

David Gil

Max Planck Institute for the Science of Human History, Jena

Antoinette Schapper

Vrije Universiteit Amsterdam / Lacito-CNRS

John Benjamins Publishing Company

Amsterdam / Philadelphia

To the memory of Scott Paauw

Table of contents

Preface

This volume is a product, albeit a belated and rather indirect one, of a panel on "Isolating Austronesian Languages", organised by David Gil and John McWhorter, which took place on the 22nd of June 2009, at the Eleventh International Conference on Austronesian Linguistics, in Aussois in the French Alps.

The production of the volume took much longer than anticipated, due largely to the sloth and ineptitude of the first editor, David Gil. In the course of the years, we lost a few papers but also gained a few others, and several of the original papers presented at the 11ICAL panel evolved substantially as their authors were able to incorporate fresh insights and understandings with the passage of time.

The focus of the volume is largely inspired by John McWhorter, who, in his writings, brought the issue of isolating Austronesian languages to the attention of the general linguistic community, and introduced the metaphor of languages undressing. For several years, editorial work on the volume was assisted by Scott Paauw, whose untimely passing was a great shock to us all – this volume is dedicated to his memory.

Finally, we are grateful to Angela Terrill for her punctilious copy editing.

Introduction

David Gil and Antoinette Schapper

Many studies of Austronesian languages have focused on aspects of their morphological structure. One of the most commonly discussed grammatical features of Austronesian languages is their voice systems, often expressed by means of verbal affixation. Other morphological features typically associated with Austronesian languages include the use of a ligature to mark nominal attribution, and the expression of a wide range of grammatical functions by means of reduplication. In addition, Austronesian languages are host to a variety of cross-linguistically unusual or quirky morphological processes, ranging from case prefixes in Nias through plural infixes in Sundanese and complex ablaut paradigms in Kerinci to the expression of a variety of grammatical relations by means of metathesis in Leti. Studies of these and other aspects of Austronesian morphology may perhaps give rise to the impression that Austronesian languages are endowed with rich morphological systems.

Contrary, however, to such an impression, many Austronesian languages actually have relatively little word-internal structure, thus meriting the characterization as isolating. Examination of Dryer's (2005) map on prefixing vs. suffixing inflectional morphology in the *World Atlas of Language Structures* shows that whereas worldwide, the proportion of languages with "little or no affixation" is 14% (122 out of 894 languages in his sample), within Austronesian the proportion of such languages rises to 47% (45 out of 95). Although not all of Dryer's languages with "little or no affixation" are appropriately considered to be isolating, the figures still bear witness to a propensity, within the Austronesian language family, for simpler-than-average morphological structures, which, in many cases, do result in languages that may warrant the appellation of isolating.

Indeed, the geographical distribution of isolating Austronesian languages points towards the existence of an Isolating Crescent, stretching, with inevitable bumps, gaps and wiggles, from Hainan and mainland Southeast Asia down the Malay peninsula into Sumatra, Java and much of Borneo, and then along the lesser Sunda islands of Nusa Tenggara and up into the Bird's Head of New Guinea. At the northwest tip of the Isolating Crescent are the Chamic languages of Hainan and Indochina, relatively new arrivals to their region (Thurgood 1999). Much of

the Isolating Crescent is dominated by numerous varieties of colloquial Malay and Indonesian exhibiting a range of sociolinguistic types, among which are transplanted dialects such as Nonthaburi Malay (Tadmor 1995), heartland varieties such as Jambi Malay (Yanti 2010), urban koinés such as Jakarta Indonesian (Sneddon 2006), and eastern varieties such as Papuan Malay (Kluge 2014). Alongside these is a wide variety of other Austronesian languages sharing the isolating profile to different degrees. These include, but are not limited to, central Sumatran languages such as Minangkabau (Crouch 2009), regional varieties of Javanese (Conners 2008), languages of Flores such as Keo (Baird 2002) and Rongga (Arka, Kosmas and Suparsa 2007), languages of Timor such as Tetun Dili (Williams-Van Klinken, Hajek and Nordlinger 2002), as well as some of the lesser-known Austronesian languages of the Bird's Head region such as Sekar. This Isolating Crescent lies at the heart of the proposed Mekong-Mamberamo linguistic area (Gil 2015).

This volume is the first to bring together studies focusing on the isolating languages of the Austronesian world. We hope that it will be of interest to a wide range of scholars, from specialists in Austronesian languages and linguistics to generalists in related fields such as historical linguistics, language contact, creolistics, morphology, and linguistic typology.

The main goal of this volume is to address the question posed in the title, namely, how and why languages become isolating. While some scholars, such as Givón (1971), view morphological loss as a natural process that can happen to languages "on their own", others lean towards the view that the development of isolating word structure is typically driven by language contact. However, in the case of Austronesian languages, there are varying proposals regarding the mechanisms underlying such language contact. In one scenario, the development of isolating word structure is an outcome of general processes of simplification associated with imperfect transmission and second-language acquisition. In an alternative scenario, Austronesian languages may have acquired their isolating word structure through assimilation to pre-Austronesian languages that they encountered during their dispersal into the region. Another variable pertains to the time frame of the contact, and whether it is relatively recent, or alternatively, reconstructable to the distant past and the original intrusion of Austronesian languages into the region. Some of the concepts that are invoked in the course of the discussion include creolisation, metatypy, *Sprachbund* effects, and Non-Hybrid Conventionalised Second Languages. Nevertheless, while recognising the significance of language contact, several chapters suggest that it is not, and cannot be, the whole story behind the rise of isolating word structure in Austronesian languages; other important factors include phonological reduction, drift and typological stability. As argued in the chapter by Donohue and Denham (this volume), there is no single story shared by all of the isolating Austronesian languages; instead, different diachronic scenarios

may have played out in different places and at different times, an hypothesis borne out by the various in-depth case-studies presented in this volume.

The first chapter of this volume addresses the fundamental conceptual question of what it means to be an isolating language. Subsequent chapters address the question of how isolating word structure developed, with reference to a range of largely new and unfamiliar data from a wide variety of Austronesian languages exhibiting differing degrees of isolation, covering most of the of the Isolating Crescent, from Mainland Southeast Asia through western Indonesia and into Wallacea: Cham, colloquial varieties of Malay/Indonesian, Minangkabau, peripheral dialects of Javanese, languages of the Lamaholot cluster, Tetun Dili, and the languages of Timor. Chapter 10 presents a broader perspective on the spread of Austronesian languages into the region and how they acquired their isolating profile.

Chapter 1, "What does it mean to be an isolating language? The case of Riau Indonesian", by David Gil, sets the stage for the remainder of the volume by proposing a definition of an isolating language. The common understanding of an isolating language, as one in which words are of minimal morphological complexity, relies crucially on a prior understanding of wordhood; however, Martin Haspelmath and others have pointed to the practical difficulties in distinguishing words from smaller units such as affixes, and have accordingly called into question the universal validity of the notion of word. The first part of this chapter proposes a definition of word as a cut-off point between two distinct levels of structure, morphology and syntax, and follows with a definition of isolating language as one lacking a robust structural unit of word. The second part of this chapter presents a detailed exploration of wordhood in Riau Indonesian. While some evidence is provided for two distinct kinds of words, the evidence in question is relatively sparse: apart from reduplication, it pertains to para-linguistic features such as ludlings (language games or secret languages), poetic metre, and naturalistic orthographies. Thus, Riau Indonesian is argued to meet the definition of an isolating language.

Chapter 2, "The loss of affixation in Cham: Contact, internal drift and the limits of linguistic history", by Marc Brunelle, provides a historical overview of Chamic languages from first-millennium inscriptions until the present day, positing three distinct stages in the development of isolating word structures. Most of the original Malayo-Polynesian morphology was already lost prior to the first inscriptions; then from the 9th to the 19th centuries the amount of affixation remained more or less constant; finally, in the last century, a second wave of simplification resulted in the largely monosyllabic and affixless profile of Colloquial Eastern Cham. The driving force behind the latter development is argued to be phonological; namely, iambic rhythm, and the reduction and subsequent loss of the initial syllable of a disyllabic word. To the extent that language contact plays a role, it is of an indirect nature here; the borrowing of periphrastic syntactic constructions from Vietnamese provides

for an alternative to the affixes that are lost through phonological erosion. Thus, it is suggested that the contemporary isolating profile of Chamic languages can be accounted for without recourse to explanations involving long-term intensive contact between Chamic and Mon-Khmer languages or populations.

Chapter 3, "Dual heritage: The story of Riau Indonesian and its relatives", also by David Gil, continues where Chapter 1 left off, but switches from a synchronic to a diachronic perspective, asking how Riau Indonesian as well as other varieties of Malay/Indonesian and other related languages acquired their isolating word structure. Previous accounts by John McWhorter and others characterise Malay/Indonesian varieties as creoles or Non-Hybrid Conventionalised Second Languages, arguing that their isolating profiles are due to contact-induced simplification that took place subsequent to the break-up of proto-Malayic. This chapter argues that the lion's share of the simplification occurred much earlier, as far back as the original intrusion of Austronesian languages into what is now western Indonesia. In addition, however, it is argued that the simplification was due not only to imperfect second-language acquisition but also to assimilation to the typological profile of the pre-Austronesian languages of the Mekong-Mamberamo region. In this sense, then, Riau Indonesian and its relatives exhibit dual heritage: Austronesian and Mekong-Mamberamo.

Chapter 4, "Voice and bare verbs in Colloquial Minangkabau", by Sophie Crouch, is concerned with Minangkabau, focussing specifically on the hitherto undescribed colloquial register, which differs in many ways from the more familiar standard variety of Minangkabau. The chapter provides a detailed description and analysis of the "bare verb" construction, in which the verb appears without the voice markers that are mostly obligatory in the standard language, suggesting that whereas standard Minangkabau exhibits an "Indonesian-type" voice system, colloquial Minangkabau is characterised by a "Sundic-type" voice system, in which the compositional semantics are based largely on a looser relationship of "associationality". The proposed analysis of Minangkabau thus brings it into line with closely related Malay/Indonesian, in which colloquial varieties are often more highly isolating than the standard language.

Chapter 5, "Javanese undressed: 'Peripheral' dialects in typological perspective", by Thomas J. Conners, also shifts the spotlight away from a familiar standard language and towards its less well-known colloquial varieties – in this case Javanese. The chapter presents a contrastive analysis of Central Javanese, which forms the basis for what is often taken as a standardised variety, and five "peripheral" dialects: Banten, Banyumasan, Pesisir Lor, Tengger and Osing. The main finding is that peripheral Javanese dialects tend to be more highly isolating than Central Javanese. More generally, the peripheral dialects tend to exhibit a greater number of features characteristic of the Mekong-Mamberamo linguistic area. Thus, Central Javanese is something of an outlier with respect to both Javanese and the larger region. The

chapter argues that the more elaborate grammatical structures of Central Javanese are innovative, and that the more highly isolating typological profile of peripheral Javanese dialects is the more conservative.

Chapter 6, "Are the Central Flores languages really typologically unusual?", by Alexander Elias, moves the discussion to south-eastern Indonesia, to the languages of the Central Flores subgroup. Over a series of publications, John McWhorter has pointed to Lio, Ende, Nage, Keo, Nga'o, Ngadha and Rongga as examples of unnaturally reduced languages. He proposes that the Central Flores languages lost their morphology under conditions of language shift, due to imperfect learning and subsequent simplification by adult learners from Sulawesi in recent times. In response to this claim, the chapter argues that the typology of the Central Flores languages is more economically explained by a single early contact event occurring at the time of the original Austronesian settlement of Flores. The author points out that most of the typological features which make Central Flores unusual for an Austronesian language are the features proposed by Gil (2015) as characterising the Mekong-Mamberamo area.

Chapter 7, "From Lamaholot to Alorese: Morphological loss in adult language contact", by Marian Klamer, continues the move eastward, treating the closely related Austronesian languages of the Lamaholot cluster, spoken from the eastern tip of Flores to the Alor archipelago. The chapter presents a comparison of the morphological profiles of sister languages Lewoingu-Lamaholot spoken on Flores, and Alorese spoken on Alor, Pantar and some small islands in between. Showing that Alorese has less morphology than Lewoingu-Lamaholot, the chapter considers the role of contact with speakers of the non-Austronesian languages of Alor and Pantar in causing simplification in Alorese. The chapter posits that trade contacts, intermarriage and slave exchanges led to adult speakers of non-Austronesian languages in Pantar and Alor acquiring a morphologically richer precursor of today's Alorese as a second language. The claim is that the morphologically simplified variety they used as adult learners of the language developed into the isolating Alorese that we see today.

Chapter 8, "Double agent, double cross? Or how a suffix changes sides in an isolating language: *dór* in Tetun Dili", by Catharina Williams-van Klinken and John Hajek, focuses on how the highly isolating *lingua franca* of East Timor, Tetun Dili, adapts an agent suffix from Portuguese. The chapter shows that Portuguese *-dór* has two instantiations in Tetun Dili, the one as *-dor* on Portuguese borrowings and with Portuguese semantics, and the other as *door* with a native-like independent word status and distinct semantics. In showing how the Portuguese suffix has been reanalysed in the direction of a lexical element within a compound comparable to other native forms with similar functions, the chapter provides a counterpoint to other chapters in the volume. Rather than dealing with how a language

becomes isolating, it looks at how a language *stays* isolating when borrowing from a morphologically-rich language.

Chapter 9, "The origins of isolating word structure in eastern Timor", by Antoinette Schapper, returns to the theme of how isolating word structure developed in Austronesian languages. The chapter methodically deconstructs McWhorter's hypothesis of the languages of eastern Timor being 'simple' due to non-native acquisition. Presenting a wealth of data from both the Austronesian and Papuan languages of Timor, the chapter demonstrates that, although isolating, the languages are by no means simple and display many features that are complex according to McWhorter's metric. The chapter shows that adequate explanation of morphological loss is found in patterns of phonological changes shared across eastern Timor languages and that there is no need for recourse to a hypothesis involving radical simplification as a result of heavy non-native acquisition to explain the existence of isolating structure in eastern Timor.

Chapter 10, "Becoming Austronesian: Mechanisms of language dispersal across southern Island Southeast Asia and the collapse of Austronesian morphosyntax", by Mark Donohue and Timothy Denham, presents a wide-ranging perspective on the spread of Austronesian languages throughout Island Southeast Asia and how they acquired their isolating profile. The chapter highlights the fact that the Austronesian languages are situated in the broader Southeast Asia region which as a whole has a preponderance for heavily isolating word structure, only tending towards greater morphological complexity as the languages approach New Guinea in the east and South Asia in the west. The chapter argues that this contemporary linguistic situation projects back in time, indicating that a large number of isolating characteristics were already present in the languages of the region prior to the dispersal of Austronesian languages. In turn, this suggests, the authors argue, that the more isolating characteristics of many contemporary Austronesian languages can be attributed to contact with languages that were already strongly isolating.

Chapter 11, "Concluding reflections", by John McWhorter, offers a personalised overview of the main issues considered in the preceding chapters of the volume. Its focus is on the hypothesis, argued for elsewhere by the author, to the effect that grammatical simplicity and isolating word structure cannot develop on their own, but only as the result of radical language contact and imperfect adult second-language acquisition. In earlier chapters, three authors, Gil (Chapter 3), Elias (Chapter 6), and Schapper (Chapter 9) argue against various aspects of McWhorter's accounts of isolating structure in Austronesian languages. In this chapter, McWhorter proposes rebuttals to these and other arguments against his hypothesis. The volume thus ends in healthy disagreement, which, we hope, will provide impetus for future investigations into Austronesian languages, and how and why some of them became isolating.

References

Arka, I Wayan, Kosmas, Jeladu & Suparsa, I Nyoman. 2007. *Bahasa Rongga: Tatabahasa* (The Rongga Language: Grammar). Jakarta: Acuan Ringkas, Penerbit Universitas Atma Jaya.

Baird, Louise. 2002. A Grammar of Kéo: A Language of East Nusantara. PhD dissertation, Australian National University, Canberra ACT.

Conners, Thomas. 2008. Tengger Javanese. PhD dissertation, Yale University, New Haven CT.

Crouch, Sophie. 2009. Voice and Verb Morphology in Minangkabau, a Language of West Sumatra, Indonesia. MA dissertation, University of Western Australia, Perth WA.

Dryer, Matthew S. 2005. Prefixing vs. suffixing in inflectional morphology. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds), 110–113. Oxford: OUP.

Gil, David. 2015. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia. The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: De Gruyter Mouton.

Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistic Society* 7(1): 394–415.

Kluge, Angela. 2014. A Grammar of Papuan Malay. PhD dissertation, University of Leiden, The Netherlands.

Sneddon, James Neil. 2006. *Colloquial Jakartan Indonesian*. Canberra: Pacific Linguistics.

Tadmor, Uri. 1995. Language Contact and Systematic Restructuring: The Malay Dialect of Nonthaburi, Central Thailand. PhD dissertation, University of Hawai'i at Manoa, Honolulu HI.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change.* Honolulu HI: University of Hawai'i Press.

Williams-Van Klinken, Hajek & Nordlinger, Rachel. 2002. *Tetun Dili. A Grammar of an East Timorese Language.* Canberra: Pacific Linguistics.

Yanti. 2010. A Reference Grammar of Jambi Malay. PhD dissertation, University of Delaware, Newark DE.

What does it mean to be an isolating language?

The case of Riau Indonesian

David Gil

Max Planck Institute for the Science of Human History

This chapter poses the question "What does it mean to be an isolating language?" and addresses it by offering a case study of such a language, Riau Indonesian. First, this chapter surveys the debate concerning the viability of the notion of word as a comparative concept, proposes a definition of word as a cut-off point between two distinct levels of structure, morphology and syntax, and then follows with a definition of an isolating language as one lacking a robust structural unit of word. Next, the chapter presents an extensive exploration of wordhood in Riau Indonesian, examining 14 potential sources of evidence for word structure. Overall, the evidence for wordhood is shown to be sparse, thereby justifying the characterisation of Riau Indonesian as an isolating language and at the same time demonstrating what an isolating language may look like.

Keywords: Riau Indonesian, isolating, wordhood, ludlings

1. Introduction: The challenge of orthography

What does it mean to be an *isolating* language? This paper proposes a general answer to this question, and then applies it to a case study of a particular isolating language: Riau Indonesian.

As commonly understood, an isolating language is one in which words are of minimal morphological complexity. A purely isolating language would be one in which each word consisted of exactly one morpheme; however, such languages are not known to exist. Instead, languages may approach the ideal isolating type to varying degrees, as their morphological complexity decreases, and the morphemes contained within each word become fewer and less distinctive.

There is a problem, though, with the above notion of isolating language, namely that it relies crucially on the notion of *word*. *Prima facie* this should not

be a problem; after all, we all know what a word is. But do we really? For sure, our word-processors do: for them a word is what occurs between two spaces. But orthography is an imperfect and potentially misleading mirror of linguistic reality. To begin with, most languages and language varieties are still largely unwritten; moreover, even for languages that are written, we cannot be sure that orthographic conventions are a reliable reflection of how the language works. On the contrary, what writing systems demonstrate is that the notion of word is actually very problematic.

These issues may be amply illustrated in the Indonesian context. As in other parts of the world, the recent rise and proliferation of social media have set the stage for a spontaneous, naturalistic, and bottom-up development of orthographic systems for colloquial varieties of Malay and Indonesian that were previously rarely or never written. Such novel orthographic systems may provide valuable insights into the mental representations of native speakers; in particular, by examining the division of the text into strings of letters separated by spaces, we may try to make inferences with regard to the writer's notion of wordhood – see Section 3.3.14 below for an application of this method. What emerges, however, is a complex picture involving some consistency but also a great deal of variation.

In Indonesia, the overwhelming majority of social media interactions take place in colloquial varieties of Indonesian, whose novel orthographies, while largely parasitic on that of Standard Indonesian, still differ from it, and of course from each other, in numerous ways. For some studies of the language of social media in Indonesia see Gil (2004a), Manns (2010), Jukes (2015), and Brugman & Conners (2018). Example (1) below is a Facebook status update in Riau Indonesian. In (1a) the text is shown the way it was actually written, exhibiting typical social-media features such as multiple repetition of letters, the use of the numeral "2" to indicate phonological repetition, and of course emoticons. In (1b) the same text is represented in a normalised orthography, based on that of Standard Indonesian, but modified to take into account the main findings of this chapter, together with interlinear gloss and translation into English.[1] The text is a humorous representation of a parent's exasperation at his infant son's unwillingness to eat and sleep on time, and of his own ultimate submission to his son's capriciousness; it consists

---

1.   The major challenge facing the normalised orthography, employed in this chapter, is of course the decision of where to insert spaces between orthographic words. Since the orthography makes reference to the typology of morphemes proposed in Section 3.2.2, it is presented there, in footnote 27 below. Apart from that, the orthography is mostly straightforward, with only the following points worthy of note. There are two digraphs, *ng* [ŋ] and *ny* [ɲ]. For some speakers, orthographic *e* may correspond to two distinct phonemes, realised as [e] and [ə] respectively; see further discussion in footnote 36 below. Final orthographic *k* is usually realised as [ʔ]; see detailed discussion in Section 3.3.8 below.

of two syntactically and semantically parallel short dialogues between father and son occupying the first three and latter three lines of the original text respectively.[2]

(1) a. *Makan siang nyaa….???????*
         Nanti ajaa…
         Okeee… 😵
         bo2 siangnya…?????
         Nanti laaahhh
         yoooo weeeesss lah nak 😵 😎

    b. *Makan **siangnya** / Nanti aja / Oke*
        eat   midday:ASSOC   FUT.PROX   NEG.FOC   okay
        'Eat your lunch / Later / Okay'
        *Bobo **siangnya** / Nanti lah / Yo wis lah nak*
        sleep midday:ASSOC   FUT.PROX   FOC   yes   PRF   FOC   HYP\child
        'Take your nap / Later / Yeah, okay son'

Of interest in the above example is the orthographic representation of the associative marker -*nya*. In keeping with the parallelism between the two halves of the text, the form occurs twice, at the end of the first and fourth lines, in each case following *siang* 'midday'. However, in spite of the parallelism, the form is written differently in both cases: while in the former, it is a separate word, *nyaa*, in the latter it is joined on to the preceding word, *siangnya*. What this example suggests, then, is that the writer cannot make up his mind with regard to whether the associative marker *nya* is an independent word or part of the word preceding it.

What is true of naive writers and speakers is equally true of professional linguists. When describing unwritten languages, linguists often make arbitrary decisions with regard to wordhood, and as a result, different scholars may end up representing the same form in different ways. The following example from Papuan Malay shows the exact same expression as cited by two different sources:

(2) a. *sa   pu   bapa*   (Kluge 2014: 377)
         1SG   POSS   father
         'my father'

    b. *sa   = pu   = bapa*   (Donohue & Sawaki 2007: 260)
         1SG = POSS = father
         'my father'

---

2. In (1) and subsequent examples, the passage of text that is the focus of discussion is indicated in **boldface**. As in many naturalistic texts in Riau and other colloquial varieties of Indonesian, Example (1) displays instances of code-switching: *yo* 'yes' is from Minangkabau, *wis* (PFCT) from Javanese, and *bobo* 'sleep' from a specialised child-directed speech register of Indonesian.

How many words are there in (2)? Kluge assumes three, while Donohue and Sawaki's representation in terms of a stem preceded by two proclitics suggests that in some respects at least it is more like a single word. However, neither source provides explicit arguments in support of their decisions, and so, a typologist wondering whether Papuan Malay is an isolating language does not know what to think. Of course, such cases are not specific to Malay/Indonesian: Dixon and Aikhenvald (2002: 30–31) provide a similar example of the conflicting word boundaries adopted by the early missionaries in their attempts to write Fijian.

The problem is not restricted to unwritten languages with their emerging orthographies and associated linguistic transcriptions. Even for written languages with conventionalised orthographies purporting to represent word boundaries, we have no assurance that the spelling does indeed reflect linguistic reality. Standard Indonesian, reflecting more than a millennium-long tradition of writing, provides numerous examples of this. Toponyms in Indonesia often consist of a juxtaposition of two typically disyllabic words, but if there is a "rule" that specifies whether the two words are to be written separately or as a single compound word, I have not been able to uncover it. For example, for the capital city of Riau Province, one sees both *Pekan Baru* and *Pekanbaru*; a google search (28 August 2018) found 2.54 million occurrences of the former as opposed to 48.8 million occurrences of the latter.[3]

So is it *siang nya* or *siangnya*, *sa pu bapa* or *sapubapa*, *Pekan Baru* or *Pekanbaru*? Obviously, the answer to this question will have a crucial bearing on whether the respective varieties of Malay/Indonesian should be classified as isolating languages. Examples such as these can be multiplied at will – see Section 3.3.14 below for some results of a corpus study of text messages in Riau Indonesian, exhibiting some consistent patterns alongside a great deal of variation. And of course, similar cases of orthographic variability are hardly specific to Malay/Indonesian; other examples can be adduced from pretty much every part of the world; see, for example, Peterson (2008: 34–39) on the Munda language Kharia.

Even when the prescriptive orthographic rules are upheld consistently and without significant variation, one cannot uncritically assume that the rule reflects

---

3.   Analogous separate and joined-on spellings are also attested for the corresponding strings in English, *New Market* vs. *Newmarket*. However, the Indonesian case differs from the English one in (at least) the following three respects: (a) the Indonesian toponym refers to a single location rather than to many different ones; (b) since *pekan* is a somewhat archaic word for 'market' (with additional related meanings 'week' and 'town'), the string *Pekan Baru* is much less likely to have the compositional non-toponymic meaning 'new market' than its English counterpart; and (c) whereas the Indonesian toponym has but a single pronunciation, the English has two significantly distinct pronunciations, depending on whether the primary stress is on the first or second term – a distinction reflecting the grammatical status of the expression as a compound or a phrase.

linguistic reality. Standard Indonesian is said to have two homophonous forms: a "locative" *di* written separately, and a "passive" *di-* written joined on to its following host word. However, I am not aware of any substantive reason to distinguish between the wordhood properties of these two forms.[4] Again, similar problems replicate themselves from around the world. Orthographic conventions do not usually represent the product of serious grammatical analysis of the language that they represent. As pointed out by Dixon and Aikhenvald (2002: 8), Haspelmath (2011b) and others, just a simple tweak in the writing system, replacing one arbitrary convention with another equally arbitrary one, can be enough to recast an isolating language as an agglutinating or polysynthetic one, or vice versa. Indeed, in some cases, such an orthographic tweak might arguably offer a more insightful way of looking at the language; see, for example, Arkadiev's (2005) suggestion that Spoken French is more appropriately written as, and accordingly also viewed as, a polysynthetic language.

Thus, orthographies do not provide an obvious answer to the difficulties in determining word boundaries. In some instances, when treated with due caution, they provide useful insights into word structure; see Section 3.3.14 for a detailed study of one such case. However, in many other instances they may actually prove to be misleading. In recognition of these issues, and several other important ones, Haspelmath (2011b) presents arguments for a radical position, namely that there is no such thing as a word, or, more precisely, that linguists have not yet developed a comparative concept of a word that can be consistently invoked as a tool for cross-linguistic typological studies. A corollary of Haspelmath's position is that there is no such thing as an isolating language, that is to say, no viable distinction between isolating languages and languages of other morphological types, such as agglutinative, fusional, and polysynthetic. And an inconvenient consequence of that would be that a volume such as the present one, whose scope is defined in terms of the notion of isolating language, is thematically incoherent.

The goal of this chapter is thus to put the discussion of isolating languages back onto the firm footing to which it is entitled, and to establish a solid conceptual framework for the discussion of morphological typology. Section 2 examines Haspelmath's critique in detail, and shows how, in spite of the many

---

4.   In Gil (2002b) it is argued that in Riau Indonesian, there is actually a single form *di-* with a broad range of functions encompassing both "locative" and "passive"; these arguments probably carry over in large part also to Standard Indonesian. And indeed, as shown in Section 3.3.14 below, they are written in more or less the same way in Riau Indonesian. To the extent that the Standard Indonesian form *di(-)* is also macrofunctional rather than homophonous, this would provide further support for the claim that their distinct spelling does not reflect any substantive differences in wordhood.

reasonable points that he makes, it is, nevertheless, possible to come up with a viable cross-linguistically valid concept of word, and as a corollary, of isolating language. Specifically, word is defined as the structural unit that serves as a cut-off point between two distinct kinds of linguistic structure, commonly referred to as morphological and syntactic. Isolating languages, then, are ones in which there is relatively little morphology, and the structural unit of word plays a limited role in the grammatical organisation of the language. Following on this, Section 3 presents a detailed exploration of wordhood in Riau Indonesian, showing how this variety provides an instantiation of the isolating language type. Of the 14 potential sources of evidence for wordhood that are considered, just 6 turn out to be relevant to the notion of word; however, even these 6 fail to converge on a single structural unit, instead pointing towards two distinct kinds of word-like structural units. This leads to the conclusion that the notion of wordhood is only weakly defined in Riau Indonesian, which in turns shows how Riau Indonesian provides an instantiation of the isolating language type.

While Section 2 is highly abstract and may be of greater interest to general linguists, Section 3 provides an extensive description and analysis of the finer details of Riau Indonesian grammar and may thus be more appealing to linguists interested in various more specialised topics. Although the two sections may seem to embody very different concerns, each would be incomplete without the other. The only way to argue for the viability of an abstract notion such as word is to look in detail at the linguistic facts; conversely, the brute facts alone are meaningless unless they are analysed in terms of more abstract concepts. Only by combining the abstract with the concrete, the general with the specific, is it possible to progress towards a better understanding of the nature of wordhood, and more generally of grammar.

2.   The notions of wordhood and isolating language

John (1:1) may have said that "In the beginning was the word" – but in terms of linguistic architecture, he got it round the wrong way. The notion of word is a derivative and emergent notion that can only be understood in terms of two ontologically prior notions: *morpheme* and *hierarchical structure*. Thus, while it is well-nigh impossible to imagine a human language without morphemes and hierarchical structure, it is actually quite easy to imagine one without words.

2.1   Morphology vs. syntax

In order to appreciate the derivative nature of wordhood, it is instructive to compare language to another complex product of human cognition, namely, tonal music. A simplified representation of some of the major similarities and differences between the structures of tonal music and language is given in Figure 1 below.
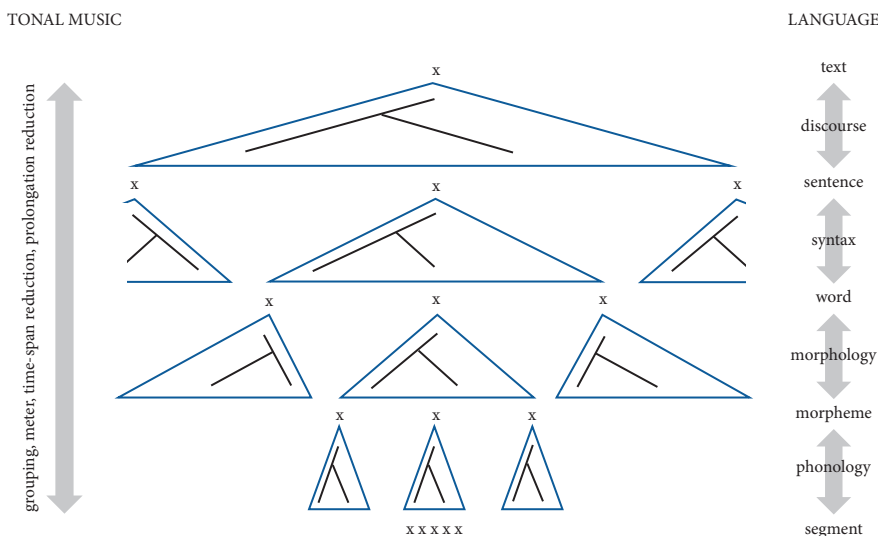
Figure 1. Tonal music and language

The big tree structure in the centre of Figure 1 represents the fact that both tonal music and language are founded on hierarchical structures in which smaller units group together to form larger ones. While in tonal music, individual notes group together all the way up to the entire musical piece, in language, individual segments group together all the way up to the entire text.

However, it is almost impossible to state this obvious fact without affording due recognition to a crucial difference between the two domains. In tonal music, structure is *vertically homogeneous*: the same principles that govern the relationship between, say, two quarter notes at the bottom of the tree also govern the relationship between the first and second halves of the entire musical piece at the top of the tree. This homogeneity of structure is a central design feature of theories of tonal music such as that of Lerdahl and Jackendoff (1983), who posit four types of structure: meter, grouping, time span reduction and prolongation reduction, each of which enters into play throughout the entire structure, from bottom to top.

In contrast, linguistic structure is *vertically heterogenous*, with different sets of principles underlying structure at different architectonic levels. These different sets

of principles are associated with distinct subdisciplines within linguistics: phonology, morphology, syntax and discourse theory. Phonology deals mostly with how phonemes group together to form morphemes, morphology largely with how morphemes combine to form words, syntax mainly with how words are put together to form sentences, and discourse theory for the most part with how sentences combine to form texts. The boundaries between different subdisciplines are each demarcated by a privileged structural unit: the morpheme sets apart phonology and morphology, the word separates morphology and syntax, while the sentence constitutes the cut-off point between syntax and discourse theory.[5]

To what extent is the vertical heterogeneity of linguistic structure an essential design feature of language rather than a contingent one? Or to recast the question: Would we expect a newly discovered extra-terrestrial language also to exhibit such vertical heterogeneity? At least one aspect of vertical heterogeneity would indeed appear to be an essential design feature: the distinction between phonological structure and structures at higher architectonic levels. Unlike tonal music, language is a tool for the conveyance of conceptual representations, or meanings, and there are an immense number of different meanings that it is called upon to convey. While compositionality allows for a potentially infinite number of meanings to be expressed through combinations of signs, language must still have access to a very large inventory of primitive signs: the morphemes of the language. However, given the physical nature of the oral speech medium in which language is embedded, it would not be practical for the tens or hundreds of thousands of individual morphemes to be associated each with its own primitive and non-decomposable individual sound. The resulting distinctions would be impossible to produce, impossible to perceive, and – even if the physical difficulties were somehow overcome – most

---

5.   The existence of different types of structures associated with different architectonic levels does not preclude the existence of other kinds of structures that are common to the different levels of structure. An example of this is provided by head-modifier, or X-bar structures, which have been posited in phonology for syllable structure (Levin 1985) and for stress (Halle & Vergnaud 1987), in morphology for word structure (Lieber 1980; Jensen 1980; Williams 1981), in syntax for phrase structure (Jackendoff 1977), and in discourse theory for the structure of narrative texts (Shen 1985, 1988, 1989). Indeed, similar structures have also been posited outside of language, among others for tonal music (Lerdahl & Jackendoff 1983; Lerdahl 1992), under the guise of time-span reductions. Haspelmath (2011b: 68–69) argues that similar shared structures pose a challenge to a putative distinction between different levels of structure, specifically, morphology and syntax; however, this need not be the case. An alternative approach, upholding the distinctions between different levels of linguistic structure, is offered by Gil (1985), suggesting that such shared head-modifier and X-bar structures should be "factored out of" tonal music and language and attributed instead to an autonomous mental faculty, which may subsequently be drawn upon by tonal music and by language – and within language independently by the distinct domains of phonology, morphology, syntax and discourse structure.

likely very difficult to learn. Analogous considerations hold also for sign languages and their motoric-gestural medium. Language solves this problem by constructing its morphemes out of combinations, mostly linear, of smaller units drawn from an inventory of typically just a few dozen building blocks that are lacking in meaning, namely phonemes. This is of course the well-known *duality of patterning* design feature of language proposed by Hockett (1960).[6]

However, while duality of patterning involves a sharp split between phonology and all the rest, it says nothing about the degree of vertical homogeneity associated with higher architectonic levels of linguistic structure. It would in fact be relatively easy to imagine a hypothetical artificial language whose structure, from the level of the morpheme upwards, was – like tonal music – vertically homogeneous, with just a single set of principles underlying linguistic structure all the way from morphemes up to entire texts. In such a language, you would just keep on putting morphemes together the same way, in accordance with a fixed set of linguistic principles, until you completed your text (be it *War and Peace* or a Facebook status update), and were done. Such a language would provide no basis for a distinction between morphology, syntax and discourse structure; putting it differently, such a language would have neither words nor sentences.

Is such a vertically-homogeneous language possible? With regard to the distinction between syntax and discourse theory, as far as I have been able to ascertain there has been little serious foundational discussion in the literature about the viability of the notion of sentence, either as a descriptive category in individual languages or as a cross-linguistically valid comparative concept.[7] We tend to take

---

6. It should be acknowledged that the boundary between meaning-bearing morphemes and meaningless submorphemic phonological units is not always clear cut. One large family of cases of apparently meaningful phonology is that subsumed under the header of *sound symbolism*. Another similar group of cases is sought out by a tradition of trying to identify more specific meanings associated with individual segments or strings of segments. A pioneering albeit highly speculative study of this kind is that of Fabre d'Olivet (1816) for Hebrew; a much more substantial recent study is that of Blust (1988), who identifies a set of submorphemic "roots" in Austronesian languages, such as, for example, *pit* 'press, squeeze together; narrow', instantiated in numerous larger forms such as the following from Riau Indonesian: *impit* 'press, squeeze', *jepit* 'squeeze between two objects', *sepit* 'squeeze between two objects', *sempit* 'narrow, crowded', *su(m)pit* 'blowpipe,' *sumpit* 'chopsticks', *Apit* [part of a toponym *Sungai Apit* 'Apit river', a town located on a long narrow peninsula between two bodies of water].

7. The only attempt that I am familiar with to examine the motivation behind the notion of sentence is that of Mithun (2005). In contrast, there is a substantial literature on the so-called "syntax/discourse interface", which, for the most part, deals with the ways in which the syntactic properties of sentences are affected by the larger discourse environments in which they are embedded. However, such discussions typically take for granted the *a priori* existence of individual sentences – whereas it is exactly that presupposition that is being questioned here.

it for granted, perhaps once again overly relying on the orthography, and in this case, the presence of a full stop marking the supposed boundaries between sentences – until, that is, we are confronted with real live naturalistic speech data, at which point the difficulty of parsing a text into sentences reveals itself. But that is the topic for a different paper.

In this chapter, my concern is with words, and the question being posed is: How viable is the distinction between morphology and syntax, and the notion of word as a cut-off point separating the two? Are words an essential design feature of the architecture of language, as presupposed by many, or a completely incoherent notion, as seems to be argued by others – or perhaps, somewhere between these two extremes, a variable parameter whose relevance may differ from one language to another, as in fact is suggested here?

The traditional view is that the internal structure of words, falling under the aegis of morphology, is of a fundamentally different nature to the ways in which words combine to form larger phrases, which is the remit of syntax. Thus, Katamba (1993: 217) writes that "the principles that regulate the internal structure of words are quite different from those that govern sentence structure", while Bresnan and Mchombo (1995: 181) assert that "words are built out of different structural elements and by different principles of composition than syntactic phrases". And indeed, there are excellent reasons not to abandon this view. To see this, it suffices to consider various design features of morphological structure that have no analogues in the domain of syntax. Following are two examples.

One is *non-linearity*. While in some cases morphemes may combine non-linearly to form words, words rarely or never combine non-linearly to form larger phrases. Imagine a hypothetical language in which a basic clause was formed via a template involving 7 positions, in which, invariably, the 2nd, 4th and 6th were occupied by the subject, and the 1st, 3rd, 5th, and 7th by the predicate. Sentences might look like this: *Jumped the over quick brown the lazy fox dog*. If the reader is totally baffled by this putative sentence, it may help to see it typographically parsed: ***Jumped*** THE ***over*** QUICK BROWN ***the lazy*** FOX ***dog***. No real language's syntax looks remotely like this, but replace subject with tri-consonantal root morpheme and predicate with inflection, and you have the well-known pattern of Semitic templatic morphology, e.g. Hebrew ***meкapecet***, where *-к-p-c-* is the root-morpheme 'jump', and ***me-a-e-et*** the inflection for PRS.SGF.

A second example is *paradigmatic syncretism*. Imagine a hypothetical language in which there were three open grammatical classes of nouns. For Class A nouns (the nice well-behaved kind) such as *ball*, *red ball* means 'red ball', *green ball* means 'green ball', and *blue ball* means 'blue ball'. But for Class B nouns such as *car*, *red car* means 'red car' but *green car* means either 'green car' or 'blue car' while **blue car*

is ungrammatical. And for Class C nouns such as *wall*, *red wall* means either 'red wall' or 'green wall', **green wall* is ungrammatical, while *blue wall* means just 'blue wall'. If this appears totally bizarre, well it is, for syntax that is – but in morphology it's commonplace. Just replace grammatical class with inflectional class, and *red*, *green* and *blue* with nominative, accusative and genitive case markers respectively, and you end up with something that looks like it's taken out of a description of a garden-variety Slavic language, with accusative-genitive syncretism for Class B nouns and nominative-accusative syncretism for Class C nouns.

The difficulty in even contemplating typical morphological design features such as non-linearity and paradigmatic syncretism being operative at the level of syntax underscores just how different morphology and syntax are from each other. Of course, it may be argued that phenomena such as non-linearity and paradigmatic syncretism are atypical of morphology, and that in general, morphology is actually much more similar to syntax than is suggested by these two features. However, what is at issue is not the existence of cross-domain similarities, such as, for example, basic principles of concatenation and constituency; of course, there are lots of such similarities. Rather, what is crucial is the existence of significant differences, design features that are uniquely morphological and that never manifest themselves in the syntax. It is precisely these latter features that provide support for the distinction between morphological and syntactic structures, and the importance of the notion of word as constituting the cut off point between these two architectonic levels.

2.2   The wordhood debate

An appreciation of the fundamental distinction between morphological and syntactic structures has been with us from the distant past all the way through to modern linguistic theories, ranging from Dixon's (2010–2012) Basic Linguistic Theory through Baerman, Brown and Corbett's (2017) typologically-oriented approach, to the generativists' "Lexicalist Hypothesis" – see Anderson (1982, 1985, 1992), Aronoff (1994) and others. Nevertheless, the distinction between morphology and syntax, and the role of the notion of word within linguistic theory, remain the subjects of lively debate and disagreement; see Dixon and Aikhenvald (2002) for a comprehensive if now slightly dated survey.

Much of this debate is couched in terms of a distinction between two analytical approaches, first referred to by Hockett (1954) as *Item and Process* and *Item and Arrangement*. Whereas an Item and Process model presupposes a distinction between morphology, which is amenable to such an approach, and syntax, which presumably is not, an Item and Arrangement model is neutral with regard to the

question whether morphology and syntax are governed by the same or different kinds of principles.[8]

Adopting an Item and Arrangement model, a number of recent scholars have chosen to downplay or disregard the significant differences between morphology and syntax, and instead propose models in which there is no place for the notion of word. In Chomsky's (1965) *Aspects* model, syntax subsumes morphology, and the terminal nodes of syntactic trees are "formatives", which correspond more or less to morphemes – within this system there is thus no such thing as a word. A similar word-less approach is maintained in the *Distributed Morphology* model of Halle & Marantz (1993); see also Embick (2015) and Bruening (2018). Of course, if there are no words, it makes no sense to ask how much internal structure they have, and hence there can be no meaning to the notion of isolating language.

Perhaps the most explicit and well-argued critique of the distinction between morphology and syntax, and the validity of the notion of word, is provided by Haspelmath (2011a,b). A central feature of his argumentation is the distinction between *descriptive categories*, which are emic properties of individual languages, and *comparative concepts*, which are etic constructs posited for the purpose of cross-linguistic typological investigations – see Haspelmath (2010, 2015, 2016, 2019a). Adopting the convention of using all-lower-case letters for cross-linguistic concepts alongside initial capitalisation for language-specific categories, Haspelmath's approach permits us to talk of, say, the *dative* marker in a certain language, dative being defined as an etic comparative concept, even if an emic analysis of the language in question does not support the existence of a language-specific *Dative* marker. For example, in Riau Indonesian, the form *sama* would instantiate a comparative concept of dative defined as the form marking the goal of a ditransitive construction based on the verb 'give', even though, as argued in detail in Gil (2004b), its broad range of functions, including among others transitive agent marker, comitative, coordinator and 'same', suggest that it is associated with a single more general meaning along the lines of 'together' – thereby precluding its characterisation as a Riau-Indonesian-specific Dative.[9]

---

8.   Note that the diagram in Figure 1 presupposes an Item and Arrangement approach to morphology; however, this is only for convenience sake; nothing in the following discussion depends on this particular aspect of the Figure 1 diagram.

9.   An obvious problem with Haspelmath's ontological distinction between language-specific descriptive categories and comparative concepts is that, modulo capitalisation, they generally make use of the same terms. For example, if the Russian Dative and the Georgian Dative are incommensurate with one another as also with the comparative concept of dative, then why are they all called Dative/dative? Using the same term for both language-specific categories and comparative concepts seems to suggest that they are related, which is precisely what Haspelmath is at pains to deny. The only terminological strategy that would uphold a strict ontological distinction between

In the case at hand, Haspelmath (2011b) is willing to countenance the possibility that some individual languages may in fact have a language-specific category of *Word* (pp. 60–62). However, he questions the viability of a cross-linguistically valid comparative concept of *word*, asserting that "[l]inguists have no good basis for identifying words across languages, and hence no good basis for a general distinction between syntax and morphology as parts of the language system" (p. 65). Or, elsewhere, "[l]inguists generally employ a range of different criteria, but these are not uniformly applicable across contexts and languages, and where they are applicable, they do not always converge. I conclude from this that we have no good basis for a general cross-linguistically viable word concept, and hence no basis for a general bifurcation between morphology and syntax." (p. 32).[10]

To support his assertion, Haspelmath identifies 10 criteria that are commonly invoked as criteria for wordhood, namely potential pauses, free occurrence, external mobility and internal fixedness, uninterruptability, non-selectivity, non-coordinatability, anaphoric islandhood, nonextractability, morphophonological idiosyncrasies, and deviations from biuniqueness; he then goes on to argue, in some detail (pp. 38–59), that "none of [these criteria] is necessary and sufficient for wordhood" (p. 38). This part of his argumentation is convincing, and perhaps not that surprising; after all, if any particular constellation of these criteria did provide for a good definition of wordhood, it would most likely have already gained wide acceptance among linguists.

Next, Haspelmath turns to examine possible ways in which criteria for wordhood might be applied in a graded or quantitative manner. One such method is that of "persuasion by test batteries" in which "a number of criteria are selected and applied, and in the published accounts usually all of them point in the same direction" (p. 59). Haspelmath objects to this method on the grounds that the criteria are often "selected opportunistically", and presents a table (p. 60) showing how 9 different published sources each select different subsets of 9 out of his 10 criteria, in their respective quests for a cross-linguistically valid definition of wordhood. However, as is argued below, such opportunism and variability need not necessarily be considered as a problem.

---

descriptive categories and comparative concepts is that of the tagmemicists, who introduce an arbitrary symbol to denote each and every language-specific descriptive category posited – but this results in prose that is unreadable to the point of uselessness.

10. Note that this is a different and much more far-reaching claim than the one argued for below, to the effect that some languages have words while other languages do not – where "word" is understood as a comparative concept whose definition is such that it can be meaningfully applied to any language in order to see whether it is met. Rather, Haspelmath is calling into question the very possibility of constructing a cross-linguistically valid comparative concept of word.

Another approach considered by Haspelmath is to define word as a "fuzzy concept" (p. 62). Building on an earlier proposal by Gil (1999a,b), Haspelmath suggests setting up a "boundness scale" (p. 63) that would rank the degree to which adjacent grammatical units are bound to each other; as noted by Haspelmath, "all languages have different degrees of tightness of minimal sign combinations" which is "quite possibly an innate design feature of language" (p. 62). Given such variation, a proponent of a cross-linguistically valid concept of word might then predict the existence of a clustering of boundness relationships, in which one cluster of stronger boundness relations would be word-internal, or morphological, while a second cluster of weaker boundness relations would be word-external, or syntactic (p. 64). In fact, such a model even allows for the possibility of a third intermediate cluster of mid-scale boundness relations corresponding to the notion of clitic. Although Haspelmath is rather sceptical with regard to such "fuzzy" approaches, dwelling on the conceptual issues involved in setting up such a scale, he concedes that "[t]he question is an empirical one, and should be approached in this way, rather than a priori" (p. 64). In fact, recent work has picked up on Haspelmath's challenge and attempted to explore some of the quantitative methods that might provide support for a comparative concept of word – for one such approach see Geertzen, Blevins & Milin (2016), for another Tallman et al. (2018), Tallman (to appear), and Tallman & Epps (2020).

Indeed, in spite of his sometimes outspoken scepticism with regard to the viability of a comparative concept of word, Haspelmath never quite slams the door shut on the possibility. Rather, it's a case of our still limited knowledge: "It is of course quite possible that we will eventually find evidence for something like a cross-linguistic word notion, but we will see that at the present state of our knowledge, we do not have the evidence." (p. 33) In fact, he invites other linguists to seek such evidence: "Linguists who believe that words exist as a cross-linguistically identifiable category should try to find ways of identifying words consistently." (p. 66)

Haspelmath is right that much work in linguistics rests on the shaky foundations of inadequately justified notions of what it means to be a word. And of course he is right that it is an open and empirical question as to whether a comparative concept of word will ultimately prove to be useful for linguistic typology – this is, after all, an inherent property of comparative concepts. However, his scepticism with regard to the prospects of a viable cross-linguistic notion of wordhood is somewhat overstated, and runs the risk of throwing the baby out with the bathwater. In order to see this, let us pick up on Haspelmath's invitation and examine the ways in which a cross-linguistically viable comparative concept of word can in fact be formulated in a way that satisfies his objections.

Haspelmath assumes that in order for a definition of wordhood to achieve the desired level of cross-linguistic validity, there must exist a universal menu of criteria

against which specific constructions in particular languages may be put to the test –
such as the 10 criteria that he lists in his paper.[11] Indeed, a similar assumption ap-
pears also to be shared by Tallman et al. (2018), Tallman (to appear), and Tallman
and Epps (2020), though Tallman (pers. comm.) is willing to countenance also the
possibility of language-specific criteria. The challenge faced by such approaches is
to determine the universal menu of criteria that underlie the comparative concept
being defined, that of word. And indeed, as Haspelmath shows convincingly, in the
case of word, such a universal menu of criteria just does not exist.

However, in order to define a comparative concept of word, there is no need
for such a universal menu of criteria. Instead, different languages can and should
be expected to be associated with different criteria for wordhood; see, for example,
Geertzen, Blevins & Milin (2016). In fact, it is precisely this move from the con-
crete idea of a fixed set of properties to the more abstract notion of a variable and
language-specific configuration of properties that makes it possible to entertain the
notion of a cross-linguistically viable comparative concept of word.

To see this, though, we must begin with a definition of Word as a language-specific
descriptive category:

(3)   Definition: *Word (language-specific descriptive category)*
      If, in a language L, a set of properties $P_1 \ldots P_n$ converges to define a structural
      unit X such that:
      a.   Xs are never smaller than individual morphemes, and in some cases are
           larger;
      b.   there is a substantial difference between the principles governing the inter-
           nal structure of Xs and those governing the combinations of Xs into larger
           units;
      Then X is a Word in L.

In the above definition, the properties $P_1 \ldots P_n$ that define a Word in L are
language-specific: different languages may make use of different sets of properties.
For example, in a language L_1, properties $P_1 \ldots P_5$ may converge, more or less, to de-
fine a language-specific structural unit which we might refer to as an L_1 Word, while
in another language L_2, properties $P_6 \ldots P_{12}$, disjoint from $P_1 \ldots P_5$, may combine,
in most part, to define a language-specific structural unit which we might similarly
refer to as an L_2 Word. Said properties $P_1 \ldots P_{12}$ may be just about any properties

---

11.   This approach would seem to be modelled on a seminal paper by Keenan (1976) dealing with
the difficulties of defining the notion of subject. Keenan proposes a list of universally relevant
subjecthood properties, and argues that a NP is subject-like to the extent that it displays a larger
subset of subject properties; in addition, a language as a whole is "subject prominent" to the extent
that it has NPs that are highly subject-like.

that apply over a certain domain; they could be drawn from any of Haspelmath's 10 properties above, but also from any number of phonological properties pertaining to stress, tone sandhi, vowel harmony, consonant assimilation and so forth, however idiosyncratic and language-specific such properties may be. Thus, in terms of substance, L_1 Words and L_2 Words are completely different kinds of objects; their similarity is of a more abstract nature, both constituting structural domains defined by the convergence of respective sets of language-specific properties.

Condition (3a) above ensures that Words are not submorphemic, nor for that matter consistently coextensive with morphemes; in at least some cases, they have to contain more than a single morpheme. Condition (3b) encapsulates the heterogeneity of linguistic structure as per Figure 1, characterising Words as cut-off points between distinct architectonic levels, morphology and syntax. If, in a language L, some alternative set of principles defines units of smaller or of larger size, such domains will fail to meet condition (3b), and we will accordingly seek other terms for them, such as, perhaps, Stem or Phrase. Thus, Stems are distinguished from Words by the presence of morphological structure not just within but also outside of them, while Phrases are distinguished from Words by the presence of syntactic structure not only between but also within them.

A further possibility needs to be acknowledged at this point. In some languages, there may exist two disjoint sets of properties converging on different structural units, each of which seems, *prima facie*, to warrant the term Word. In such cases, there is no reason whatsoever not to characterise both of these structural domains as Words, which, in lieu of more revealing terms, one might refer to as Word 1 and Word 2 respectively.[12] One commonly cited instance of this is the distinction, argued for in some languages, between Phonological and Grammatical (or Morphosyntactic) Words. Of course, in addition to Phonological and Grammatical Words, there could be any number of alternative language-specific Word 1 and Word 2 domains based on different sets of properties.

The definition of the language-specific descriptive category of Word in (3) sets the stage for a straightforward definition of word as a comparative concept:

(4)   Definition: *word (comparative concept)*
      A word is a structural unit that is a Word in some particular language.

---

12. An analogous use of such terminology can be found in the practice of referring to two quite different syntactic constructions in Malay/Indonesian as Passive 1 and Passive 2, predicated on the assumption that they both constitute instantiations of a comparative concept of passive – see, for example Dardjowidjojo (1978) and Sneddon (1996).

In accordance with (4), the comparative concept of word is identified with language-specific descriptive categories of Words. In a nutshell: words are Words. Haspelmath's (2010, 2015, 2016) definition of comparative concept automatically precludes such an identification. According to Haspelmath, language-specific descriptive categories and comparative concepts are ontologically distinct entities; whereas the former are social categories that exist in the minds of native speakers, the latter are abstractions whose only existence is in the mind of the comparative linguist. However, in Gil (2016) it is argued that, *pace* Haspelmath, comparative concepts may, in some cases, be identified with language-specific descriptive categories. Such an identification then makes it possible to assert that L_1 Words and L_2 Words constitute two language-specific instantiations of a comparative concept of word. And indeed, more recently, Haspelmath (2019a) seems to tentatively acknowledge the possibility of using the same descriptive categories to describe different languages, at least in the case of languages that are closely related to each other.

We are thus led back to a notion of wordhood akin to the original one outlined at the top of this section. As descriptive categories, Words are units defined by a convergence of possibly language-specific properties, and constituting cut-off points between principles of significantly different kinds. And as a comparative concept, word is simply the class of objects that are analysed as Words in individual languages. Thus, a language may be said to instantiate a comparative concept of word to the extent that there exists a set of criteria converging, more or less, on a structural domain that constitutes a cut-off point between two levels of structure differentiated by the kinds of principles that are generally associated with morphology and syntax respectively.[13]

These definitions are hardly new; they merely codify and give expression to the kinds of notions and understandings that most linguists have been working with all along. However, they do so in a way that makes reference to the distinction between language-specific categories and comparative concepts, while addressing Haspelmath's concerns regarding unwarranted appeals to the notion of word and the morphology/syntax distinction.

---

13. The above definition of the comparative concept of word does not preclude the possibility that other, alternative definitions may prove to be of equal value – for example, definitions based on quantitative analyses such as that of Tallman et al. (2018), Tallman (to appear), and Tallman and Epps (2020). The purpose of this section is simply to argue that a comparative concept of word is not an incoherent notion. If in fact competing definitions turn out to be viable, then this would only serve to underscore the relevance of a comparative concept of word for cross-linguistic typological studies.

2.3   Isolating languages

With the above understanding of the notion of wordhood in hand, we may now address the issue of morphological typology and isolating languages, beginning with the following definition:

(5)   Definition: *Isolating language*
      A language is isolating to the extent that it fails to instantiate the comparative concept of word.

In accordance with (5), a purely isolating language would be one without words, while familiar isolating languages are ones in which the comparative concept of word is only weakly applicable. Putting the definitions in (4) and (5) together, a language is isolating to the extent that it lacks a distinctive and robust language-specific descriptive category of Word. This definition follows in the footsteps of several others scholars, such as Hockett (1944: 255) for Chinese, who address the difficulty of identifying a viable word in isolating languages. It differs, though, from the position adopted by Dixon and Aikhenvald (2002: 32) and others, who maintain that all languages, even isolating ones, have words. The definition in (5) acknowledges that there is no such thing as a purely isolating language; rather, languages are isolating to various degrees, and some may be more isolating than others. Such variability in the degree to which various languages instantiate the isolating type is a theme that runs through several of the chapters in this volume.[14]

A language may fail to possess a category of Word for two interrelated reasons. As a conceptual starting point, consider the ideal case of a purely isolating language.

---

14.   In view of the fuzzy nature of the definition in (5), some scholars would like to have a quantitative yardstick making it possible to measure the degree to which a language is isolating. One often-cited yardstick is that of *morpheme-to-word ratio*, as calculated for a typical text in the language; however, this measurement is problematic as it presupposes a parsing of the text into words, which is precisely what is at issue here. Other quantitative methods, such as those of Tallman et al. (2018), Tallman (to appear), and Tallman and Epps (2020) overcome this particular issue, though as noted above, they are premised on somewhat different criteria to those adopted here. In this chapter, however, no such quantitative yardstick is proposed. Typological studies are full of comparative concepts that are fuzzy rather than discrete, have no agreed-upon quantitative yardstick for measurement, but have nevertheless proven their worth and led to new insights into the nature of language. For example, in the realm of word-order typology, categories such as SOV basic word order are commonly used, even though the notion of basic word order is fuzzy and potentially problematical, with some languages more strictly SOV but others less so, and SOV languages shading off into non-SOV ones with flexible or indeterminate word order. Word order studies ranging from Greenberg (1963) to Dryer (2005) are considered to be valuable contributions to the field even though they do not offer quantitative yardsticks for the determination of basic word order.

By assumption, in such a language there is no evidence whatsoever for a domain larger than the morpheme that might plausibly be referred to as a Word, and hence no morphological structure – everything from the morpheme upwards is syntax. As noted above, even languages that we might wish to characterise as isolating will deviate from this ideal to varying degrees. However, what is important to take note of is that they may do so in two different ways. In the first, a language may have a clearly-defined Word domain, but one that is only minimally larger than that of the morpheme. In such a language, the typical Word will contain a small number of morphemes, and will be endowed with minimal morphological complexity: this corresponds to the traditional understanding of the notion of isolating, alluded to in the beginning of this chapter. However, there is a second way in which a language may deviate from the ideal isolating type. In accordance with this second path, the Word domain will itself be weakly defined, due to a quantitative paucity of relevant properties, and their lack of convergence on a single unique domain. In Section 3 below, it is shown that Riau Indonesian takes both of these paths: to the extent that it has Words, they are both small and of limited internal morphological complexity, and also relatively weakly defined, since the relevant properties are both few and divergent.

The notion of isolating language acquires its meaning within the broader framework of *morphological typology;* see, for example, Sapir (1921). Although varying in several details, morphological typology typically revolves around the following four types:

(6) *Morphological typology*
   a. Isolating
   b. Agglutinative
   c. Fusional
   d. Polysynthetic

Morphological typology generally makes reference to two main structural features. The first is the degree to which adjacent morphemes within the word are phonologically bound to each other, ranging from low, in the case of agglutinative languages, to high, in the case of fusional and polysynthetic ones. The second is the complexity of word-internal structure, ranging from low in the case of isolating languages to high in the case of polysynthetic ones.

The usefulness of morphological typology has been called into question for several interrelated reasons, over and beyond its reliance on the notion of word. Some scholars have argued against the idea that variation in word-internal structure can form the basis for a holistic typology that makes predictions about other linguistic features. Going further, it is sometimes suggested that morphological typology is the product of an overly exoticising view of non-European languages. In

this view, morphological typology takes the classical fusional type of Latin/Green/ Sanskrit as its point of departure, and then attributes undue significance to particular languages as prototypes for the other respective types – Chinese and Mainland Southeast Asian languages for the isolating type, Turkish for the agglutinative type, and certain Amerindian languages for the polysynthetic type. For some cogent criticisms of morphological typology along these lines, see Haspelmath (2009) for the notion of agglutinative language, and, more recently, Haspelmath (2018) for that of polysynthetic language. Notwithstanding such criticisms, however, this paper suggests that the distinction between isolating and non-isolating languages, at least, is a well-defined and important parameter along which languages may vary, thus meriting a central role in linguistic typology.

Whereas the distinction between isolating languages and fusional or polysynthetic ones is relatively easy to draw, based on the first structural feature above, namely phonological boundedness, the distinction between isolating and agglutinative types is somewhat more challenging, relying as it does on a judicious application of the second structural criterion, involving wordhood. Haspelmath (2011b: 67) invites us to consider a typical Atlantic creole language with its verbal complexes containing pre-verbal tense-aspect, modality and polarity marking, complaining, not unreasonably, that such pre-verbal markers are typically written as separate words, although they "might well be prefixes" in which case an ostensibly isolating language would be recast as one that is agglutinative or polysynthetic. In order to adjudicate between the two alternative analyses, the grammatical patterns of the language must be considered in painstaking detail, in order to address the following question: Does there exist a set of linguistic properties $P_1 \ldots P_n$ that converge to pick out a structural domain satisfying the criteria in (3), and hence meeting the definition of Word? If such a set of properties can be identified, then the language would not be isolating. Conversely, if no such set of properties can be identified, then the language would indeed merit the characterisation as isolating.

But what should we do in the meantime, before we have had a chance to conduct an extensive grammatical analysis of the language in the quest for a set of potential linguistic properties that might support a structural unit of Word? Faced with a string of morphemes in a possibly new and unfamiliar language, the linguist is immediately confronted with a decision how to transcribe the form in question, and specifically, whether to insert spaces between the morphemes. As illustrated in (2) above, different linguists sometimes make different decisions. However, the discussion in this section suggests that, in the absence of positive evidence to the contrary, the default decision should be to write each morpheme separately, as for example in (2a). The reason for this is, simply, Occam's Razor: a language without words, and without a distinction between morphology and syntax, is simpler than one with words and with a morphology/syntax distinction. Or to put it differently,

the burden of proof is on the linguist who wishes to posit words, not on the one who tries to manage without them. For example, in the case of the creole language with its pre-verbal markers of tense-aspect, modality and polarity considered in the preceding paragraph, in the absence of specific evidence that these markers are indeed prefixes, they should indeed be written as separate words, and the language accordingly characterised as isolating. The same, of course, is also the case for the colloquial varieties of Indonesian of concern to us here. The default nature of isolating languages would hold true even if it were to turn out – as is argued, among others, by McWhorter (this volume) – that a large majority of the world's languages do indeed have words and a concomitant morphology/syntax distinction, and hence are not themselves isolating.

3. A case study: Riau Indonesian

Having established the viability of the notions of word and isolating language as comparative concepts, we are now in a position to embark on a detailed and extensive case study of a specific language, Riau Indonesian, in order to show how it instantiates the notion of isolating language.

3.1 Riau Indonesian

Riau Indonesian is the variety of Malay/Indonesian spoken in informal situations by the inhabitants of Riau and Kepulauan Riau provinces in east-central Sumatra, Indonesia. The population of these two provinces are linguistically and ethnically heterogeneous. Although the indigenous population is mostly Malay, a majority of the present-day inhabitants are migrants from other provinces, speaking a variety of other languages. Riau Indonesian is acquired as a native language by most or all children growing up in these two provinces, whatever their ethnicity – see Gil (this volume, Chapter 3) for more details. It is the language most commonly used as a lingua franca for inter-ethnic communication, and in addition is gradually replacing other languages and dialects as a vehicle for intra-ethnic communication.

Riau Indonesian is quite different from the Standard Indonesian familiar to many general linguists from a substantial descriptive and theoretical literature. It is also distinct from the dialects spoken by the indigenous ethnic-Malay communities, collectively known as Riau Malay. Riau Indonesian is one of many regional varieties of colloquial Indonesian that function as koinés in multi-ethnic communities, of which the most well described are perhaps Jakarta Indonesian (Sneddon 2006 and others), Ambonese Malay (van Minde 1997) and Papuan Malay (Kluge 2014). Although different from each other in numerous details, such varieties share much

of their typological ground plans. In particular, as suggested in Gil (this volume, Chapter 3) the analysis of wordhood in Riau Indonesian presented here is probably applicable, *mutatis mutandis*, to many other colloquial varieties of Indonesian, totalling tens of millions of native speakers.

The Riau Indonesian data presented in this paper are the product of several years of field work reported on in Gil (1994, 2001, 2002a,b, 2004a,b, 2005a,b, 2006b, 2009c, 2013 and elsewhere). Most of the data presented in Section 3 makes use of a naturalistic corpus of actual utterances, either jotted down right away into a notebook, or else recorded and subsequently transcribed.[15]

3.2   Wordhood in Riau Indonesian

Riau Indonesian comes close to not having words. To the extent that Riau Indonesian has a language-specific descriptive category of Word, the domain in question is not substantially larger than the domain of the morpheme, and in addition, it is supported by relatively few features and is thus of minor significance to the overall grammatical organisation of the language.

The gist of this section is thus a claim of non-existence, that of a robust and distinctive structural domain that might plausibly be characterised as a Word in Riau Indonesian. In principle, the defence of such a non-existence claim should consist of no more than a single sentence, to the effect that the investigator has conducted a detailed examination of the relevant patterns in the language, and found no evidence for its existence – this should be the end of the discussion. However, for better or for worse, the field requires more than that single sentence; it expects that the assertion of non-existence be backed up by a more detailed and fleshed-out story showing explicitly how the relevant domains of the grammar manage to get by and function without recourse to the non-existing item.[16] In this spirit, then, the remainder of Section 3 presents an extensive exploration of potential word structure in Riau Indonesian.

---

15. Some of the data presented here involves code-switching between Riau Indonesian and its various contact languages, primarily Minangkabau and Siak Malay – a variety of Riau Malay. Such data is commonplace in the multi-glossic linguistic landscape of Indonesia, and does not impact negatively on the validity of the argumentation.

16. This same methodological point is also made in Gil (2013) with respect to the noun/verb distinction in Riau Indonesian as well as, for good measure, other patently non-existent constructs such as the dual in English, and also in Conners, Bowden & Gil (2015) with respect to valency classes in Jakarta Indonesian.

3.2.1   *Word structure*

Evidence is provided for two competing systems of basic word structure in Riau Indonesian, which we shall refer to as P-Words and G-Words respectively.[17] Both systems share a common structural entity, the Core Foot:

```
                    Core Foot
                   /         \
                ($)           $
               /   \         /   \
            (O)     R     (O)     R
```
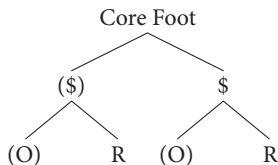
Figure 2. Core Foot Structure in Riau Indonesian

The Core Foot consists almost always of two syllables, $, though in rare instances the first of the two syllables is absent. Each syllable in turn consists of a Rhyme, R, preceded by an optional Onset O. Although the Core Foot is robustly supported in Riau Indonesian, it is not a viable candidate for a cross-linguistic comparative concept of word for the simple reason that it is sometimes wholly contained within a single morpheme, thereby violating condition (3a) and hence also (4) above.

The Core Foot is embedded within both P-Word and G-Word structures. P-Word structure is relatively simple, with the Core Foot flanked by optional P-Pre-Foot and P-Post-Foot positions, as represented in Figure 3 below.

```
                     P-Word
               /       |        \
       (P-Pre-Foot) Core Foot (P-Post-Foot)
```
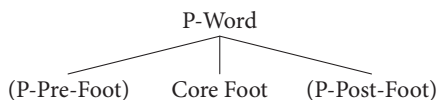
Figure 3. P-Word structure in Riau Indonesian

G-Word structure is somewhat more complex. As indicated in Figure 4, the Core Foot combines with an optional G-Pre-Foot position to form an G-Word.

```
                  G-Word
                 /      \
        (G-Pre-Foot)  Core Foot
```
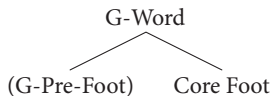
Figure 4. G-Word structure in Riau Indonesian

---

17. The modifications P- and G- allude, somewhat tentatively, to the often-made distinction between Phonological and Grammatical (or Morphosyntactic) Words; however, we leave open the question of whether the proposed distinction between P-Words and G-Words in Riau Indonesian corresponds to a cross-linguistically viable distinction between putative comparative concepts of phonological and grammatical word. As shown below, the sources of evidence for P-Words and G-Words do not split neatly into phonological and grammatical classes.

Moreover, as suggested in Figure 5, a G-Word may, in some cases, consist of two (or more) constituents each of which itself is a G-Word.

```
                        G-Word
              ┌───────────┴───────────┐
           G-Word                   G-Word
        ┌─────┴─────┐           ┌─────┴─────┐
  (G-Pre-Foot)  Core Foot  (G-Pre-Foot)  Core Foot
```
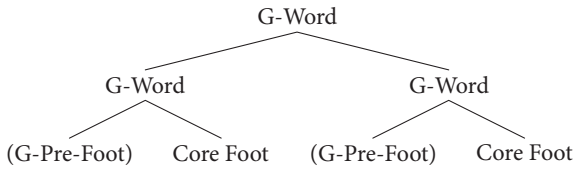
Figure 5. G-Word structure in Riau Indonesian (Recursion)

The tree structures in Figures 3, 4 and 5 make reference to two potential candidates for language-specific descriptive categories instantiating the comparative concept of word: P-Word and G-Word. However, as we shall see below, the evidence for these two language-specific descriptive categories, while substantial, is hardly overwhelming.

3.2.2 *Typology of morphemes*

The structures posited in Figures 2–5 above are supported by a typology of morphemes, as represented in Table 1 below with selected items:

Table 1. Typology of morphemes in Riau Indonesian

	Free		**Bound**	
Footed	*ayam* 'chicken'	*makan* 'eat'	*dari* 'from'	*dengan* 'with'
	merah 'red'	*aku* 1SG	*tentang* 'about'	*untuk* 'for'
	sama 'together'	*udah* PRF	*pada* 'than'	*kalau* TOP
	binatang 'animal'	*pel* 'mop'	*tiap* 'every'	*aja* NEG.FOC
Light	*blong* 'broken'	*wey* 'hey'	*di-* PAT, LOC	*ke-* ALL, ORD
	woooot ONOM	*kring* ONOM	*-kan* EP	*-nya* ASSOC
Suprasegmental	*e* Q	*kan* Q	*lah* FOC	*ma* EMPH
	ha DEIC	*yuk* EXHRT	*kek* DISJ.ASSOC	*do* NEG.POL

The first distinction, represented by the rows of Table 1, is between Footed morphemes, associated with a Core Foot, as specified in Figures 2–5 above, and morphemes lacking a Core Foot, which are, in turn, of two types, Light and Suprasegmental, the latter consisting of morphemes associated with a distinctive intonation contour. The second distinction, represented by the two columns of Table 1, is between Free morphemes, which can stand on their own, and Bound morphemes, which can only occur in construction with a host morpheme.[18] Resulting from

---

18. For Footed and Light morphemes, the distinction between Free and Bound morphemes corresponds to the distinction, made in Gil (2000, 2005b, 2013), between the two syntactic categories

these distinctions is a six-way morpheme classification. Of these six types, the Free Footed type is by far the largest, containing the overwhelming majority of morphemes in Riau Indonesian. In what follows, we shall be concerned primarily with the Free Footed and Bound Light types, as it is these two types that are of most direct relevance to the determination of Word structure in Riau Indonesian.

Among Bound Light morphemes, two further distinctions are shown in Table 2 below:

Table 2. Typology of Bound Light morphemes in Riau Indonesian

	Weakly Bound	Weakly/Strongly Bound	Strongly Bound
Preceding	*di-* PAT, LOC *ter-* NON.AG	*N-* AG (*me-*) *ber-* NON.PAT *ke-* ALL, ORD *se-* 'one', *i-* DEM *s-* LOC *g-* 'like'	*N-* AG (other allomorphs) *si-* PERS
Following	*-kan* EP *-in* EP *-nya* ASSOC	*-an* AUG	

The first distinction, represented by the rows of Table 2, is between Bound Light morphemes that precede their host and those that follow it. The second distinction, represented by the columns of Table 2, pertains to their relationship to the G-Word structure in Figure 4 above. Whereas Weakly Bound Light morphemes can only occur outside the G-Word, Strongly Bound Light morphemes can only occur within the G-Word. However, there is also an intermediate set of Weakly/Strongly Bound Light morphemes whose behaviour is ambivalent, alternatively occurring either outside of or within the G-Word.[19]

---

of Riau Indonesian. Free morphemes belong to the single open syntactic category S^0, consisting of expressions that may stand on their own as complete non-elliptical utterances, and combine freely with other such expressions. In contrast, Bound morphemes belong to the single closed syntactic category S^0/S^0, consisting of forms that can only occur in construction with members of S^0 to yield another S^0 – as suggested by the slash symbol in the categorial-grammar notation.

19. Since Weakly Bound Light morphemes occur outside of the G-Word, their relationship to their hosts is governed by principles pertaining to two other systems: P-Word structure (Weakly Bound Light morphemes share the same P-Word as their hosts), and syntactic structure, as per the preceding footnote.

The Bound Light morphemes that precede their host contain four markers of "generalised voice": patient-orientation *di-*, agent-orientation *N-*, non-patient-orientation *ber-* and non-agent-orientation *ter-*.[20] Two others are associated with a flagging function (in the sense of Haspelmath 2019b), locative *di-* and allative *ke-*, while yet others include the personal marker *si-*, the ordinal numeral marker *ke-* and the form *se-*, whose literal meaning is 'one' but which occurs in a number of other constructions. Finally, a set of Bound Light morphemes consisting of basic *i-*, locative *s-* and manner *g-* combine with the proximal and distal demonstratives *ni* and *tu* to form a mini-paradigm of basic, locative and manner deictic forms.[21] The Bound Light morphemes that follow their host consist of two generalised voice markers, *-kan* and *-in*,[22] the associative marker *-nya*,[23] and the form *-an*, associated with a wide range of meanings.[24]

---

20. In the literature on Malay/Indonesian, the counterparts of these four markers are often referred to as "passive", "active", "medial" and "involuntary passive" respectively; however, as argued in Gil (2002b) and Conners, Bowden & Gil (2015), their functions are primarily semantic rather than syntactic. The marker *N-*, often referred to as "prenasalisation", has a number of distinct morphophonemic realisations, discussed in detail in Section 3.2.7. See also discussion of the corresponding forms in closely-related Minangkabau in Crouch (this volume).

21. The present description differs from most traditional descriptions of the corresponding forms in other varieties of Malay/Indonesian, which characterise forms such as *ni, tu, ini, itu, sini, situ, gini, gitu*, as monomorphemic, in spite of their obvious transparent compositionality.

22. The markers *-kan* and *-in* share the function of generalised voice markers associated with a thematic role of end-point (glossed as "EP") or target; their functions thus correspond to those of causatives and applicatives in other languages. The choice between the two markers is governed by sociolinguistic factors. Even though *-kan* is also the Standard Indonesian form of the end-point marker, its use in Riau Indonesian is basilectal, being associated with more rural, lower-class, and socially inward-oriented speech. In contrast, the use of *-in* is generally characteristic of more urban, middle-class and socially outward-oriented speech, suggesting that it is a borrowing from Jakarta Indonesian, in which it functions as the sole end-point marker – see Connors, Bowden & Gil (2015) for details.

23. The marker *-nya*, described in more detail in Gil (2009a), asserts that its host is associated with some other entity, whose reference is determined contextually, by a combination of linguistic and extra-linguistic factors. For example, in (1), *siangnya* (midday:ASSOC) means something like 'midday associated with the given context', resulting in an interpretation resembling that of a definite article; whereas in (64), *SMSnya* (SMS:ASSOC) is understood as 'SMS associated with you', that is to say as a 2nd person possessive.

24. The marker *-an* occurs relatively infrequently and in mostly non-productive contexts. While its interlinear gloss suggests a shared semantic component of augmentative, evident in forms such as *jual* 'sell' ~ *jualan* 'sell habitually'/'trade', *cium* 'kiss' ~ *ciuman* 'kiss each other repeatedly', and *duri* 'thorn' ~ *durian* 'durian' (a kind of fruit with lots of thorns), its functions are so variegated and idiosyncratic that they are probably most appropriately described on an individual

Table 2 provides an exhaustive listing of Bound Light morphemes in Riau Indonesian, subject to the following three qualifications. First, the class of Bound Light morphemes includes two others that are devoid of segmental material and hence do not fit readily into the classification in Table 2: Reduplication, discussed in detail in Section 3.3.8, and Truncation, used to form hypocoristic expressions. Secondly, a small number of additional morphemes, exclusively monosyllabic, exhibit variable behaviour, in some cases that of Bound Light morphemes. In particular, the negative marker *tak*, the abbreviated perfect marker *dah*, the demonstratives *ni* and *tu*, and truncated terms of address such as *bang*, although primarily Free Footed, occasionally exhibit behaviour of the kind associated with Bound Light morphemes. Similarly, the relative marker *yang*, although primarily Bound Footed, sometimes exhibits properties of a Bound Light morpheme. Thirdly, a handful of additional Bound Light morphemes, occasionally occurring in texts, are either instances of code-switching with Standard Indonesian (or other colloquial varieties of Malay/Indonesian), or else borrowings from Standard Indonesian (or other varieties), in which latter case the morphemes are at least partly non-productive and opaque in Riau Indonesian.[25]

The morpheme types proposed in Tables 1 and 2 correspond imperfectly to traditional classifications of morphemes that make reference to potential comparative categories. What are sometimes referred to as "content" morphemes are exclusively of the Free Footed type, whereas their opposite number, so-called "grammatical" morphemes, may belong to five out of the six types – the exception being Free Light morphemes, which are of an ideophonic or onomatopoeic nature. Suprasegmental morphemes, either Free or Bound, are forms of the kind often referred to in the literature as "pragmatic particles" or "discourse particles".[26] Finally, Bound Light

---

item-by-item basis. In this respect, Riau Indonesian differs from other varieties such as Jakarta Indonesian, in which the use of *-an* is more frequent, and is at least partly productive, being associated with various grammatical functions as such as comparative. Most if not all forms containing *-an* in Riau Indonesian probably represent instances of borrowing, either from Jakarta Indonesian or from the standard language.

25. Examples of such morphemes cited in this chapter include *meN-* in *membawa* in (30), *membeli* in (33), and *membalas* in (64); *peN-* in *pengantin* in (33); *per-* in *pertemannya* in (i), footnote 43 – all from Standard Indonesian; and also *kat-* in *katSungai Apit* in (32), from Siak Malay. Other examples involve circumfixes such as *ke- -an* and *per- -an* – circumfixes being absent from the core list of native productive Bound Light morphemes presented in Table 2.

26. As argued in Gil (2002a), Suprasegmental morphemes are not part of the morphosyntax proper, and are not associated with any syntactic category; instead, they lie, together with their intonation contours, on a independent suprasegmental tier. Most commonly, they occur at the end of a phonological phrase; with their own distinctive intonation contours, they appear to constitute an island of tonality in an otherwise non-tonal language. In terms of their meanings,

morphemes are what are commonly referred to as "clitics" and/or "affixes", with the distinction between the two possibly corresponding to that between Weakly and Strongly Bound Light morphemes respectively.[27]

3.3   Sources of evidence

The word structure of Riau Indonesian posited in Figures 2–5 above is supported by the various sources of evidence shown in Table 3 below. In Table 3, columns represent distinct structural levels in accordance with Figures 2–5, while rows represent linguistic properties: an X in a cell means that the linguistic property in question provides evidence for the structural level as indicated.[28]

As suggested by Table 3, the viability of P-Word and G-Word levels in Riau Indonesian is only weakly supported. Of the 14 sources of evidence, the first eight pertain exclusively to the Core Foot, while only the remaining six are of direct relevance to the various notions of Word. Of these six, just one, Reduplication, is a hardcore grammatical property, while the remaining five are of a para-linguistic nature, involving poetic meter, ludlings, and orthography – see below for explication of these terms. All in all, it would be fair to say that support for the two Word levels in Riau Indonesian, while present, is hardly overwhelming; for the most part, the language gets by very well without recourse to any structural level corresponding to a cross-linguistic comparative category of word.

Suprasegmental morphemes are a very mixed bag, expressing idiosyncratic combinations of logico-semantic features such as deixis, focus and negative polarity, pragmatic functions such as emphasis, and affective value.

27. The typology presented in Table 1 forms the basis for the practical orthography for Riau Indonesian adopted in this chapter. Specifically, Bound Light morphemes are written joined on their hosts, while other morpheme types are written as separate orthographic words. While this corresponds largely to the conventions of Standard Indonesian orthography, there are some salient differences, involving morphemes whose Standard Indonesian counterparts are written separately but are joined on here: locative *di-*, allative *ke-*, personal marker *si-*, and truncated terms of address such as *bang-*. For Compound and Reduplication constructions, the individual terms are written separately, except in the case of Partial Reduplication, where they are joined on. No orthography is perfect; the present one represents a reasonable compromise between the structure of Riau Indonesian argued for in this chapter and the familiar orthographic conventions of Standard Indonesian.

28. In Table 3, parentheses around an X suggest that the source of evidence in question provides only weak support for the relevant level of structure. Coindexing of two Xs represents the state of affairs in which the source of evidence is ambivalent, supporting one of the two relevant levels of structure without it being possible to adjudicate between the two.

Table 3. Sources of evidence for word structure in Riau Indonesian

		Core Foot	P-Word	G-Word
1	the canonical morpheme	X		
2	focus intonation	X		
3	fast speech reduction	X		
4	epenthesis and spreading	X		
5	loan form expansion	X		
6	truncated name expansion	X		
7	realisation of *N-* as *nge-*	X		
8	final [k] → [ʔ]	X		
9	reduplication			X
10	Pantun prosody		X	
11	Sabaha ludling			X
12	Warasa ludling	X	X	
13	Bahasisa ludling	X (X_i)	(X)	(X_i)
14	naturalistic spelling		(X)	X

In the remainder of Section 3.3, we consider, in turn, each of the 14 sources of evidence listed in Table 3.[29] We begin with the eight sources of evidence that pertain exclusively to the Core Foot.

3.3.1   *The canonical morpheme*

Whereas the evidence for Words in Riau Indonesian is relatively weak, that for the Core Foot is more robust. The first and most straightforward source of evidence is quantitative, deriving from the lexicon: an overwhelming majority of morphemes are disyllabic, occupying the entirety of the Core Foot. These comprise a large majority of the Free Footed morphemes. e.g. *ayam* 'chicken', *merah* 'red', *makan* 'eat', *sama* 'together' and numerous others, plus also the members of the smaller category of Bound Footed morphemes. The large numerical predominance of disyllabic

---

29. Some of the sources of evidence summarised in Table 3 above have already been discussed elsewhere. Gil (1993) mentions Pantun prosody and its relationship to word structure (#10 in the above table). Gil (2002a) discusses in detail the Sabaha, Warasa and Bahasisa ludlings and how they provide evidence for foot and word structure (#11,12,13). Gil (2005a) discusses the relevance of reduplication to word structure (#9). And Gil (2006b) presents the major suprasegmental patterns and shows how focus intonation, fast speech reduction, epenthesis and spreading, loan word expansion, truncated name expansion and the Warasa ludling join forces to support the existence of a Core Foot (#2,3,4,5,6,12). The discussion in this section brings these different sources of evidence together for the first time, while adding additional and novel sources of evidence, in order to paint a unified and comprehensive picture of word structure in Riau Indonesian.

morphemes constitutes the most obvious and straightforward piece of evidence in support of the Core Foot in Riau Indonesian.[30]

Bearing this in mind, it is worth taking a closer look at the non-canonical cases, the exceptions to disyllabicity. One small residue of morphemes is tri-syllabic or more; these all belong to the class of Free Footed morphemes, e.g. *binatang* 'animal', *pramuka* 'scouts', *puskesmas* 'clinic' (abbreviated from *pusat kesehatan masyarakat* 'centre health society'), *Ramayana* [name of supermarket chain], and are largely if not exclusively neologisms or borrowings from other languages. As argued below, for these morphemes, the last two syllables occupy the Core Foot, while the preceding syllable or syllables spill over to preceding positions.

Another somewhat more heterogeneous residue of morphemes is mono-syllabic; these belong to three types, Footed, Light and Suprasegmental. Footed monosyllabic morphemes are Free morphemes which, as argued below, occupy the second syllable of the Core Foot, and are therefore preceded by an empty first-syllable position, e.g. *pel* 'mop', *tes* 'test', and *cat* 'paint'. In contrast, monosyllabic morphemes of the Light and Suprasegmental types constitute the entirety of their respective classes.

Nevertheless, monosyllabic morphemes remain in a minority, in terms of both types and tokens. The overwhelming predominance of disyllabic morphemes provides strong quantitative support for the existence of the Core Foot. Moreover, as we shall see below, the evidence for the Core Foot is not just quantitative but also qualitative. Riau Indonesian does not "like" morphemes that do not conform to the dominant disyllabic pattern; when it encounters them, it tries as hard as it can to beat them into shape, and fit them in to the Core Foot. The strategies that it employs for this purpose are discussed in the following subsections.

3.3.2   *Focus intonation*

In Riau Indonesian, stress falls on the final syllable of an intonational group that, depending on context, may range from a single disyllabic morpheme to a much longer string of morphemes, often an entire utterance. Thus, unlike in many other languages, stress does not provide any insights into Wordhood or the Core Foot.

What is of relevance, however, is an optional intonational feature, focus intonation, which assigns semantic and pragmatic prominence to the form to which it applies. Phonetically, focus intonation assigns suprasegmental prominence to the entire Core Foot, spanning both of its syllables in a distinctive way. Typically, the rhyme of the first syllable, and sometimes also the consonant in-between the two syllables, are significantly extended in duration, while the rhyme of the second

---

30. As noted by Blust (2009: 224–226) and others, the preference for disyllabicity is widespread across the Austronesian language family.

syllable is associated with a high falling pitch contour.[31] Thus, focus intonation selects the entire Core Foot, singling it out and marking it off against the surrounding material which remains less prominent.

Examples (7)–(10) below illustrate the occurrence of focus intonation in naturalistic speech, the prominent foot being marked in upper-case letters:[32]

(7) *Wooot* ***DATANG*** *rajawali* *tadi,* *mau nikam dia,* ***CUCUKnya***
ONOM arrive sparrowhawk PST.PROX want AG:stab 3 skewer:ASSOC
sat caaap
ONOM ONOM
[From tale about a sparrowhawk]
'Whoosh, the sparrowhawk came, he wanted to stab it, he skewered it wham'

(8) ***BiNATANG*** *kau*
animal 2SG
[cursing friend]
'You animal'

(9) *Aku bilang* ***BELOKkan*** *lagi*
1SG say turn:EP ADD.FOC
[Playing billiards on laptop computer; telling friend to rotate the cue]
'I said turn it around'

(10) *Makan* ***ramBUTAN***
eat hair:AUG
[Reproachfully commenting on friends eating rambutan without inviting speaker]
'Eating rambutan (a fruit named for its hairy appearance)'

The structure of the forms bearing focus intonation in Examples (7)–(10) is shown in Figure 6 below:[33]

---

31. For many speakers lengthening of the consonant between the two syllables applies only if the consonant in question is a continuant or a sonorant, as in the [l] of *BELOKan* in (9). However, for speakers of Riau Indonesian with a Minangkabau accent, lengthening may apply for any consonant, such as for example the [t] of *DATANG* in (7).

32. Interestingly, the convention of marking the prominent Core Foot with upper-case letters has been adopted by the ubiquitous Indonesian wireless carrier Telkomsel, most saliently with their "Simpati" packages, characteristically written *simPATI* – at the time of writing, numerous instances of this could readily be observed by searching for "Simpati Telkomsel" in Google Images. The success of this orthographic convention is, arguably, due, at least in part, to the pleasing alignment of the upper-case letters with the Core Foot in Indonesian.

33. In Figure 6, as well as subsequent Figures 7–12, Word structure above the level of the Core Foot is not indicated, as the relevant phenomena do not provide any evidence for it, and more specifically for whether the structure in question is that of the P-Word or the G-word.
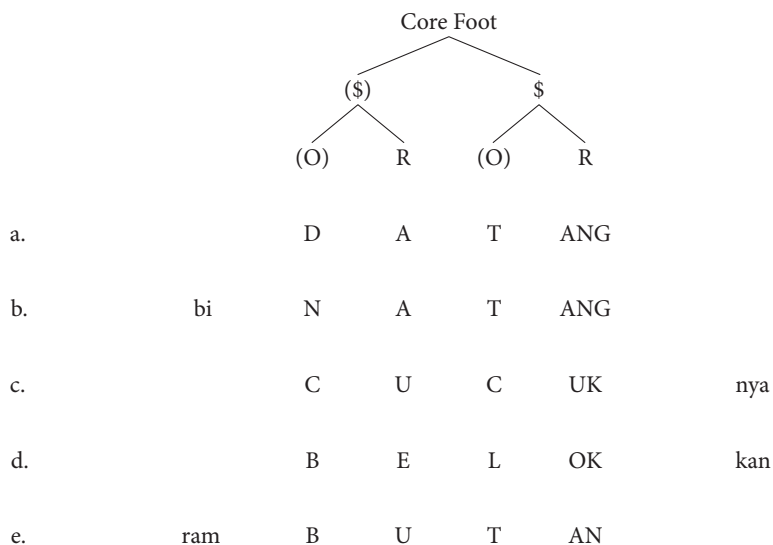
Figure 6. Focus intonation

In Figure 6 above, the first two examples show monomorphemic forms. Example (a) represents the simplest and most common case of a disyllabic morpheme that is coextensive with the Core Foot; focus intonation spreads across the entire form. Example (b) shows what happens with a trisyllabic morpheme; here, focus intonation occurs across the last two syllables, occupying the Core Foot, while the first syllable, outside the Core Foot, does not receive any prominence.[34] The next three examples show forms consisting of a disyllabic morpheme followed by a monosyllabic Bound Light morpheme, *-nya, -kan* and *-an*, and present an important contrast. While in (c) and (d), the Weakly Bound Light morphemes *-nya* and *-kan* fall outside the Core Foot and hence do not receive focus intonation, in (e), the Strongly Bound Light morpheme *-an* occurs within the Core Foot, thereby "pushing back" its host disyllabic morpheme, in order to share focus intonation with the second syllable of its host morpheme.[35]

---

34. Examples such as in (b) create an acoustic effect that is very different from that in languages such as English, in which, in a trisyllabic word with final stress, such as *macaroon*, secondary stress would fall on the first syllable. In contrast, in Riau Indonesian, in trisyllabic words with focus intonation, it is the adjacent penultimate and final syllables that are both more prominent than the first syllable.

35. It should be acknowledged that there is considerable cross-dialectal variation with regard to the realisation of focus intonation in the context of the end-point markers *-kan* and *-in* and the associative marker *-nya*. For example, in Jakarta Indonesian, *-in* and *-nya* behave just like *-an*, occurring within the Core Foot and accordingly sharing focus intonation.

In summary, then, focus intonation provides clear and straightforward support for the Core Foot in Riau Indonesian, assigning it greater prominence than other positions outside the Core Foot.

3.3.3 *Fast speech reduction*

While focus intonation increases the prominence of Core Foot positions, another phonological process, fast speech reduction, decreases the prominence of positions outside of the Core Foot. As shown in Figure 7 below, in fast speech, in trisyllabic monomorphemic forms such as *belanja* 'shopping', *komputer* 'computer' and *Bengkalis* [toponym], the antepenultimate syllable is optionally reduced, as in the (ii) variant, or even deleted, as in the (iii) variant:

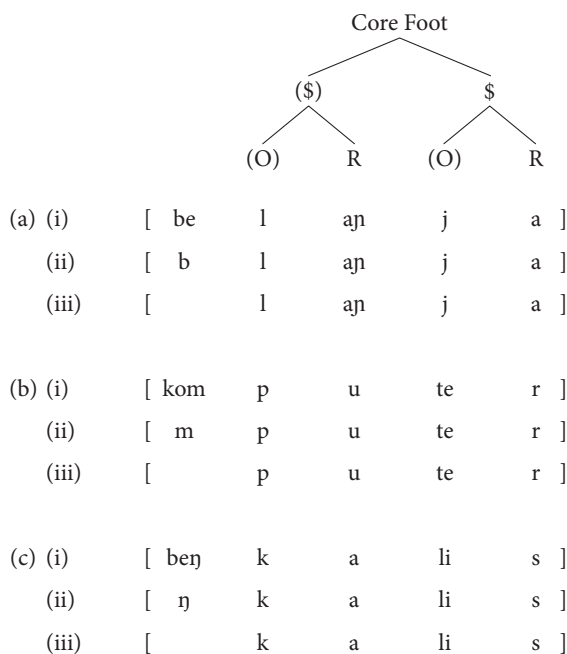| | | Core Foot | | | |
		($)		$	
		(O)	R	(O)	R
(a) (i)	[be	l	aɲ	j	a]
(ii)	[b	l	aɲ	j	a]
(iii)	[	l	aɲ	j	a]
(b) (i)	[kom	p	u	te	r]
(ii)	[m	p	u	te	r]
(iii)	[	p	u	te	r]
(c) (i)	[beŋ	k	a	li	s]
(ii)	[ŋ	k	a	li	s]
(iii)	[	k	a	li	s]

Figure 7. Fast speech reduction

The effect of fast speech reduction is thus to bring overly long morphemes into line with the predominantly disyllabic shape of Riau Indonesian morphemes, based on the Core Foot.

3.3.4   *Epenthesis and spreading*

A mirror-image process has the effect of beefing up monomorphemic forms that are too small, thereby increasing even further the predominance of disyllabic morphemes. In Riau Indonesian there are two subdialects, which may be referred to as the *schwa dialect* and the *schwa-less dialect* respectively.[36] In the former dialect, the schwa [ə] is part of the phonemic inventory, even though, in this dialect, it never occurs in the final syllable. Of concern here however is the second or schwa-less dialect, in which there is no phonemic schwa. Figure 8 below shows the way in which forms containing a penultimate schwa in the schwa dialect, such as *enam* 'six', *kering* 'dry' and *ketan* 'sticky rice', are realised in the schwa-less dialect. In each of the three examples, the first line shows the distribution of the segmental melody within the Core Foot, while the subsequent lines show the variant phonetic realisations of the forms in question. Whereas the realisation in (i) directly reflects the segmental melody, the subsequent realisations add phonetic material: in (ii) an epenthetic [ə], in (iii) an epenthetic [e], and in (iv), for *enam* and *kering,* a lengthening and syllabification of the sonorant consonant.

These alternative pronunciations are best accounted for in terms of a Core Foot in which most or all of the segmental melody is associated with the second syllable. Whereas in (a) the first syllable does not contain any material from the segmental melody, in (b) and (c) the first syllable contains just a single consonant in the onset. By filling in the otherwise empty rhyme of the first syllable, and enabling phonetic material to extend across the entirety of the Core Foot, epenthesis and spreading in (ii), (iii) and (iv) serve to beef up an impoverished segmental melody, and thereby contribute further to the maintenance of the predominantly disyllabic shape of Riau Indonesian morphemes, founded on the Core Foot.[37]

---

36. The distribution of these two dialects reflects a combination of ethnic and sociolinguistic factors. Ethnically, speakers of the schwa dialect are more likely to belong to ethnic groups whose associated languages have a phonemic schwa, as is the case in dialects of Malay and Javanese, while speakers of the schwa-less dialect are more likely to belong to ethnic groups whose associated languages lack a phonemic schwa, such as Minangkabau. Sociolinguistically, speakers may alternate between schwa and schwa-less dialects, the presence of schwa being associated with a higher speech register, bearing a closer resemblance to Standard Indonesian, which also has a phonemic schwa.

37. Forms such as *ketan* and *kering*, which are plentiful in Riau Indonesian, contrast with a much smaller set of forms whose segmental melody is also of the form CCVC, but which do not allow alternative pronunciations involving epenthesis or spreading. This much smaller set consists largely if not entirely of ideophones of a clearly expressive nature, such as *kring* and *blong*. For example, *kring* is invariably realised as a simple monosyllabic [kriŋ], never *[kəriŋ], *[keriŋ], or *[kr̩ŋ] (with syllabic [r̩]); similarly, *blong* is invariably realised as a simple monosyllabic [bloŋ],

Core Foot

	(O)	R		(O)	R
			($)		$
a.				n	am
(i)	[			n	am]
(ii)	[	ə		n	am]
(iii)	[	e		n	am]
(iv)	[		←	n	am]
b.	k			r	iŋ
(i)	[k			r	iŋ]
(ii)	[k	ə		r	iŋ]
(iii)	[k	e		r	iŋ]
(iv)	[k		←	r	iŋ]
c.	k			t	an
(i)	[k			t	an]
(ii)	[k	ə		t	an]
(iii)	[k	e		t	an]
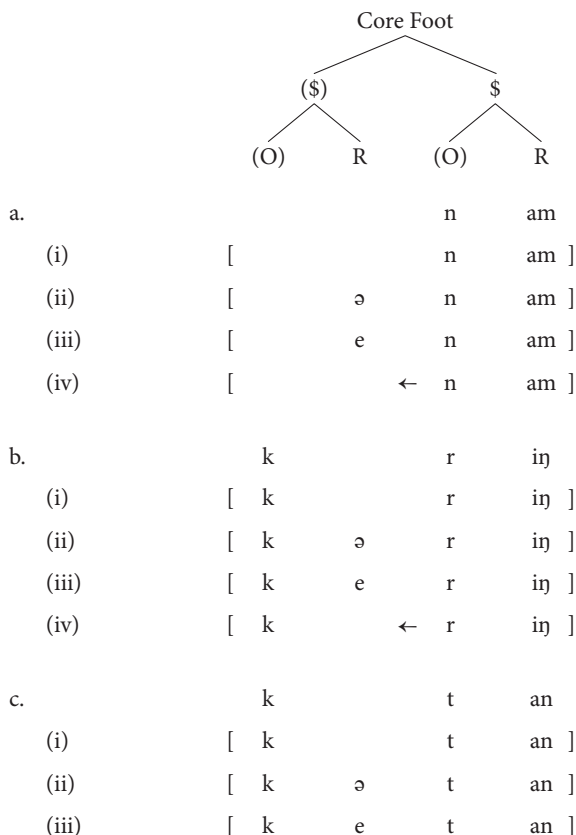
Figure 8. Epenthesis and spreading

3.3.5 *Loan form expansion*

The synchronic processes of epenthesis and spreading discussed in the preceding section are mirrored by a diachronic process of loan form borrowing, whereby a monosyllabic form in the source language is expanded to a disyllabic form occupying the entire Core Foot in Riau Indonesian. Depending on the phonological properties of the source form, three different patterns of expansion may be distinguished, as shown in (a), (b) and (c) respectively in Figure 9 below:

---

never *[bəloŋ], *[beloŋ], or *[bl̩oŋ] (with syllabic [l̩]). In the absence of any evidence in support of the existence of a preceding empty syllable, these forms may be accounted for assigning the entirety of the CCVC segmental melody to a single syllable, without recourse to the Core Foot; such morphemes are accordingly assigned to the class of Free Light morphemes.

Core Foot

		($)			$	
		(O)	R		(O)	R
a.		k	a		os	< *kous* (Dutch)
(i)		[k	a		os]	
b.		s	a	l	ap	< *zalf* (Dutch)
(i)		[s	a	l	ap]	
c.		g		l	as	< *glas* (Dutch)
(i)		[g		l	as]	
(ii)		[g	ə	l	as]	
(iii)		[g	e	l	as]	
(iv)		[g		← l	as]	
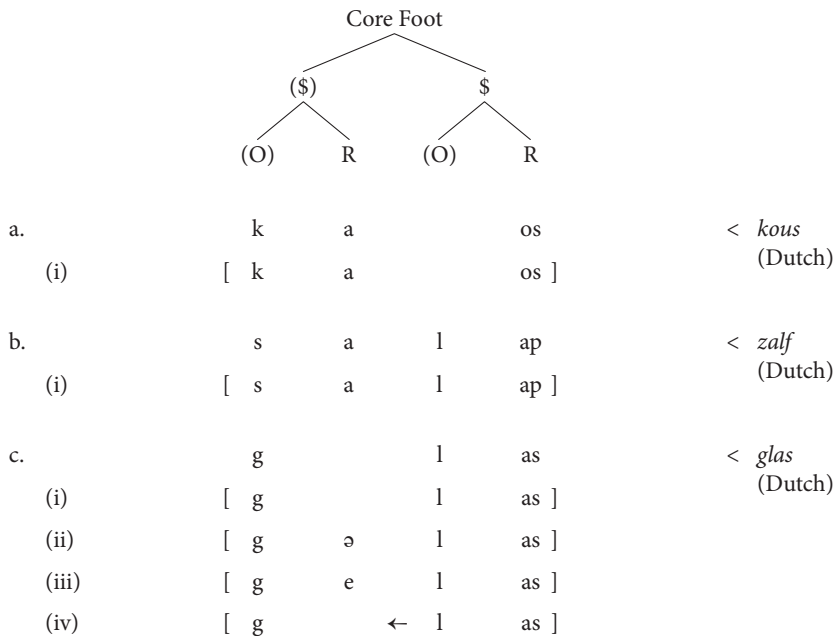
Figure 9. Loan form expansion

The first pattern of expansion involves source-language forms containing a diph-thong, which is expanded into a sequence of two vowels in Riau Indonesian, for ex-ample *kaos* 'T-shirt' < Dutch *kous* 'stocking'; *duit* 'money' < Dutch *duit*, the name of an old Dutch coin; *koin* 'coin', 'token' < English *coin*. The second pattern of expansion comprises source-language forms of CVCC structure, in which an epenthetic vowel is inserted between the two final consonants, for example *salap* 'ointment' < Dutch *zalf* 'ointment'; *helem* 'helmet' < Dutch *helm* 'helmet'; *filem* 'filem' < English *film*. The third pattern of expansion consists of source-language forms of CCVC structure, which end up with the same structure as the native forms discussed in Section 3.3.4, for example *gelas* 'glass', 'cup' < Dutch *glas* 'glass'; *keran* 'tap' < Dutch *kraan* 'tap'; *sepit* 'speedboat' < English *speed(boat)*. As the form is borrowed, its segmental melody is expanded to occupy the entire disyllabic Core Foot; then, once already in Riau Indonesian, it undergoes optional epenthesis and spreading just like in Figure 8 pre-viously. In all three patterns, then, the process of borrowing involves an expansion of an original monosyllabic form to a disyllabic form in Riau Indonesian, thereby providing further evidence for the existence of a disyllabic Core Foot.

It must be acknowledged, however, that in cases such as these, it is not always possible to trace the precise path of the borrowing; in fact, it is more than likely that in at least some cases, the form in question was borrowed from the donor language

first into some other variety of Malay/Indonesian, one that was in more direct contact with the donor language, and only later made its way, through inter-dialectal borrowing and diffusion, to Riau Indonesian. Obviously, this is the case for borrowings that are known to have taken place at some earlier stage in the history of Malay/Indonesian.[38]

To the extent that this is the case also for the above examples, the argument presented above would appear to apply not to Riau Indonesian but rather to whatever dialect of Malay/Indonesian first borrowed the form. Nevertheless, given the fluidity and indeterminacy of dialect boundaries, it is not unreasonable to consider the form in question as being borrowed into a superordinate language variety also including Riau Indonesian, in which case the above argument could legitimately be construed as applying also to Riau Indonesian.

More recently, though, with the advent of mass media followed soon after by that of social media, it has become possible to observe clear cut cases of borrowing from foreign languages directly into Riau Indonesian and other similarly peripheral varieties, without the mediation of more central and prestigious varieties of Malay/Indonesian. One example of this is the form *semek*, meaning, roughly, to 'lift someone up and then smack them to the ground'. This form was observed to enter Riau Indonesian in the late 1990s, via television, following the introduction into US professional wrestling of the brand name *Smack Down;* given its ubiquity and popularity throughout Indonesia, it is reasonable to assume that the form *semek* was borrowed directly from English into Riau Indonesian – which would then provide further support for the reality of the diachronic process of expansion into the Core Foot in Riau Indonesian.

---

38. For example, the *Loanword Typology* database of Tadmor (2009) for Standard Indonesian contains several instances of forms present also in Riau Indonesian that were borrowed into Early Malay during a period said to extend from 680 to 1620 AD, which also exhibit expansion from an original monosyllabic form into a disyllabic one. One example is *(e)mas* 'gold', borrowed from Middle Khmer *mās* 'gold' (which in turn was borrowed from the disyllabic Sanskrit *māṣa*). Several examples come from Arabic, e.g. *akal* 'mind', 'reason', 'intellect', 'intelligence' < ʕ*aql* 'sense', 'reason', 'mind'; *sihir* 'black magic' < *siḥr* 'sorcery', 'witchcraft', 'magic'; *fajar* 'dawn' < *fajr* 'dawn', 'daybreak'. Other examples involve the borrowing of an originally disyllabic form in Arabic, but one in which the second syllable is a case-marking suffix, e.g. *waktu* 'time' < *waqt-u* (time-NOM); *Sabtu* 'Saturday' < *sabt-u* (Saturday-NOM). As pointed out by Tadmor, in Malay/Indonesian words of ultimate Arabic origin, the case marker is "usually lost", as in fact is also the case in most colloquial varieties of Arabic; accordingly, its retention in examples such as these may also be attributed to the pressure to uphold the disyllabic Core Foot. Examples such as these and others suggest that the relevance of the Core Foot to the structure of Malay may date back a significant amount of time.

3.3.6   *Truncated name expansion*

While the above three arguments for the Core Foot pertain to single morphemes, the next two sources of evidence involve the combination of a disyllabic morpheme with a Bound Light morpheme.

The first of these involves a productive process of truncation in which typically disyllabic names and address terms are reduced to monosyllabic variants, for example *Saipul > Pul* (below), *Dapit > Pit* (in several subsequent examples), *anak > nak* in (1), and *oom > om* in (16). The syllable retained is most often the second, though in some cases it is the first. Truncated names are most commonly used vocatively; however, they often occur in non-vocative constructions, where they are more fully integrated into the syntactic structure of the sentence. It is in these non-vocative contexts that the effects of the Core Foot may be observed.

Truncation creates monosyllabic forms, but as we have already seen, Riau Indonesian does not like monosyllabic forms; accordingly, it seeks ways to undo the "damage" wreaked by truncation and restore the forms to their preferred disyllabicity, thereby replenishing the Core Foot. Two different strategies are available for this. The first is the personal marker *si-*, a Bound Light Morpheme which may occur in front of proper names, for example *siSaipul* 'Saipul', and in front of other forms in order to convert them into names, for example *siompong* 'Gap-tooth', *sikancil* 'Mouse deer'. The second is a set of kinship terms, such as *abang* 'elder brother', *bapak* 'father' and others, which, most commonly in their own truncated forms, may be placed in front of names or other expressions, for example *bangSaipul, bangompong* and *bangkancil*. (The two strategies are in complementary distribution: the personal marker *si-* cannot co-occur with a kinship term.)

The crucial fact, pertaining to the non-vocative contexts, is shown in Figure 10 below: Whereas the personal marker *si-* and the kinship terms are optional in the case of full disyllabic names, as in (a), the presence of one of them is obligatory, or near obligatory, in the case of non-vocative truncated names, as in (b). Thus, the ungrammaticality (or near-ungrammaticality) of monosyllabic truncated names in non-vocative contexts provides additional support for the preference for disyllabicity and, *ipso facto*, for the Core Foot.[39]

In some cases, expansions of truncated monosyllabic names appear to have undergone lexicalisation, and are reinterpreted as new monomorphemic names.

---

39. It should be noted that, compared to some of the other sources of evidence discussed in Section 3.2, truncated name expansion would appear to be less straightforwardly generalisable to other varieties of Malay/Indonesian. For example, Jakarta Indonesian seems to be more tolerant of monosyllabic truncated names in non-vocative contexts. Yet other dialects, such as Papuan Malay, do not make use of truncated names at all, so the "problem" of monosyllabicity does not arise in the first place.

Core Foot

		($)		$	
		(O)	R	(O)	R
a.		s	ai	p	ul
	si	s	ai	p	ul
	bang	s	ai	p	ul
b.	?*			p	ul
		s	i	p	ul
		b	ang	p	ul

Figure 10. Truncated name expansion

Most commonly, this involves a reduced variant *i* of the personal marker *si-*. For example, the truncation of *Saipul* and subsequent addition of the personal marker resulting in *siPul* is subsequently reanalysed as a "new" name, *Ipul*. Another strategy for the expansion of truncated monosyllabic names involves initial CV- reduplication, for example *Ulis > Is > Iis*, *Putra > Put > Puput*. These strategies differ from the ones discussed previously, involving the addition of *si-* or a term of address, in that they are idiosyncratic rather than productive. Nevertheless, their effect is the same, namely to restore disyllabicity, thereby providing further support for the Core Foot.

3.3.7   N- *realised as* nge-

Like many other Austronesian languages, Riau Indonesian has a Bound Light morpheme, commonly written as *N-*, whose function is to mark its host as being oriented towards an associated agent, and whose form typically involves some kind of *prenasalisation*; see Gil (2002b) for detailed description. While the occurrence of *N-* is licensed by semantic and pragmatic factors, its distribution is also governed by phonological features of the host form (see below), as well as by arbitrary lexical properties: while some forms frequently host prenasalisation, others do so less commonly, others rarely, and yet others never – for reasons that do not seem amenable to any generalisations.

The form that *N-* assumes depends on the phonological properties of its host. For disyllabic and longer hosts, the form of prenasalisation varies in accordance

with the first segment of its host, as follows. If the first segment is an unvoiced stop or continuant, it is replaced with a homorganic nasal, for example *putar > mutar* 'revolve', *tulis > nulis* 'write', *curi > nyuri* 'steal', *kejar > ngejar* 'chase', *sampe > nyampe* 'reach', *hisap > ngisap* 'suck'. If the first segment is a voiced obstruent, the host rarely or never undergoes prenasalisation. If the first segment is a nasal or an approximant, the host never undergoes prenasalisation. If the first segment is a liquid, then it is preceded by *me-*, for example *lempar > melempar* 'throw', *rokok > merokok* 'smoke'. Finally, if the first segment is a vowel, then it is preceded by *ng-*, for example *angkat > ngangkat* 'lift'. However, in contrast with disyllabic and longer hosts, when the host is monosyllabic, *N-* assumes an invariant form, *nge-* (realised as either [ŋə] or [ŋe]), regardless of the first segment of the host, for example *pel > ngepel* 'mop', rather than **mel*; *tes > ngetes* 'test', not **nes*.

The rationale behind the realisation of *N-* as *nge-* is evident in Figure 11 below. In Figure 11, in each pair, bare forms in (i) are contrasted with prenasalised forms in (ii). Examples (a) and (b) contrast disyllabic and monosyllabic hosts in which the first segment is *p*, while (c) and (d) compare disyllabic and monosyllabic hosts in which the first segment is *t*. As can be seen in (b/ii) and (d/ii), the *nge-* allomorph has the function of expanding a monosyllabic form to a disyllabic one occupying the entirety of the Core Foot. Thus, the realisation of *N-* as *nge-* with monosyllabic hosts provides yet additional evidence for the existence of the Core Foot.[40]
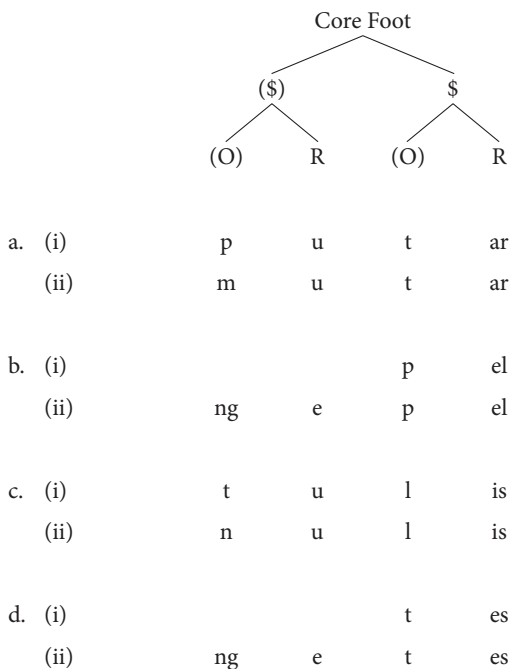
3.3.8   *Final [k] → [ʔ]*

The final source of evidence for the Core Foot derives from a phonological rule that changes the unvoiced velar stop [k] to a glottal stop [ʔ] when it occurs the end of the appropriate structural domain.[41]

The predominantly disyllabic nature of morphemes in Riau Indonesian makes it convenient to talk of three positions, initial, medial and final, corresponding to the three C slots in the typical CVCVC template. In Riau Indonesian, [k] occurs frequently in initial position, in forms such as [kulit] 'skin', [kilat] 'shiny', [kejar] 'chase' and many others, as well as in medial position, in forms such as [rakit] 'raft',

---

40. It should be acknowledged that whereas the *nge-* allomorph of *N-* brings its host into conformity with the Core Foot, the *me-* allomorph has the opposite effect of taking a disyllabic host and adding on a third syllable lying outside the Core Foot. However, this adverse effect is often mitigated by subsequent application of the rule of fast speech reduction discussed in Section 3.3.3 above.

41. Outside of Riau Indonesian, the distribution of this rule is rather variable. For example, amongst Malay/Indonesian varieties, it is present in Kuala Lumpur Malay but not Papuan Malay. Similarly, in Javanese, it is present in many varieties while absent in Banyumasan – see Conners (this volume).

Core Foot

		(O)	R	(O)	R
a.	(i)	p	u	t	ar
	(ii)	m	u	t	ar
b.	(i)			p	el
	(ii)	ng	e	p	el
c.	(i)	t	u	l	is
	(ii)	n	u	l	is
d.	(i)			t	es
	(ii)	ng	e	t	es

Figure 11. *N-* realised as *nge-*

[nakal] 'naughty', [makan] 'eat', and numerous others; however, in final position, its occurrence is limited to a small number of loan forms, such as [ojek] 'motorcycle taxi', from Javanese, and [lotek] (a kind of salad), possibly from Sundanese. In contrast, [ʔ] occurs frequently in final position, in forms such as [budaʔ] 'child', [busuʔ] 'rotten', [tabraʔ] 'collide' and many others, whereas in non-final position, it is much less frequent: medially, it occurs in just a few loan forms such as [maʔap] 'forgive' from Arabic, baby-talk forms such as [eʔẽ] 'poop', and nicknames formed by phonological copying such as [aʔãp] "Aap", while initially, it does not occur, except as an automatic phonetic gesture before forms beginning with a vowel. On the basis of these distributional facts, it is reasonable to conclude that initial [k], medial [k] and final [ʔ] are instantiations of a single phoneme. Moreover, given the universal preference for lenition over fortition, the most plausible analysis is to characterise this phoneme as *k*, and to posit a rule changing [k] to [ʔ] in final position.[42]

---

42. The question remains how to analyse the forms that do not conform to the generalisation concerning the distribution of [k] and [ʔ]. Final [k] is probably best analysed as an instance of "loan phonology" – the cross-linguistically well-attested state of affairs in which borrowed forms are subject to different principles than their native counterparts. As for non-final [ʔ], this is perhaps most appropriately considered to represent a somewhat marginal phoneme, occurring in

The psychological reality of the [k] → [ʔ] rule in Riau Indonesian is supported by two further meta-linguistic arguments. First, in the more prestigious Jakarta dialect that many speakers are familiar with from television and elsewhere, final [k] is retained, in forms such as [budak] 'child', [busuk] 'rotten', [tabrak] 'collide', and others. Secondly, in the Standard Indonesian orthography that many speakers learn in school, these forms are also, for the most part, written with a *k: budak, busuk, tabrak*.[43] Although we are "not supposed to" use evidence from other dialects, or from orthography, in support of linguistic analyses, there is no doubt that in the Indonesian context, these factors may potentially play a significant role.

The question arises what is the precise structural domain that constitutes the licensing environment of the [k] → [ʔ] rule. Given forms such as [kulit], [rakit] and [budaʔ], the realisation, in the latter form, of *k* as [ʔ] could equally well be attributed to its position in the coda of the syllable, at the end of the Core Foot, or at the end of a larger Word unit. So far, I have not been able to find any arguments specific to Riau Indonesian that would make it possible to adjudicate between these three alternatives.[44]

---

loan forms, as well as in other special registers such as baby talk and nicknames. In addition, this phoneme may arguably also be present in final position, in a ordinary register of the language, in at least two sets of high-frequency forms associated with specific semantic domains: negation, e.g. [taʔ], [ndaʔ], [ŋaʔ] 'no'; and kinship terms, e.g. [bapaʔ] 'father', [ibuʔ] 'mother', [adiʔ] 'younger sibling' and others. Nevertheless, it remains to be shown that these forms cannot be derived from a phonemic *k* and the [k] → [ʔ] rule – an argument that lies beyond the scope of the present paper.

43. A small class of exceptions to this generalisation consists of a handful of forms that end with a [ʔ] but whose Standard Indonesian spelling does not contain a *k*, for example, [kenaʔ] 'undergo', written as *kena*; [bukaʔ] 'open', written as *buka*; [mintaʔ] 'request', written as *minta*. Crucially, speakers of Riau Indonesian often "mis-spell" such forms by adding a final orthographic *k*, writing *kenak, bukak, mintak*. The following Facebook chat example, formatted in the same way as Example (1) above, shows an instance of the spelling *kenak*:

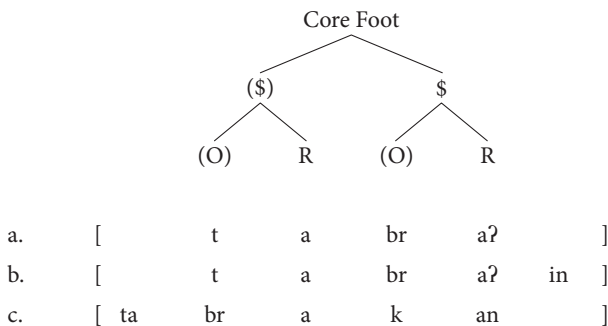(i) *fb yendi kenak belokir pertemanan nya vid selama 30 hari ni*
     Facebook Yendi **kenak** *blokir pertemannya   Vid   selama*
     Facebook Yendy undergo block CAUS:friend:ASSOC HYP\David one:long.time
     tiga   puluh hari ni
     three ten   day   DEM.PROX
     [Speaker named Yendy explaining why he hasn't been in touch]
     'My Facebook friending has been blocked, David, for the last thirty days.'

Novel orthographic representations such as these thus provide additional support for the claim that speakers of Riau Indonesian associate final [ʔ] with the *k* phoneme.

44. One obvious place to look is in the coda of the first syllable of forms exhibiting a CVCCVC template; if forms such as CVkCVC were consistently realised as CV[ʔ]CVC then this would provide conclusive evidence for the claim that it is the syllable that is relevant, not the Core Foot or a larger Word unit. However, according to Tadmor (2004), there are no native forms of either

Nevertheless, in spite of this indeterminacy, the [k] → [ʔ] rule still provides evidence for the existence of a Core Foot in Riau Indonesian. Such evidence derives from two vowel-initial Bound Light morphemes that may attach to the end of a host form with final [ʔ], Loosely Bound *-in* and Strongly Bound *-an*. As expected, the two Bound morphemes behave differently in the above-mentioned phonological context. When *-in* is added to a form ending in [ʔ], the [ʔ] is retained, for example [tabraʔ] 'collide' ~ [tabraʔin] ' collide with', [masaʔ] 'cook' ~ [masaʔin] 'cook for', and many others. In contrast, when *-an* is added, the [k] resurfaces, for example, [tabraʔ] 'collide' ~ [tabrakan] 'collide with each other', [masaʔ] 'cook' ~ [masakan] 'cuisine', and a handful of others.[45]

An analysis of these facts, illustrated with the [tabraʔ] ~ [tabraʔin] ~ [tabrakan] alternation, is presented in Figure 12 below:

```
                        Core Foot
                    ┌───────────┴───────────┐
                   ($)                       $
                ┌───┴───┐               ┌────┴────┐
               (O)      R              (O)        R

     a.    [         t      a        br      aʔ            ]
     b.    [         t      a        br      aʔ    in      ]
     c.    [  ta     br     a        k       an            ]
```

Figure 12. Final [k] → [ʔ]

In (a), the monomorphemic *tabrak* occupies its usual position in the Core Foot. In (b), too, *tabrak* remains in place, while *-in* is added outside the Core Foot in a structurally higher position. However, in (c), *-an* occupies the final position in the Core Foot pushing *tabrak* back to earlier positions in the structure.

CV[k]CVC or CV[ʔ]CVC structure. Loan forms do exist, though, instantiating both possibilities, for example *waktu* [waktu] 'time' < Arabic *waqtu*; *takbir* [takbir] 'profession of Islamic faith via the pronunciation of *Allāhu akbar*' < Arabic *takbīr;* but *rakyat* [raʔyat] 'public' < Arabic *raʕīyat*; *bakso* [baʔso] 'meatball soup'< Hokkien *bah-so·*. However, as noted above, there is similar variability also in the final slot of the CVCVC template; the phonology of loan forms differs from that of native forms, and is not relevant to the analysis of the [k] → [ʔ] rule.

45. Alternations such as [tabraʔ] ~ [tabrakan] are of relatively limited geographical distribution across Malay/Indonesian dialects. For example, in Jakarta Indonesian the corresponding forms are [tabrak] and [tabrakan], while in Sumatra Barat Indonesian they are [tabraʔ] ~ [tabraʔan] – the latter providing a relatively rare example of a contrast between Riau Indonesian and the variety of Indonesian spoken in the neighbouring and ethnically not too dissimilar province of Sumatra Barat.

The above analysis accounts straightforwardly for the alternative realisations of *k*, whatever the conditioning domain is taken to be: syllable, Core Foot, or a larger Word unit. Whereas in (a) and (b), *k* occurs in the coda of its syllable, the end of the Core Foot and also the end of the P-Word, and therefore changes to [ʔ], in (c) *k* occurs in the onset of its syllable and in a non-final position, and accordingly retains its realisation as [k].

As shown above, in order to account for the distinct behaviour of Weakly Bound Light morpheme *-in* and Strongly Bound Light morpheme *-an*, it is necessary to posit distinct structural positions, the former outside of the Core Foot, the latter within it. Thus, the [k] → [ʔ] phonological rule provides evidence for the existence of the Core Foot.

So far, in Sections 3.3.1–8, eight distinct sources of evidence were shown to converge in support of a pervasive disyllabic structure. The first made reference to the canonical shape of morphemes, the second gave the two privileged syllables an extra oomph, while five others involved processes that conspired towards supporting disyllabicity – the first by whittling down forms that are too big, the remaining four by beefing up forms that are too small. And the final source of evidence, in this subsection, made reference to distinct structural positions defined in terms of the Core Foot. These eight sources of evidence thus provide independent support for the reality of the Core Foot in Riau Indonesian. Two additional sources of evidence for the Core Foot are presented in Sections 3.3.12–13 below.

As suggested in Figures 6–12, in very many cases, the Core Foot is actually coextensive with the Word. However, as is equally evident in the same figures, from a structural point of view, the Core Foot is just part of a larger Word. We now turn to consider the sources of evidence in favour of these larger units, beginning with one source of evidence for the G-Word in Section 3.3.9, followed by a contrasting source of evidence for the P-Word in Section 3.3.10.

3.3.9   *Reduplication*

Reduplication occurs frequently in Riau Indonesian and is associated with a wide range of functions, some clearly iconic, such as plurality, distributivity, intensification and reciprocity, others less obviously so, for example atelicity, depreciation, concessivity and negative polarity. Formally, Reduplication most commonly applies to a single Footed morpheme along with some associated Bound Light morphemes, and results in a construction consisting of two copies of the reduplicated material, though other less common variants may also occur.[46]

---

46. Less frequently occurring variants of the basic Reduplication construction may differ in the following two ways. First, the number of copies may be three or more. Secondly, the reduplicated

The basic and most common variant of Reduplication in Riau Indonesian follows the general rule in (11) below:

(11) Reduplication
General Rule
Reduplication combines two G-Words to yield a superordinate G-Word

Thus, in accordance with the General Rule in (11) above, Reduplication is a source of evidence for the existence of G-Words in Riau Indonesian.

Example (12) below illustrates the simplest and most common form of Reduplication, involving a disyllabic morpheme with no additional Bound Light morphemes, *mandi mandi*. (It also happens to contain an instance of multiple partial reduplication, *mumumuak*.)

(12) *Ala Rip,   tak mumumuak   engkau **mandi mandi?***
EXCL HYP\Arip NEG DISTR~nauseous 2SG   DISTR~ wash
[Complaining to friend who washes too much]
'Gosh, Arip, aren't you sick of washing all the time?'

Examples (13)–(24) below show the different ways in which Reduplication may interact with Bound Light morphemes: whereas Strongly Bound Light morphemes undergo Reduplication together with their hosts, Weakly Bound Light morphemes remain outside the scope of Reduplication. In Examples (13)–(16), Reduplication applies only to the host Footed morpheme, not to the preceding Weakly Bound Light morphemes, *N-, di-* (twice), *ber-* and *ter-*.

(13) *Ayo,   berangkat, jangan **melambat lambat***
EXHRT NON.PAT:lift NEG.IMP AG:DISTR~ slow
[Getting ready to leave]
'Come on, let's leave, don't be slow'

(14) *Bodoh, **disimpan simpan,** tak mau **dibagi   bagi***
stupid PAT:DISTR~ put.away NEG want PAT:DISTR~ share
[At night market, complaining about friend who was pouring Milo, bit by bit, into bottle, instead of giving it to the rest of his friends to share]
'Stupid, he's putting it away, he won't share any of it'

---

material may be smaller, consisting of just the initial syllable of the reduplicated material, or alternatively, it may comprise a larger string of two or more Footed morphemes along with some associated Bound Light morphemes. For the most part, this subsection focuses on the most common variant, consisting of two copies of a single Footed morpheme plus, optionally, one or more Bound Light morphemes.

(15) *Om,   tadi   om   e, budak ini   e, Yendy,*
      HYP\uncle PST.PROX HYP\uncle Q child DEM:DEM.PROX Q Yendy
      berak   om   e, tak disiram   om,   taiknyo   om,
      defecate HYP\uncle Q NEG PAT:spray HYP\uncle shit-ASSOC HYP\uncle
      berlepet   lepet   om
      NON.PAT:DISTR~ spread.out HYP\uncle
      [Complaining to his uncle about friend who didn't flush the toilet]
      'Uncle, just before this kid, Yendy, shat and didn't flush, his shit was all over
      the place'

(16) *Ini   dia, yang **terbesar   besar** itu*
      DEM:DEM.PROX 3   REL NON.AG:DISTR~ big   DEM:DEM.DIST
      [Passing a car display]
      'There they are, the biggest ones'

In contrast, in Examples (17)–(20), Reduplication applies to the host together with
a preceding Strongly Bound Light morpheme, *N-, ber-, se-,* and *ke-* respectively:

(17) *Main yang **nembak nembak** itu   Pit   ha,*
      play REL DISTR~ AG:shoot DEM:DEM.DIST HYP\Dapit DEIC
      enak Pit
      nice HYP\Dapit
      [About a game we had played a short while before]
      'Let's play that shooting game David, it's fun David'

(18) *Remi   kita **bertaruh bertaruh***
      rummy 1.2 DISTR~   NON.PAT:bet
      [Suggesting a card game]
      'Let's play rummy with bets'

(19) *Dia minta **sebelah sebelah***
      3   request DISTR~ one:cleave
      [Two shoeshine boys having just approached me, speaker explains]
      'They want to polish one shoe each'

(20) *Ngapain   mister   **keSingapore keSingapore** terus*
      AG:what:EP white.person DISTR~   ALL:Singapore continue
      [After I say that I'll be going to Singapore in a few days time]
      'Why do you keep on going to Singapore?'

Similarly, in (21)–(23), Reduplication applies only to the host Footed morpheme,
not to the following Weakly Bound Light morphemes, *-an, -kan* and *-in:*

(21) *Dia **cium   ciuman***
      3   DISTR~ kiss:AUG
      [About two doves]
      'They're kissing'

(22) *Saya **masuk masukkan** semua*
    1SG   DISTR~ enter:EP   all
    [Playing billiards, speaker brags]
    'I'll pocket them all.'

(23) *Singapore mister   **ngapa   ngapain** aja?*
    Singapore white.person AG:DISTR~ what:EP NEG.FOC
    [I say that I'm going to Singapore next week; speaker then asks]
    'What all will you be doing in Singapore (referring to me as *mister* 'white person')'

In contrast, in (24) below, Reduplication applies to the Footed morpheme together with a following Strongly Bound Light morpheme -*an*:

(24) *Tak **kasihan kasihan** Pit?*
    NEG DISTR~ pity:AUG HYP\Dapit
    [Pleading for sympathy]
    'Aren't you sorry for me, David?'

The structure of selected examples representing the variegated patterns in (15)–(24) is shown in Figure 13 below:
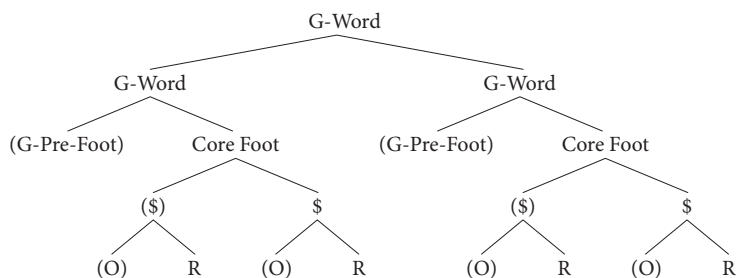
Figure 13. Reduplication

While example (a) shows the simple case of reduplication applying to a basic Free Footed morpheme, subsequent examples illustrate the variable interactions of Reduplication with Bound Light morphemes, accounting for them in terms of the distinction between Weakly and Strongly Bound morphemes and the different structural positions that they occupy. Specifically, whereas in (b), (c), and (f), a Weakly Bound Light morpheme occurs outside the G-Word, and hence remains outside the scope of reduplication, in (d), (e) and (g), a Strongly Bound Light morpheme occurs within the G-Word and hence undergoes reduplication together with its host Footed morpheme.[47]

So far, all the examples of Reduplication have involved a single Footed morpheme. However, in some cases, sequences of two or more Footed morphemes may undergo Reduplication. Most commonly, such examples involve numerals, as in (25)–(28) below:

(25) ***Enam puluh enam puluh** biar lah Pit*
  DISTR~   six   ten   let   FOC   HYP\Dapit
  [In hotel, I wonder where to do my laundry; speaker says I should have it done in the hotel; I say last time six pieces cost sixty thousand, which is too expensive; speaker responds]
  'Sixty or whatever, let it be, David'

(26) ***Dua telur dua telur***
  DISTR~   two   egg
  [In hotel room, interlocutor on phone to room service, ordering soft boiled eggs in cups; speaker wants him to ask for two eggs in each cup]
  'Two eggs each'

(27) ***Dua batang dua batang** semua*
  DISTR~   two   long.object   all
  [Playing cards, complaining about hand just dealt]
  'They're all in twos'

---

47. Note that, in accordance with the typology of Bound Light morphemes in Table 2, while some morphemes are unambivalently Weakly or Strongly Bound, others may belong to either of the two classes. Thus, whereas *ber-* is Weakly Bound in (c), it is Strongly bound in (e); similarly, whereas *-an* is Weakly Bound in (f), it is Strongly Bound in (g). While the existence of such ambivalent morphemes might appear to weaken the empirical force of the distinction between Weakly and Strongly Bound morphemes, closer inspection of the cases in point reveals that the conditioning factors governing the alternative behaviours of such morphemes are themselves largely systematic, providing yet additional support for the distinction. Thus, for example, while the Weakly Bound Light morpheme *-an* typically occurs in semi-productive contexts such as the reciprocal construction in (f), its Strongly Bound counterpart *-an* is generally found in constructions exhibiting a low degree of compositionality and correspondingly high degree of lexicalisation.

(28)   *Orang **dua**   **puluh kaleng dua puluh kaleng*** tak ada   mabuk
       person DISTR~   two ten   can   NEG exist drunk
       [Teasing me for getting tipsy over half a can of beer]
       'People drink twenty cans each and don't get drunk.'

Another class of longer forms that may undergo Reduplication, but for which I currently have no examples in the naturalistic corpus, is that of Compounds – constructions consisting of two Footed morphemes, whose semantics is partly or completely non-compositional. One common class of Compounds is that of toponyms, such as Pekan Baru; another common class involves constructions with a class term, such as *ikan* 'fish' in *ikan hiu* 'fish shark'. Such compounds may also undergo reduplication, yielding forms such as *Pekan Baru Pekan Baru* and *ikan hiu ikan hiu*.

The structure of Reduplications of such longer sequences of two (or more) Footed morphemes is shown in Figure 14 below:
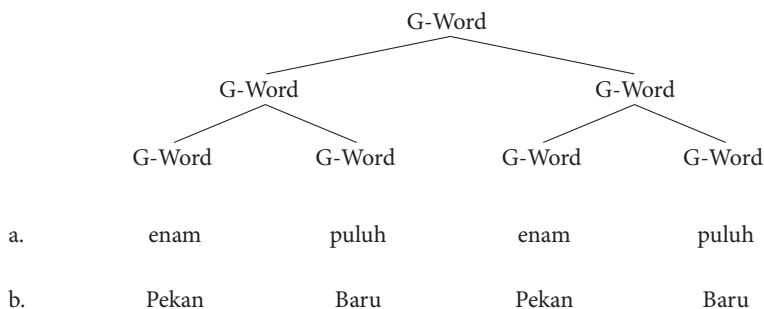

Figure 14. Reduplication (Compounds)

As suggested in Figure 14, in cases such as these, Reduplication applies to a non-terminal G-Word. Examples such as these show the potentially recursive nature of the G-Word, though at present I am not familiar with any evidence in favour of more than the above two levels of embedding.

Thus, Reduplication, via the General Rule in (11), provides evidence for the existence of G-Words in Riau Indonesian. At the same time, it does not allow for an alternative analysis making reference to P-Words; this is because P-Word structure does not offer distinct structural positions making it possible to account for the different behaviour of Weakly and Strongly Bound Light morphemes with respect to Reduplication.[48]

---

48. Specifically, while the present account distinguishes between Weakly Bound Light morphemes occurring outside the G-Word and Strongly Bound Light morphemes occurring within it, an analogous account in terms of the P-Word would not be possible because, as shown below, both Weakly and Strongly Bound Light morphemes occur within the P-word.

Reduplication is in fact the only hard-core grammatical process providing evidence for Word structure of any kind in Riau Indonesian. Nevertheless, Reduplication is complemented by a variety of phenomena of an extra-grammatical nature involving poetic meter, ludlings, and natural orthography, each of which provides additional evidence for some kind of Word structure in Riau Indonesian. The remainder of Section 3.3 considers five such extra-grammatical sources of evidence, beginning with one that provides evidence for P-Word structure, namely Pantun prosody.

3.3.10   *Pantun prosody*

The Pantun is a ubiquitous feature of Malay and Indonesian culture, a poetic verse form spanning the range from high to low, from written to oral, and from canonical to non-canonical – it is both the Shakespeare and the nursery rhymes of Malay/ Indonesian. As is often the case in studies of metered verse, it is the "low" forms – unencumbered by the conscious manipulations of an expert and learned poet – that offer the most direct, straightforward and unmediated window into our tacit prosodic and grammatical knowledge. It is, therefore, such naturally-occurring forms of the Pantun that we shall consider here.

The rules of versification of the Pantun make clear and extensive reference to the P-Word. Most notably, the P-Word features in the following general rule of well-formedness:

(29)   Pantun:
       General Rule
       The Pantun consists of 16 P-Words

It is no coincidence that 16 equals 2^4. The prosodic structure of the Pantun is binary all the way down, consisting of one quatrain, 2 couplets, 4 lines, 8 hemistiches, and 16 feet – the latter each containing exactly one P-Word.[49] This binary structure provides the scaffolding for numerous optional embellishments to the Pantun structure, involving features such as rhyme, sound patterns, repetition, syntactic and semantic parallelism, and others – see Gil (1993) for details.[50] In particular, an important organisational feature of the Pantun is that its "main" content occurs in the second couplet, while the first couplet consists of formally echoing material ranging from randomly nonsensical to subtly metaphorical.

---

49. Note that the above-mentioned 16 feet are *prosodic* feet, of relevance to the prosodic organisation of the Pantun. They are thus distinct from the notion of Core Foot discussed at length in this chapter.

50. Stanzaic structures consisting of four lines, each containing four prosodic feet, are ubiquitous in canonical and non-canonical verse worldwide, ranging from limericks and nursery rhymes all the way to the medieval Arabic *muwaššah* and English ballad metre – see Gil (1990, 1991) for discussion.

Examples (30)–(33) are from a corpus of Pantun, recorded during several sessions in which speakers engaged in friendly competition, showing off their skills in the impromptu improvisation of Pantun on a variety of topics of immediate relevance – often mentioning the names of people present (including myself). Examples (30)–(33) are formatted so as to highlight the prosodic structure, consisting of 4 lines, each containing 4 feet, each of which contains a single P-Word:[51]

(30) | *Kalau* | *anda* | *ingin* | *melayang* |
| --- | --- | --- | --- |
| TOP | 2 | desire | AG:kite |
| *Jangan* | *lupa* | *membawa* | *panci* |
| NEG.IMP | forget | AG:carry | pot |
| *David* | *datang* | *ingin* | *belajar* |
| David | arrive | desire | NON.PAT:learn |
| *belajar* | *bahasa* | *Sungai* | *Selari* |
| NON.PAT:learn | language | Sungai | Selari |

[Sungai Selari is the name of the village where the speaker and many of his friends come from]
'If you want to fly a kite
Don't forget to bring a pot
David has come and wants to learn
To learn the language of Sungai Selari'

(31) | *Jalan* | *jalan* | *keKuala* | *Deli* |
| --- | --- | --- | --- |
| DISTR~ | go | ALL:Kuala | Deli |
| *[a] Lewat* | *jalan* | *berliku* | *liku* |
| pass | road | NON.PAT:DISTR~ | curve |
| *Kalau* | *anda* | *temannya* | *Elly* |
| TOP | 2 | friend:ASSOC | Elly |
| *Singga* | *dulu* | *diWisma* | *Ratu* |
| stop.over | first | LOC:Wisma | Ratu |

[Kuala Deli is a distant town; Elly is a friend of the speaker, Wisma Ratu is the hotel where Elly used to work]
'Going to Kuala Deli
Along a road full of curves
If you want to be friends with Elly
Stop off first at Wisma Ratu'

---

51. Examples (30)–(33) were recorded in Sungai Pakning. In (30)–(33), material judged to be extraneous to the structure of the structure of the Pantun is represented in square brackets: this consists of [a], representing a filler vowel, and [x3], denoting that the word preceding it was repeated three times, reflecting either hesitation or a desire to create a dramatic effect. An additional example of a Pantun, sent by SMS, can be found in (63) below.

(32) | *Jalan* | *jalan* | *kerumah* | *Rudy* |
|---|---|---|---|
| DISTR~ | go | ALL:house | Rudy |
| *Singga* | *sebentar* | *katSungai* | *Apit* |
| stop.over | one:while | near:Sungai | Apit |
| *David* | *datang* | *melenggang* | *lenggok* |
| David | arrive | AG:DISTR~ | sway |
| *Nampak* | *Susi* | *langsung [x3]* | *terdiam* |
| see | Susi | straight | NON.AG:quiet |

[Rudy is a friend of the speaker; Sungai Apit is the nearby village where Rudy lives; Susi is the name of a girl]
'Going to Rudy's house
Stopping over at Sungai Apit
David arrives all topsy turvy
Seeing Susi immediately falling quiet'

(33) | *Jalan* | *jalan* | *kePekan* | *Baru* |
|---|---|---|---|
| DISTR~ | go | ALL:Pekan | Baru |
| *Jangan* | *lupa* | *membeli* | *rambutan* |
| NEG.IMP | forget | AG:buy | hair:AUG |
| *Sungguh* | *enak* | *pengantin* | *baru* |
| true | nice | AG:bride(groom) | new |
| *Dalam* | *kamar* | *cubit* | *cubitan* |
| inside | room | DISTR~ | pinch:AUG |

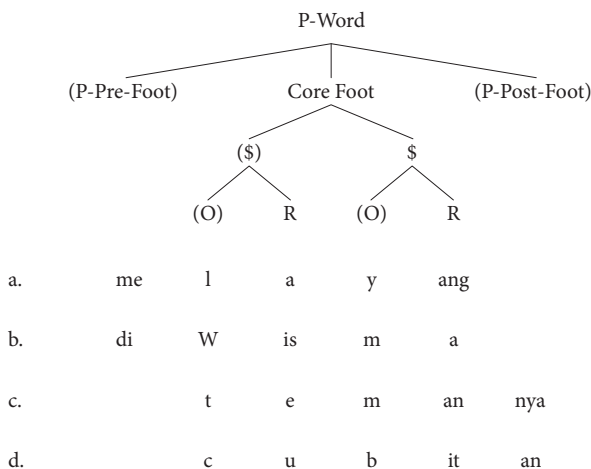[Pekan Baru is the provincial capital]
'Going to Pekan Baru
Don't forget to buy rambutan
How nice are the newlyweds
Pinching each other in the room'

The above examples vividly illustrate the centrality of the P-Word to the prosodic organisation of the Pantun. Indeed, the rather doggerel feel of these verse forms is due precisely to the need to conform to the General Rule in (29). Thus, for example, in (30), the speaker could do no better than to simply repeat the form *belajar*, at the end of the third line and then at the beginning of the fourth.

More often than not, each prosodic foot contains exactly one morpheme; this is the case in 46 out of the 64 prosodic feet in the above four examples. However, it is the remaining cases, involving Bound Light morphemes, that provide evidence in support of P-Word as the relevant structural unit for the General Rule in (29). Thus, *me(n)-, be(r)-, ter-, ke-, kat-, di-, peN-* and *se-* all share a prosodic foot with the following morpheme while *-nya* and *-an* share a prosodic foot with the preceding morpheme. Accordingly, these Bound Light morphemes form a single P-Word structural unit with their adjacent morphemes, as indicated for selected examples in Figure 15 below:

P-Word

(P-Pre-Foot)   Core Foot   (P-Post-Foot)

($) $

(O)   R   (O)   R

a.	me	l	a	y	ang	
b.	di	W	is	m	a	
c.		t	e	m	an	nya
d.		c	u	b	it	an

Figure 15. Pantun prosody

Compounds and Reduplications show why the General Rule in (29) could not have admitted an alternative formulation making reference to G-Words. As suggested in (30)–(33), Compounds such as *Sungai Selari, Kuala Deli*, *Sungai Apit*, *Pekan Baru* and *Wisma Ratu*, and Reduplications such as *jalan jalan* and *melenggang lengok*, constitute two distinct P-Words with respect to Pantun prosody. In particular, as shown in Figure 15 (b), in the expression *di Wisma Ratu*, *di Wisma* forms a P-Word constituent to the exclusion of *Ratu*, thereby cross-cutting the G-Word constituent *Wisma Ratu*; similarly, as shown in Figure 15 (d), in *cubit cubitan*, *cubitan* forms a P-Word constituent to the exclusion of the first *cubit*, cross-cutting the G-Word constituent *cubit cubit*. Thus, it is the P-Word rather than the G-Word that is relevant to Pantun prosody.

    In summary, then, each of the 16 prosodic feet of the Pantun contains a single P-Word, consisting of an obligatory disyllabic morpheme plus one or more optional Bound Light morphemes.[52] In doing so, Pantun prosody provides a simple and straightforward source of evidence for the existence of P-Words.

    Having presented evidence in support of distinct G-Words and P-Words in the last two subsections, we now consider some additional sources of evidence for these two kinds of Word structure. Much of this evidence is in the form of three very different ludlings, revealing contrasting kinds of Word structure, discussed in the next three subsections.

---

52. A prima facie similar system involving the association of apparent phonological words with prosodic feet is argued to be one of the principles underlying much Biblical Hebrew poetry; see Shoshany (1986) and references therein.

3.3.11   *Sabaha ludling*

The value of ludlings, also known as secret languages, language disguises, language games, or backwards languages, as a window into linguistic structure has long been recognised, perhaps most notably in the pioneering work of Bagemihl (1988, 1989), though others, such as Nevins and Endress (2007), have questioned the utility of ludlings in providing evidence for such structures. For other varieties of Malay/Indonesian, ludlings have been described by Azizul Rahman (1995), Chambert-Loir (1998) and others. The next three subsections examine three different ludlings in Riau Indonesian, named in accordance with the way in which they render the Indonesian word *bahasa* 'language'.

The first ludling to be considered is the Sabaha ludling, which provides evidence for the existence of G-Words, in accordance with the general rule given below:[53]

(34)   Sabaha Ludling:
General Rule
The last syllable of the terminal G-Word is moved to the front of the terminal G-Word

An example of how this works is evident in the name of the ludling: *bahasa →
sabaha*. Some examples of naturalistic sentences employing the Sabaha ludling are given in (35)–(40) below:[54]

(35)   *Lakga   lakga mika*
Galak   galak kami
DISTR~ laugh 1
[From a long narrative]
'We all laughed'

---

53. The Sabaha ludling would seem to represent one of the most common types of ludling cross-linguistically. Some ludlings operating on apparently similar principles to Riau Indonesian Sabaha include French *Verlan* (Lefkowitz 1991), Serbo-Croatian *Šatrovački* (Rizzolo 2007), and Oromo *Bird Talk* (Hordofa and Unseth 1986).
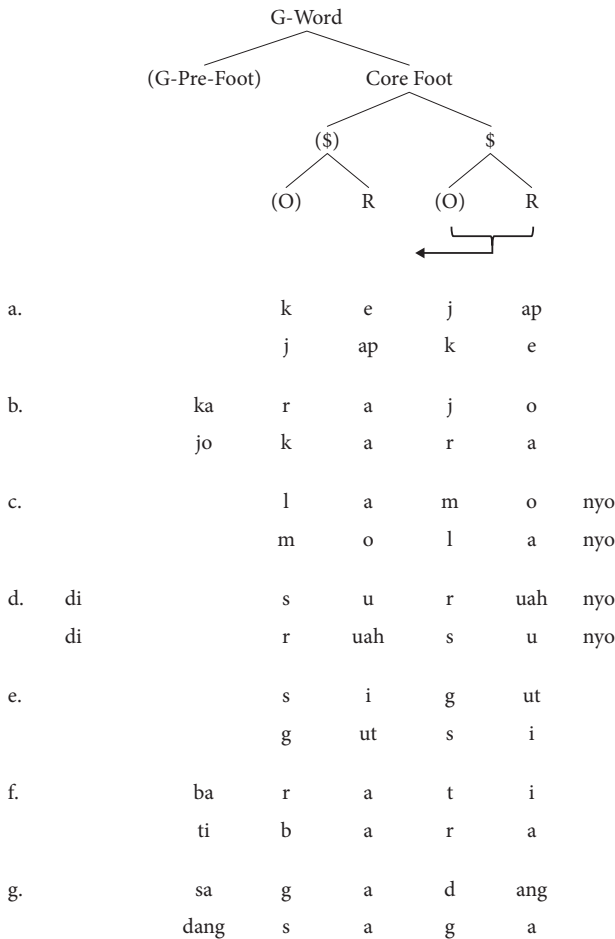
54. Examples (35)–(40) were recorded in Pekan Baru. In these examples, as well as subsequent ones involving ludlings, the first line presents the actual utterance, while the second one shows the ludling "undone", that is to say, the source utterance to which the ludling applies. Subsequent lines provide the usual interlinear glosses, context, and free translation into English. The Sabaha ludling data involve code-switching between Riau Indonesian and closely related Minangkabau, which is represented in the interlinear glosses as follows: no underline – unambiguously Riau Indonesian; single underline – both Riau Indonesian and Minangkabau; double underline – unambiguously Minangkabau.

(36) *Doa, dangga kani hanlounyo*
Ado gadang ikan louhannyo
<u>exist</u> <u>big</u>   fish <u>louhan</u>:ASSOC
[Arguing whether a certain place has a kind of fish called "louhan"]
'Yes there are, they're big, their louhan fish'

(37) *Diroahsunyo   likba   japke*
Disuruahnyo   balik   kejap
PAT:<u>order</u>:ASSOC return wink
[Speaker worried he should be back at work]
'He said we should return for a bit'

(38) *Wekce angwak Gutsi   kan*
Cewek wa'ang siGut   kan
girl   <u>2SGM</u>   HYP\<u>XXX</u>[55] Q
[Teasing friend]
'Your girlfriend is Gut'

(39) *Molanyo   jokara Ja- Jakarta Nda'i?*
Lamonyo   karajo   Jakarta Indak?
<u>long.time</u>:ASSOC <u>work</u>   Jakarta NEG
[Asking me about my travel plans]
'Are you going to be working in Jakarta for a long time?'

(40) *Tibara   yang dangsaga tui   yang wa'a hekli tui*
Barati   yang sagadang itu   yang awak lihek itu
NON.PAT:<u>meaning</u> REL <u>one</u>:big DEM:DEM.DIST REL <u>1.2</u>   see DEM:DEM.DIST
[Discussing different kinds of fish]
'So the ones that are as big as that, the ones that we saw'

The Sabaha ludling thus provides direct evidence for the terminal G-Word. Some sample analyses of particular forms in (35)–(40) above are given in Figures 16 and 17 below. In each example, the original form is indicated first, following by the attested ludling form directly beneath it.

---

55. Unfortunately, I was not able to ascertain the full name for which *Gut* is the truncated hypocoristic form.

Figure 16. Sabaha ludling

	(G-Pre-Foot)	(O)	R	(O)	R	
a.		k	e	j	ap	
		j	ap	k	e	
b.	ka	r	a	j	o	
	jo	k	a	r	a	
c.		l	a	m	o	nyo
		m	o	l	a	nyo
d.	di	s	u	r	uah	nyo
	di	r	uah	s	u	nyo
e.		s	i	g	ut	
		g	ut	s	i	
f.	ba	r	a	t	i	
	ti	b	a	r	a	
g.	sa	g	a	d	ang	
	dang	s	a	g	a	

In Figure 16, (a) and (b) show the ludling applying to simple Free Footed morphemes of disyllabic and trisyllabic structure respectively, the latter case showing that the final syllable may move forward to a position that is outside the Core Foot. The remaining examples in Figure 16 show the application of the ludling to Footed morphemes in construction with a Bound Light morpheme. In general, the Sabaha ludling makes reference to the same distinction as does Reduplication, namely that between Weakly and Strongly Bound Light morphemes, the generalisation being that the ludling applies to the latter but not to the former. In examples (c) and (d), Weakly Bound Light morphemes remain outside the scope of the ludling. In both examples, the associative marker -*nyo* remains in final position, while the syllable that is fronted is the one preceding it. In addition, in (d), the patient-orientation
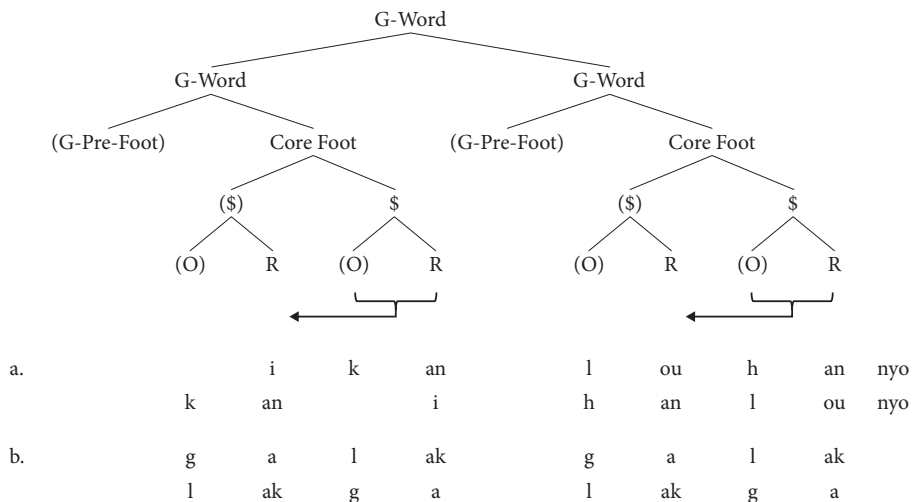
G-Word

G-Word · G-Word

(G-Pre-Foot) · Core Foot · (G-Pre-Foot) · Core Foot

($) · $ · ($) · $

(O) · R · (O) · R · (O) · R · (O) · R

	(O)	R	(O)	R		(O)	R	(O)	R	
a.		i	k	an		l	ou	h	an	nyo
	k	an		i		h	an	l	ou	nyo
b.		g	a	l	ak		g	a	l	ak
	l	ak	g	a		l	ak	g	a

Figure 17. Sabaha ludling

marker *di-* remains in front, while the syllable fronted by the ludling occurs right after it. These examples show that the ludling applies to the G-Word, whereas *-nyo* and *di-* are located outside it. In contrast, in examples (e), (f) and (g), the final syllable is moved in front of a Strongly Bound Light morpheme, personal marker *si-* (discussed in Section 3.3.6 above), non-patient-orientation marker *bar-*, and *sa-* 'one'.[56] In conjunction, then, cases (c)–(g) show that the ludling applies to the G-Word, including whatever Strongly Bound Light morphemes it may contain, while not applying to any Weakly Bound Light morphemes lying outside the G-Word.

In Figure 17, the two examples show the ludling applying to a complex G-Word consisting of two terminal G-Words. In (a), *ikan louhan* is a Compound in which the first term is a class term, while in (b), the form *galak* undergoes Reduplication. In both cases, the Sabaha ludling applies to each terminal G-Word separately, showing that the relevant domain for the Sabaha ludling is the terminal G-Word.

---

56. Although the forms *bar-*, *sa-* and *-nyo* are Minangkabau, it should be kept in mind that many of the sound correspondences between Riau Indonesian and Minangkabau are straightforwardly transparent, to the extent that even uneducated and illiterate persons are aware of them, and able to make use of them when code-switching between the two language varieties. For this reason, when dealing with a text that involves code-switching between the two varieties, it is not unreasonable to make inferences from the behaviour of Bound Morphemes such as *bar-*, *sa-*, and *-nyo*, pronounced the Minangkabau way, with regard to that of their respective cognates in Riau Indonesian, *ber-*, *se-*, and *-nya*.

The above examples highlight the conflicting constituencies associated with P-Words and G-Words respectively. Consider a string of the form *X Y -nya*, where *X Y* is a Compound or Reduplication structure. As evidenced by *ikan louhannya* in Figure 17 (a) above, the Sabaha ludling supports an *[X Y] -nya* constituency. Earlier, however, we saw that Pantun prosody treats semantically closely-knit units such as *Sungai Selari* as separate P-Words, while considering the associative marker *-nya* to be part of the same P-Word as its preceding host; thus, Pantun prosody supports a conflicting *X [Y -nya]* constituency. In conjunction, then, these two sources of evidence underscore the difference between the two distinct types of Word structure in Riau Indonesian.

In summary, as evidenced by the above examples, and many others, the Sabaha ludling provides strong support for the existence of a G-Word in Riau Indonesian, thereby joining forces with Reduplication, previously, and naturalistic orthography in Section 3.3.14 below, in reflecting the G-Word level of structure.

3.3.12   *Warasa ludling*

The second ludling to be considered is the Warasa ludling. Like the sources of evidence considered in Sections 3.3.1–8 previously, the Warasa ludling provides evidence for Core Foot structure. In addition, it also provides evidence for Word structure. However, rather than G-Words, as is the case for the Sabaha ludling considered above, the Warasa ludling instead supports the existence of P-Words.

The Warasa ludling operates in accordance with the following rule:

(41)   Warasa Ludling
    General Rule
    Replace the onset of the first syllable of the Core Foot and any material preceding it in the P-Word with the fixed sequence *war*.

Some examples of naturalistic utterances illustrating the Warasa ludling are given in (42)–(49) below:[57]

---

57. Examples (42)–(49) below, were recorded in Sungai Pakning, where the Warasa ludling was used mostly by teenage boys. The Warasa ludling data involve code-switching between Riau Indonesian and closely related Siak Malay, which is represented in the interlinear glosses as follows: no underline – unambiguously Riau Indonesian; single underline – both Riau Indonesian and Siak Malay; double underline – unambiguously Siak Malay. A large majority of the forms are marked as belonging to both Riau Indonesian and Siak Malay (though a closer phonetic analysis might make it possible to distinguish between the two in at least some cases).

(42) *Warami waredap Warelat Wararu waruduk warano waredap warisap a*
      Kami   sedap   Selat   Baru   duduk   sano   sedap   hisap   a
      1   nice   Selat   Baru   sit   there   nice   suck   uh
      warinum warinum warakan warakan e, warapit e
      minum   minum   makan   makan   e Dapit   e
      DISTR~   drink   DISTR~   eat   Q David   Q
      [Discussing planned beach trip to Selat Baru]
      'I like it in Selat Baru, sitting around there, it's nice, smoking, drinking, eating,
      right David'

(43) *W- warabtu warajo,   waraku warisa   waringgu warami waruo*
          Sabtu   ajo,   aku   takbisa minggu   kami   duo
          Saturday NEG.FOC   1SG   NEG:can Sunday   1   two
      warowok waruka
      Bowok   pramuka
      Bowok   scouts
      [Discussing planned beach trip, arguing whether to go on Saturday or Sunday]
      'Make it Saturday, I can't go on Sunday, me and Bowo have scouts'

(44) *Waraku, wara- waraku warendak waridur do,   waraku wara- warari*
      Aku,   aku   tendak   tidur   do,   aku   hari
      1SG   1SG   NEG.want sleep   NEG.POL   1SG   day
      warejam warailan warain warai- walai- warelai wareson
      jam   sembilan main   Pelai   Steson
      hour   nine   play   Play   Station
      [Discussing planned beach trip]
      'I don't want to sleep, I want to play Play Station at nine o'clock'

(45) *Warapit warajuk*
      Dapit   merajuk
      David   AG:sulk
      [Discussing planned beach trip; my silence, intended to avoid having my voice
      be part of the recording, is misinterpreted]
      'David's sulking'

(46) *Waripai,   warengkau warapo waranyak warawat?*
      SiPai,   engkau   kenapo banyak   jerawat?
      HYP\Paisal 2   why   lots   pimple
      [Teasing friend]
      'Paisal, why do you have so many pimples?'

(47)   *Wariam, warowokni   warau waridur*
       Diam   Bowokni   mau   tidur
       <u>quiet</u>   Bowok:dem.prox <u>want</u>   sleep
       [Berating his friends for making too much noise]
       'Quiet, Bowok wants to sleep'

(48)   *Warapit e, warengkau warangunkan waraku lah*
       Dapit   e engkau   bangunkan   aku   lah
       <u>David</u>   q 2   get.up:ep   1sg   foc
       [Discussing planned beach trip]
       'David, you wake me up'

(49)   *Dadanyak dadetul dadulunyo   e*[58]
       Banyak   betul   bulunyo   e
       <u>lots</u>   <u>really</u>   feather:assoc q
       [Talking about me]
       'He's got lots of body hair, hasn't he?'

The structure of selected forms in Examples (42)–(49) is represented in Figure 18 below.

While examples (a)–(d) consist of a monomorphemic form, the remaining examples show a disyllabic morpheme in construction with a Bound Light morpheme – preceding the disyllabic morpheme in (e)–(g), following it in (h)–(j). Examples (a) and (b) show the most common application of the Warasa ludling to a single disyllabic morpheme, with *war* replacing the first consonant of the morpheme. Based on the large number of examples following this pattern, learners of the Warasa ludling are sometimes led to suspect that the Warasa rule simply involves adding *war* in front of the relevant form, replacing the first consonant if it begins with a consonant; however, such an impression is belied by the next two examples. Example (c) shows that for monosyllabic morphemes of the Footed variety, *war* occurs before the first consonant of the morpheme, with an epenthetic vowel inserted between the *r* and the first consonant, while example (d) shows that for trisyllabic morphemes, *war* replaces the entire antepenultimate syllable and the first consonant of the penultimate syllable. As suggested in Figure 18, the position of *war* is structurally determined, replacing all material up to and including the first

---

58. Example (49) represents a variant of the Warasa ludling in which the fixed sequence *dad-* is used instead of *war-* . It dates from a brief period during which the ludling speakers were playing around with various alternative fixed sequences – all of which, however, followed the common template provided by the General Rule in (41).
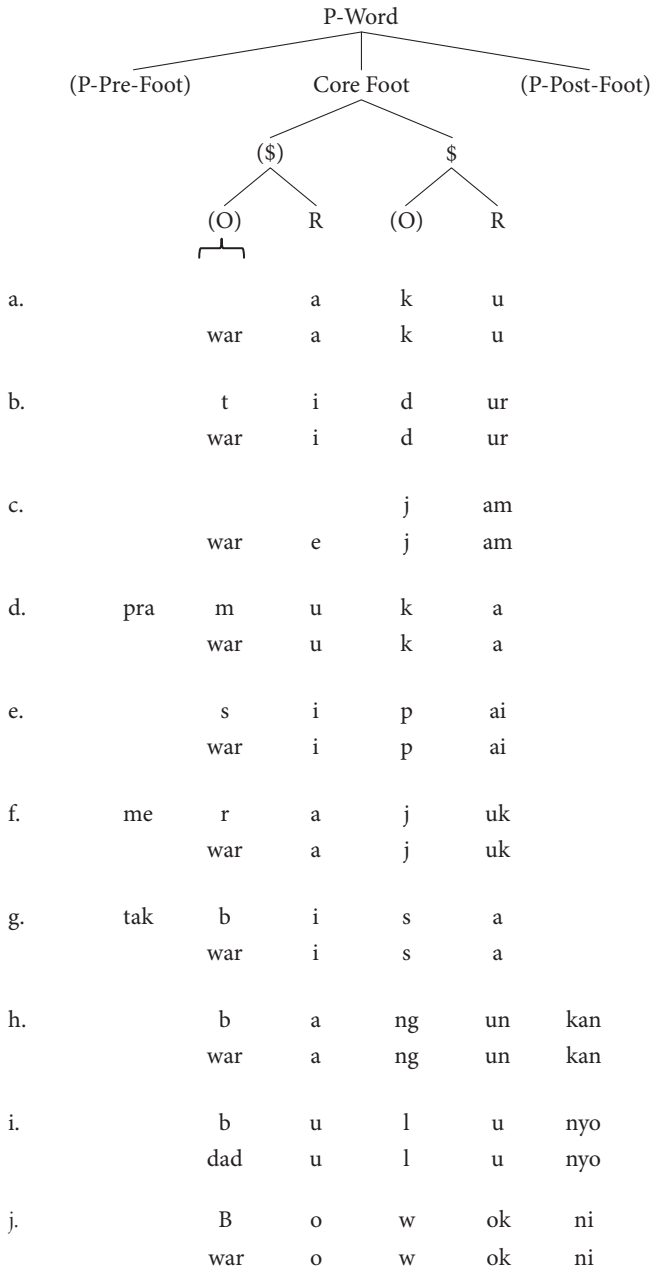
	P-Word			
(P-Pre-Foot)	Core Foot			(P-Post-Foot)
	($)		$	
	(O)	R	(O)	R
a.		a	k	u
	war	a	k	u
b.	t	i	d	ur
	war	i	d	ur
c.			j	am
	war	e	j	am

a.
| | | a | k | u |
| war | | a | k | u |

P-Word

(P-Pre-Foot) Core Foot (P-Post-Foot)

($) $

(O) R (O) R

a.
  a   k   u
war   a   k   u

b.
  t   i   d   ur
war   i   d   ur

c.
      j   am
war   e   j   am

d.   pra   m   u   k   a
     war   u   k   a

e.   s   i   p   ai
     war   i   p   ai

f.   me   r   a   j   uk
     war   a   j   uk

g.   tak   b   i   s   a
     war   i   s   a

h.   b   a   ng   un   kan
     war   a   ng   un   kan

i.   b   u   l   u   nyo
     dad   u   l   u   nyo

j.   B   o   w   ok   ni
     war   o   w   ok   ni

Figure 18. Warasa ludling

onset of the Core Foot. Thus, the Warasa ludling provides strong further evidence for the existence of the Core Foot in Riau Indonesian.[59]

Examples (e)–(g) show three distinct cases involving a Bound Light morpheme preceding the host Footed morpheme. Example (e) involves the personal marker *si-*, which, when used with a truncated name, occurs in the erstwhile empty first syllable of the Core Foot, as argued in Section 3.3.6 above; in this case, *war* replaces the *s* in the onset of the first syllable of the Core Foot, while leaving the *i* in the syllable's rhyme. In most other cases, however, when a Bound Light morpheme precedes the host Footed morpheme, it is completely obliterated by *war*. One example of that is in (d), where *war* replaces the *me-* allomorph of the marker *N-*, and another is in (e), where *war* replaces the negative marker *tak*.[60] Finally, examples (h)–(j) show three similar instances involving a Bound Light morpheme following its host Footed morpheme; in these cases, *war* replaces the first consonant of the Footed morpheme, thereby showing that the following Bound Light morphemes are situated outside of the Core Foot. In summary, then, in all of the examples in Figure 13, the location of *war* is defined in terms of the Core Foot, thereby providing evidence for the positioning of various Bound Light morphemes in relation to the Core Foot.[61]

---

59. The placement of *war* in the Warasa ludling bears an interesting resemblance to that of the Bound Light morpheme *N-* discussed in Section 3.3.7. Like *N-*, for both disyllabic and monosyllabic forms, *war* applies to the onset of the first syllable of the Core Foot: right in front of the relevant morpheme if it is disyllabic and beginning with a vowel, as in (a); replacing the first consonant of the morpheme if it is disyllabic and beginning with a consonant, as in (b), and creating an extra syllable with an epenthetic vowel if the morpheme is monosyllabic, as in (c) – thereby supporting the existence of the Core Foot in similar ways. However, the placement of *war* differs from that of *N-* in trisyllabic morphemes: whereas *N-* replaces the first consonant of the morpheme even if it is in Pre Core position, e.g. *telepon* 'telephone' > *nelepon* 'make a phone call', *war* retains its position attached to the first onset of the Core Foot, replacing a longer stretch of material, as in (d).

60. Since most "grammatical" markers in Riau Indonesian are optional, it is often difficult to make the case that a purported Bound Light morpheme was ever present in the input to the ludling, before being obliterated by it. However, in the present examples, such a case can indeed be made. In (d), *rajuk* is one of those forms that rarely occurs without prenasalisation – see Gil (2002b) for detailed discussion of the lexical variability associated with the marker *N-* – and hence the input to *warajuk* may plausibly be reconstructed as *merajuk*. And in (e), the context of the conversation makes it clear beyond doubt that the intended meaning of *warisa* is the negative *takbisa,* not the affirmative *bisa*. (The neutralisation of affirmative/negative interpretations by the Warasa ludling may seem surprising, but it is a robust and frequently occurring feature of the ludling. Nevertheless, when the speaker wishes to make sure that a negative interpretation is understood, an alternative available strategy is to co-opt the negative marker *tidak* from Standard Indonesian, which, as a Free Footed morpheme, is rendered by the ludling as *waridak*.)

61. Note that the above examples provide two instances of Free Footed morphemes that are treated by the ludling as Bound Light, in accordance with the discussion in Section 3.2.2: the negative marker *tak-* in (43) and the demonstrative *-ni* in (47).

Moreover, as suggested in the formulation of the General Rule in (38), the Warasa ludling also makes reference to the P-Word. The crucial cases are those involving a Bound Light morpheme whose structural position is before the Core Foot – examples (f) and (g) in Figure 18. The fact that *war* does not attach independently to each of the two morphemes, producing forms such as *wareme warajuk and *waretak warisa, could potentially be accounted for by the observation that *war* requires a Core Foot, which *me-* and *tak-* lack. However, the prediction would then be that these Bound Light morphemes occur in front of *war*, which would result in forms such as *mewarajuk and *takwarisa. But these do not occur; instead, as shown above, the preceding Bound Light morphemes *me-* and *tak-* are simply obliterated. This then shows that these Bound Light morphemes form part of a larger structural unit to which the Warasa ludling applies as a whole – a Word.

Remaining to be determined is whether the Word in question is the P-Word or the G-word. Constructions involving Compounds and Reduplications provide the answer. As was the case for the Sabaha ludling in the preceding subsection, the Warasa ludling applies individually to each term of the construction. Examples of Compounds are *Warelat Wararu* and *Warelai Wareson*, while examples of Reduplications are *warinum warinum* and *warakan warakan*. Recall, however, that Compounds and Reduplications involve the combination of terminal G-Words into a single non-terminal G-Word, forming a constituent to the exclusion of adjacent Weakly Bound Light morphemes, for example, *diSelat Baru,* with G-Word constituency *di- [Selat Baru]*. Now if the ludling were to apply to each of the terminal G-Words individually, the result would be *diWarelat Wararu, while if it were to apply as a whole to the non-terminal G-Word, the result would be *diWarelat Baru – but neither of these are attested. Instead, the observed output, *Warelat Wararu*, is accounted for straightforwardly by reference to the P-Word, as per General Rule (41).

Thus, in addition to providing further evidence for the Core Foot, the Warasa ludling also joins forces with Pantun prosody in providing evidence for the P-Word. In fact, these constitute the only two unambiguous sources of evidence for the P-Word in Riau Indonesian.

3.3.13   *Bahasisa ludling*

The last of the ludlings considered here, the Bahasisa ludling, is somewhat more complex than the preceding two, providing evidence for the now familiar Core Foot, and, in addition, to a more limited extent, also for Word structure. However, whereas the Sabaha ludling made reference to the G-Word, and the Warasa ludling to the P-Word, evidence for Word structure from the Bahasisa ludling is somewhat more ambivalent.

The main rules for the Bahasisa ludling are summarised in (50) below:

(50) Bahasisa Ludling
   a.   General Rule
        The last syllable of the Core Foot / terminal G-Word, C_1VC_2, is replaced
        with the sequence C_1iC_1a
   b.   Minor Rule 1: *u* retention
        If the last vowel of the Core Foot / terminal G-Word is *u*, then the last
        syllable of the Core Foot / G-Word, C_1uC_2, is replaced either with C_1iC_1a,
        as per (50a), or with C_1uC_1a
   c.   Minor Rule 2: *i* insertion
        If the first syllable of the Core Foot is empty, then an *i* is inserted in front
        of the final C_1iC_1a (or C_1uC_1a) sequence.
   d.   Minor Rule 3: end-point markers *-kan* and *-in*
        The end-point markers *-kan* and *-in* are retained, but their final consonant
        is deleted.

A couple of examples of naturalistic utterances making use of the Bahasisa ludling
are given in (51) and (52) below:[62]

(51) *Akika bisisa basisa   itita   sorira   ajija*
      Aku   bisa   basa   itu   sore   aja
      1SG   can   language DEM:DEM.DIST afternoon NEG.FOC
      [A foreign film-making crew wishes to record the Bahasisa ludling, but the
      speaker is coy]
      'I can only speak that language in the afternoon'

(52) *Janginga malila, santita ajija,   anggingga orira   samima kitita*
      Jangan   malu   santay aja   anggap   orang sama   kita
      NEG.IMP shy   relax NEG.FOC consider person together 1.2
      semuwiwa ajija
      semuwa   aja
      all   NEG.FOC
      [A foreign film-making crew wishes to record the speaker and her friend, but
      her friend is shy]
      'Don't be shy, just relax, just consider them to be people the same as us'

Examples (51) and (52) above illustrate the application of the General Rule in
(50a), which, in practice, is sufficient to account for the large majority of occurring
Bahasisa forms. Select cases of its application are represented in Figure 19 below.
In (a), the General Rule applies to *orang*, replacing the final syllable *rang* with *rira*,

---

62. Examples (51) and (52) were recorded in Sungai Pakning, where the Bahasisa ludling was
used mostly by teenage girls.
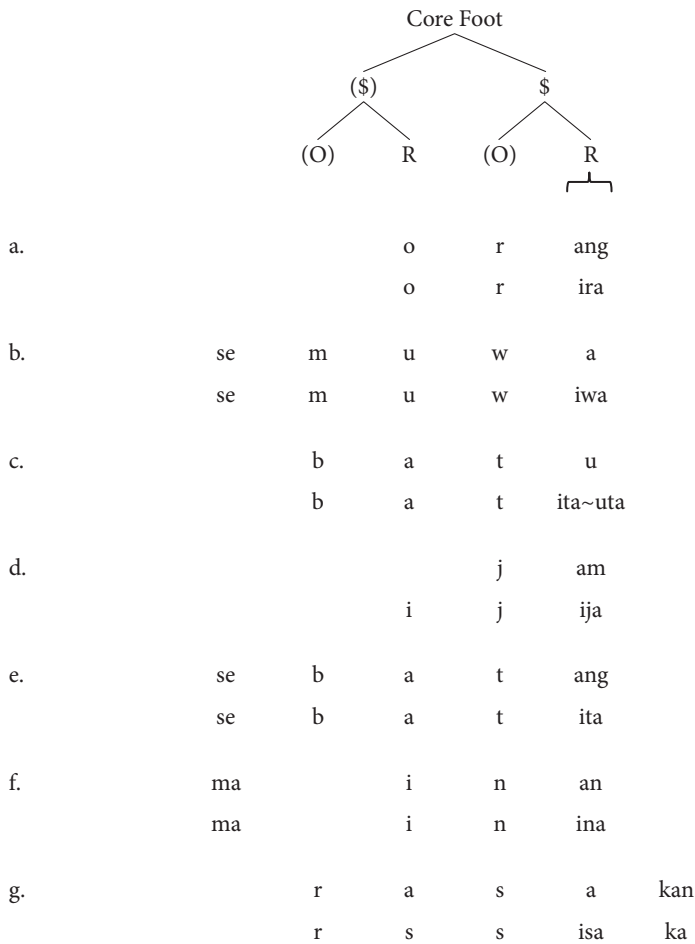
Figure 19. Bahasisa ludling

while in (b), the same rule applies to *semua*, replacing the final syllable [wa] with [wiwa]. Note that since the last syllable of the Core Foot is always the last syllable of the terminal G-Word, the rule in (50a), as also that in (50b), cannot distinguish between these two structural levels.

In order to gain further insight into the workings of the Bahasisa ludling, it is necessary to supplement the naturalistic data with data obtained from elicitation, focussing on some of the less frequently occurring cases. Example (53) shows cases of monomorphemic forms whose final vowel is *u*:

(53) a. *batu* → *batita* ~ *batuta*
　　　 stone
　　　 'stone'

b.   *hidung*   →   *hidida ~ hiduda*
       nose
       'nose'
c.   *peluk*   →   *pelila ~ pelula*
       hug
       'hug'

For these forms, two alternative ludling forms are available: the *u* is either replaced with *i*, as per the General Rule in (50a), or else retained as *u*, in accordance with Minor Rule 1 in (50b); see Figure 19 example (c).

Example (54) shows how the Bahasisa ludling applies to monosyllabic Free Footed morphemes:

(54)   a.   *jam*   →   *ijija*
              hour
              'hour'
        b.   *song*   →   *isisa*
              rummy
              'rummy'
        c.   *yang*   →   *iyiya*
              REL
              [relative marker]

On its own, the General Rule in (50a) would predict the forms **jija*, **sisa* and **yiya*. However, in the case of monosyllabic forms such as these, an additional vowel, *i*, is added at the beginning of the form, as specified in Minor Rule 2 in (50c). The actually occurring forms, *ijija, isisa* and *iyiya*, may be analysed as resulting from a two-stage process. At stage 1, *i* is added to the original forms, yielding the intermediate forms *ijam*, *isong* and *iyang*.[63] At stage 2 the General Rule then applies, to result in the actually occurring forms. As shown in Figure 19 example (d), the effect of *i* insertion in accordance with Minor Rule 2 is in fact very similar to that of the ordinary-language rules discussed in Sections 3.3.4–7, epenthesis and spreading, loan form expansion, truncated name expansion, and realisation of *N-* as *nge-*, namely to add phonological content to an empty first-syllable position, and thereby beef up an otherwise defective Core Foot. Thus, examples such as those in (54),

---

63. Stage 1 represents a particular case of a wider phenomenon whereby an ordinary-language form is modified or selected with the specific purpose of constituting a more appropriate input into a ludling rule. Another example of this phenomenon, mentioned in footnote 60 above, in the context of the Warasa ludling, is the choice of the Standard Indonesian negator *tidak* in place of its Riau Indonesian counterpart *tak*, motivated by the desire to preserve an overt expression of the negative meaning.

and Minor Rule 2 in (50c) that accounts for them, provide additional evidence in support of the Core Foot in Riau Indonesian.

The following examples show how the Bahasisa ludling applies to Free Footed morphemes in construction with a Bound Light morpheme:

(55) a. *sebatang*   → *sebatita*
          one:long.object
          'one piece'
      b. *dibukak*   → *dibukika*
          PAT:open
          'open'

(56) a. *mainan*   → *mainina*
          play:AUG
          'toy'
      b. *takbiran*   → *takbirira*
          profession.of.faith:AUG
          'profession of faith'

(57) a. *rasakan*   → *rasisaka*
          feel:EP
          'feel'
      b. *ngapain*   → *ngapipai*
          AG:what:EP
          'do what'

In (55), where the Bound Light morphemes precede their hosts, the ludling applies straightforwardly, and does not distinguish between Strongly Bound *se-* and Weakly Bound *di-* – see Figure 19 (e). However, in (56) and (57), where the Bound Light morphemes follow their hosts, a more complex picture emerges. In (56), the Strongly Bound Light morpheme *-an* is located within the Core Foot and therefore falls within the scope of the ludling, as shown in Figure 19 (f). In contrast, in (57), the Weakly Bound Light morphemes *-kan* and *-in* are located after the Core Foot, and therefore the ludling applies to the host morpheme within the Core Foot, as indicated in Figure 19 (g).

The examples in (57) also illustrate the application of Minor Rule 3 in (50d): the final *n* of each of the two end-point markers is deleted, and instead of the expected forms **rasasikan* and **ngapipain*, the actually occurring forms are *rasisaka* and *ngapipai*. Recall, however, that the two end-point markers are external to the G-Word; in terms of G-Word structure, they are indistinguishable from other Loosely Bound Light morphemes that might happen to precede the following host morpheme. However, whereas such other Loosely Bound Light morphemes preceding the following host morpheme are unaffected by the ludling, *-kan* and *-in* are

affected, in accordance with Minor Rule 3. A possible account for this difference would therefore be in terms of P-Word structure: a revised and more explanatory version of Minor Rule 3 might specify that Bound Light morphemes occurring in the P-Post-Foot position undergo reduction from C_1VC_2 to C_1V.[64]

Examples (58) and (59) show how the ludling applies to Compounds and Reduplications:[65]

(58) a. *tiga   belas*   → *tigiga belila*
          three over.ten
          'thirteen'
      b. *korek   api*   → *korira apipa*
          match   fire
          'match'

(59) a. *datang datang* → *datita datita*
          DISTR~ arrive
          'arrive'
      b. *kupukupu*   → *kupipa kupipa ~ kupupa kupupa*
          butterfly
          'butterfly'

Finally, (60) shows how the ludling applies to constructions involving Reduplication plus a Bound Light morpheme:

(60) a. *dibukak   bukak*   → *dibukikak bukikak*
          PAT:DISTR~ open
          'open'
      b. *sebatang sebatang*   → *sebatita sebatita*
          DISTR~   one:long.object
          'one piece each'
      c. *masuk   masukkan*   → *masisa masisaka*
          DISTR~ go.in:EP
          'put in'

---

64. Given that the class of Bound Light morphemes occurring in the P-Post-Foot position consists of just three items, *-kan, -in,* and *-nya*, it would not seem unreasonable to entertain alternative analyses under which the class of items to which Minor Rule 3 applies, namely *-kan* and *-in*, is defined phonologically, with reference to the final consonant *-n*, or even lexically, thereby avoiding reference to the structural unit of P-Word. Given the paucity of the data, I see no way of adjudicating between such alternative accounts.

65. Note that whereas *datang datang* in (59a) is a bona fide instance of Reduplication, *kupukupu* in (59b) is a case of "phonological copying", in that **kupu* alone is unattested. Nevertheless, the ludling treats both forms in the same way.

As was the case for the two preceding ludlings, the Bahasisa ludling applies to each term separately. Unlike the preceding ludlings, however, constructions involving Compounds and Reduplication provide no evidence for adjudicating between alternative analyses based on P-Words or G-Words.

In summary, then, the Bahasisa ludling provides a rather mixed bag of evidence for various levels of Word structure. Unambiguous support for the Core Foot is provided by Minor Rule 2, in which a monosyllabic Footed input to the ludling is beefed up by the addition of the vowel *i*. Tentative support for the P-Word is provided by Minor Rule 3, which applies to a small class of items defined in terms of P-Word structure. And finally, ambivalent support for either the Core Foot or the G-Word is provided by the General Rule for the Bahasisa ludling, which can be described as applying to either of the two structural levels, with no obvious way to adjudicate between the two analyses.

Comparing the three ludlings examined in the preceding three subsections, the emerging picture is one of variability, with one ludling making reference to P-Words, another providing evidence for G-Words, and the third pointing somewhat more tentatively towards analyses couched in terms of both types of Words.

3.3.14 *Naturalistic spelling*

The final source of evidence for Word structure considered here is naturalistic spelling, which provides support for the existence of G-Word structure, and more weakly also for P-Word structure, in Riau Indonesian.

This chapter began with a critique of the role of orthography in linguistic analysis, arguing that, when invoked inappropriately, it may reflect and then further reinforce misconceptions about grammatical structure and the relevance of a notion of word. Nevertheless, once a hypothesis about Word structure, such as that in Section 3.2 above, is put forward, it may be examined in the light of an empirical study of how people actually choose to represent their language in writing. The focus here is thus on naturalistic spelling, and in particular on how it deviates from the prescriptive norms of the standard language orthography.

The study presented here is based on a corpus of 4142 SMS messages sent, mostly to me, during 2003 and early 2004. The year 2003 was notable in that it was the year during which an early generation of mobile phones became ubiquitous, and many speakers were using one for the very first time. More often than not it was the first time they were called upon to commit speech to writing since their school days, and the very first time that they attempted to write in their own colloquial language, Riau Indonesian. A decade or more later, with the advent of smartphones and social media, pan-Indonesian orthographic norms gained wide acceptance across the nation, involving conventionalised spellings embellished with emoticons, camel case, and other optional stylistic accoutrements – as exemplified in (1). However, the

present corpus dates from an earlier era in which speakers were actually inventing their own orthography, on the fly, combining their often imperfect recollections of Standard Indonesian orthography with their own native-speaker linguistic intuitions for Riau Indonesian, in order to create a new system of writing from scratch.

Examples (61)–(68) below show some SMS messages from the corpus:

(61) *Udah puas aku ngirim tapi takmasuk2 juga, entah apa sebab nya tak tau lah dapit*
      Udah puas   aku ngirim   tapi tak   masuk masuk juga,   entah
      PRF   satiated 1SG AG:send but NEG DISTR~ enter ADD.FOC uncertain
      apa   sebabnya,   tak tau   lah, Dapit
      what reason:ASSOC NEG know FOC David
      [Complaining that his messages don't get through to me]
      'I've already had enough of my messages not getting through, I don't know what the reason is, I just don't know, David'

(62) *Pit anton udah datang ke pakning,dapit datang kepakning kapan pit*
      Pit,   Anton udah datang kePakning, Dapit datang kePakning
      HYP\David Anton PRF   arrive   ALL:Pakning David arrive   ALL:Pakning
      kapan, Pit
      when   HYP\David
      [Asking about my next trip to Sungai Pakning]
      'David, Anton's already arrived in Pakning, when are you arriving, David?'

(63) *buah salak di tepi perigi jangan sampai jatuh sebiji kenapa david rindu kan pii spai2 lupakan adi*
      Buah salak ditepi   perigi, Jangan   sampai jatuh sebiji,   Kenapa David
      fruit salak LOC:edge well NEG.IMP reach   fall   one:seed why   David
      rindukan Pi'i, Sampai sampai lupakan Adi
      miss:EP   Pi'i DISTR~ reach   forget:EP Adi
      [Pantun, speaker Adi accusing me of paying more attention to his friend Pi'i]
      'Salak fruit on the edge of the well, Do not let one happen to fall, Why is it you miss Pii, so much so that you forget Adi'

(64) *Sayang maaf ya,sur baru sekarang membalas sms nya habisnya hp sur di tarok di lemari jadi ngak,i miss you.*
      Sayang   maaf ya, Sur baru sekarang membalas SMSnya
      compassion forgive yes Sur new now   AG:reply   SMS:ASSOC
      habisnya   hp   Sur ditarok dilemari,   jadi
      finish:ASSOC handphone Sur PAT:put LOC:cupboard become
      nggak? I miss you
      NEG
      [From husband named Sur to wife]
      'Sorry, love, I'm only just responding to your message, because I put the handphone in the cupboard, so will it happen? I miss you'

(65)   *Dapit jagan lupa nelepon natimalam kalau takmasuk berakti battarai adi abis*
     Dapit   jangan   lupa   nelepon   nanti   malam kalau tak   masuk
     David NEG.IMP forget AG:telephone FUT.PROX night   TOP   NEG go.in
     berarti   baterai Adi   habis
     NON.PAT:mean battery Adi   finish
     [Asking me to call]
     'David, don't forget to call later tonight, if you don't get through it's because
     my battery is dead'

(66)   *Nomorape dapit di jerman*
     Nomor   hape   Dapit   diJerman
     number handphone David LOC:Germany
     [Asking how he can contact me in Germany]
     'Your mobile phone number in Germany?'

(67)   *malam vid dahsampai jakarta telpon aku cepat*
     Malam Vid   dah sampai Jakarta telepon   aku cepat
     night   HYP\David PRF arrive   Jakarta telephone 1SG fast
     [Asking me to call]
     'Good evening David, when you've arrived in Jakarta call me right away'

(68)   *DAPIT KENCANA LIMABELASMENIT LAGI*
     Dapit   Kencana lima belas   menit   lagi
     David Kencana five   over.ten minute ADD.FOC
     [On pier, about boat soon to arrive]
     'David, the Kencana will be here in fifteen minutes'

Examples (61)–(64) show how the same Light morpheme can be written either separately or joined on to its host even within the same SMS message: this is true of the Free Light morpheme *tak-* in (61), and the Bound Light morphemes *ke-* in (62), *-kan* in (63), and *-nya* in (64) – the latter precisely mirroring the variation shown in the Facebook status update in (1) at the beginning of this chapter. Examples (65)–(68) provide instances of a sequence of two or more Free Footed morphemes being written joined on to each other, rather than separately as is usually the case: *natimalam* in (65), *nomorape* in (66), *dahsampai* in (67), and *LIMABELASMENIT* in (68).

    The variation observed in SMS examples such as the above is not random. Table 4 below presents data for 16 selected morphemes, showing how they are written in the 4142 SMS message corpus. A final row provides similar data summing over the class of two-term toponyms, including forms such as as *Pekan Baru*, *Sungai Pakning* and the like. The first column presents the morpheme;[66] the second

---

66. In this column, "*di-* PAT" and "*di-* LOC" represent the two usages of *di-* without taking a stand on whether they should indeed be distinguished in Riau Indonesian (see further discussion

Table 4. The SMS corpus: Joined and separate spellings

Form	Standard orthography	Morpheme type		Tokens	Joined	Separate	% separate	% deviation
yang	separate	B	F	267	1	266	<100	<0
nggak	separate	F	F	214	2	212	99	−1
dari	separate	B	F	245	6	239	98	−2
tak	separate	F	L	758	39	719	95	−5
ni	separate	F	L	144	9	135	94	−6
di- (PAT)	joined	WB	L	221	51	170	77	+77
di- (LOC)	separate	WB	L	597	181	416	70	−30
-nya	joined	WB	L	569	170	399	70	+70
ke- (ALL)	separate	W/SB	L	309	133	176	57	−43
-kan	joined	WB	L	236	109	127	54	+54
si-	separate	SB	L	17	8	9	53	−47
-in (other)	joined	WB	L	53	50	3	6	+6
ter-	joined	WB	L	81	79	2	2	+2
ber-	joined	W/SB	L	209	207	2	1	+1
(ngapa)-in	joined	WB	L	249	248	1	>0	>0
se-	joined	W/SB	L	217	217	0	0	0
toponyms	variable			55	12	43	78	

column shows how its counterpart is written in Standard Indonesian; the third and fourth columns indicate the classification of the morphemes in accordance with the two dimensions of Table 1 in Section 3.2.2;[67] the fifth column displays the total number of tokens of the morpheme in the corpus; the sixth and seventh columns present the number of tokens that are written joined on and separately respectively; and the eighth column presents the percentage of tokens that are written separately – the rows of Table 4 are ordered in accordance with this column, in descending percentages, from separate to joined on. The ninth and final column indicates the extent, in percentages, to which the naturalistic orthography deviates from that of the standard orthography, and the direction of the deviation: "+" for

---

below). Similarly, "*ke-* ALL" represents just the allative usage of *ke-*, ignoring various other functions associated with *ke-*. Finally, for the end-point marker *-in*, a distinction is made between its usages in the form *ngapain*, where it is probably most appropriately considered to be lexicalised, and its more productive albeit somewhat less common usages in construction with other hosts.

67. In the first of these two columns, F stands for Free, B for Bound, W for Weakly, and S for Strongly; some morphemes are indicated as occurring in two guises, W/S. In the second of these columns, F stands for Footed and L for Light.

deviations in the direction of being written separately, "−" in the direction of being written joined on.[68]

Inspection of Table 4 above reveals clear tendencies, which may be summarised as follows, the ">" sign meaning "more likely to be written separately in the SMS corpus":

(69) a. written separately in Standard Indonesian > written joined in Standard Indonesian
     b. Free > Bound
     c. Weakly Bound > Strongly Bound
     d. Footed > Light

While (69a) represents the obvious fact that the SMS orthography is at least partially modelled after that of Standard Indonesian, (69b–d) reflect the typology of morphemes proposed in Table 1. This poses the question how to distinguish between these different but converging factors. The issue is complicated by the fact that Standard Indonesian orthography is itself presumably a reflection, albeit an imperfect one, of a linguistic reality associated with one or more earlier varieties of Malay/Indonesian not entirely dissimilar to Riau Indonesian.

The relative contributions of the standard orthography and the typology of morphemes in Table 4 may be teased apart through inspection of the figures in the final two columns, showing the percentage of separate spellings followed by percentage of deviation of the naturalistic spellings from the standard orthography. The data in Table 4 divide relatively neatly into three distinct chunks, set apart by the horizontal lines. In terms of their spelling in the SMS corpus, the top chunk is characterised by largely consistent separate spelling (94% to near 100% separate), the middle chunk by variable spelling (53% to 77% separate), and the bottom chunk by mostly consistent joined on spelling (0% to 6% separate). In terms of their deviation from the standard orthography, the top and bottom chunks exhibit relatively small degrees of deviation (−6% to +6%), whereas the middle chunk exhibits much more significant degrees of deviation (in the direction of joined spelling −30% to −47%, in the direction of separate spelling +54% to +77%). It is thus the middle chunk of Table 4 that provides the strongest evidence for the relevance to naturalistic spelling of the typology of morphemes in Table 1 and the Word structure underlying it.

---

68. The figures in this column are calculated in the following way. First, the values for the standard orthography are converted to percentages: 100% for separate, 0% for joined. Then these values are subtracted from the corresponding values for the SMS corpus. For example, for *si-*, in the SMS corpus it is written separately in 53% of the cases, in the standard orthography it is written separately in 100% of the cases, so 53% − 100% = −47%, which represents the deviation from the standard orthography in the direction of being written joined on.

Before examining the data in more detail, a specific issue, pertaining to the "two" Bound Light morphemes *di-* needs to be dealt with. As suggested in Gil (2002b), patient-oriented and locative *di-* may perhaps be more appropriately analysed as instances of a single polyfunctional or even macrofunctional form *di-*, making reference to a generalised thematic role encompassing both patient and location roles. Whatever the merits of that analysis, it is likely that actual speakers of Riau Indonesian find it difficult to distinguish between the putative distinct markers, as is necessary in order to meet the demands of the standard orthography. And indeed, as shown in Table 1, patient-oriented and locative *di-* are actually spelled similarly in the SMS corpus, (77% and 70% separate respectively). Thus, a significant proportion of the observed deviation in the naturalistic spellings of patient-oriented and locative *di-* from their prescriptive spellings (+77% and −30% respectively) might be due to confusion on the part of speakers with regard to the distinction between the two. Since there is no obvious way to assess the extent to which such confusion contributes to the observed deviation of the naturalistic spellings of these forms, we shall put aside these two forms, and focus on the remaining four forms in the middle chunk of Table 4: associative *-nya*, allative *ke-*, end-point *-kan*, and personal marker *si-*.

With respect to these latter four forms, the relevance of the typology of morphemes to the naturalistic spellings of the SMS corpus is striking. Although they are written joined on in the standard orthography, Weak Bound morphemes *-nya* and *-kan* are spelled separately in 70% and 54% of cases in the SMS corpus; conversely, although written separately in the standard orthography, Strong Bound morphemes *ke-* and *si-* are spelled joined on in 43% and 47% of cases in the SMS corpus. The emerging generalisation is that Weak Bound morphemes tend to be written separately, while Strong Bound morphemes tend to be written joined on.

In turn, this generalisation provides evidence for Word structure. For Weakly Bound *-nya* and *-kan*, their separate spellings run counter not only to the standard orthography but also to P-Word structure: the reason for them being spelled separately in the SMS corpus is that they are not part of the G-Word. For *ke-* and *si-* the picture is slightly more ambivalent: while their joined on spellings conflict with the standard orthography, they are consistent with Word structure, but fail to distinguish between alternative analyses based on G-Words and P-Words, since, as Strongly Bound morphemes, they occur within both types of Word.

Some of the other, less-frequent deviations from the standard orthography evident in Table 4 are also revealing. The non-standard spellings of Free Light *tak* and *ni* (written 5% and 6% joined on respectively) point towards a Wordhood effect; however, since *tak* and *ni* lie outside the G-Word, the relevant level of structure must be the P-Word. Conversely, the non-standard spelling of Weakly Bound *-in* (in constructions other than *ngapain*) (written 6% separately) also suggests a Wordhood

effect, except that in this case, since *-in* is part of the P-Word, the relevant level of structure here is the G-Word.

In summary, then, pretty much all of the significant deviations from the standard orthography observable in the SMS corpus can be accounted for in terms of Word structure, with the G-Word playing the major role, alongside a smaller contribution on the part of the P-Word.

Additional evidence for the G-Word is provided by the data for toponyms in the last row of Table 4. In 22% (100%−78%) of the cases, these toponyms are written joined on; such spellings are a direct reflection of the non-terminal G-Word. Joining force with these spellings are the ones in (65)–(68), in which a sequence of two or more Free Footed morphemes are written joined on to each other; in these cases too, the spellings constitute a reflection of a non-terminal G-Word.[69]

Thus, in this subsection, we have seen how naturalistic spellings, in the ways that they deviate from the prescribed orthography of Standard Indonesian, provide additional support for the existence of G-Word structure in Riau Indonesian, plus, somewhat less strongly, evidence for P-Word structure. Rather than allowing writing systems to lead us astray in our quest for a better understanding of linguistic structures, such systems can instead, when studied appropriately, actually provide an additional source of evidence for grammatical structures.

3.4   Rounding up the evidence

In Section 3.3 we examined 14 sources of evidence for the existence of Core Foot, P-Word and G-Word levels of structure in Riau Indonesian. These 14 sources of evidence were all I could think of, in a good-faith attempt to gauge the extent to

---

69. While the preponderance of the evidence from naturalistic spelling points towards the G-Word as the most relevant structural level, it must be acknowledged that there is significant variation in such spellings, reflecting both individual styles and rapidly changing fashions. In particular, Gil (2015: 314–317) discusses another kind of commonly-observed deviation from the conventions of Standard Indonesian orthography, involving what appears to be a "maximal-word" preference, whereby potentially polysyllabic orthographic words tend to be broken up into smaller units of at most two syllables: for example, *Manokwari* (a monomorphemic toponym) is written as *mano kuari*, *berangkat* (NON.PAT:lift) as *berang kat*, *Bengkalis* (another monomorphemic toponym) as *beng kalis*. Note that in the preceding examples, the spelling breaks up a single morpheme into two different orthographic units. This would seem to suggest that the operative principle in such cases is of a phonological nature. However, although seemingly related to the strong preference for disyllabicity associated with the Core Foot, these disyllabic orthographic words are not always congruent with the Core Foot; for example, in *berangkat*, it is *rangkat* rather than *berang* that occupies the Core Foot. Clearly, the principles involved in such cases are different to the ones focussed on in this subsection.

which the grammar of Riau Indonesian requires recourse to categories such as these. It is of course possible that there may be further sources of evidence that have escaped my attention. Ludlings alone provide a promising hunting ground for such evidence; in addition to the three discussed here, I am aware of the existence of several others, for which I still have insufficient data to be able to determine if and how they may be relevant to Word Structure in Riau Indonesian. And there may well be other sources of evidence beyond these, still awaiting discovery.

It is perhaps worthy of note that, of the 14 sources of evidence for Word structure considered here, 5 are para-linguistic, involving metred verse, ludlings, and naturalistic orthography. One might possibly argue that, by dint of their extraneous nature and less frequent usage, such sources of evidence should carry less weight than their more hard-core grammatical counterparts. On the other hand, one could, equally plausibly, make the opposite case: precisely because such para-linguistic phenomena do not figure in ordinary round-the-clock language usage, the sometimes quite subtle patterns that they reveal are not likely to have been learned through direct exposure to the phenomenon in question, but must instead constitute a reflection of some deeper properties of grammatical organisation. Looking at things this way, such para-linguistic phenomena might actually be afforded even greater weight than their ordinary counterparts.

With the above qualifications in mind, the detailed examination of Word structure presented in Section 3.3. paints a reasonably faithful picture of the roles and extent of Word structure in Riau Indonesian. The conclusion that emerges is that Word structure is indeed present, but not really all that pervasive. While 10 of the 14 sources of evidence converge to provide strong motivation for a well-defined Core Foot, just 6 of the 14 sources provide evidence for some higher level of structure; however, they do so without converging on a uniquely defined Word. Instead, 4 of the 6 sources support the P-word, 2 strongly, the remaining 2 more weakly, while 3 or 4 of the 6 sources provide support for the G-Word. Thus, evidence for the P-Word and the G-Word in Riau Indonesian while substantial, is hardly overwhelming.[70]

---

70. It is worth noting at this point that several of the other Malay/Indonesian varieties spoken in Riau and Kepulauan Riau provinces alongside Riau Indonesian may offer additional *prima facie* sources of evidence for Word structure of some kind. To cite just a few examples: In Siak Malay, Word-final [a] changes to [o]; although the rule is mostly lexicalised, an underlying [a] may occasionally surface, e.g. in the first term of reduplicated forms such as [apa-apo] (DISTR~what). In Akit, voiced obstruent stops [b], [d], [j] and [g] trigger raising of a following [a] to [ə] which may apply at long distance within a Word domain, e.g. [durian] → [duriən] (thorn:AUG, 'durian'). And in Bangkinang, application of the end-point marker [-in] triggers reduction of [a] to [o] in Pre-Foot but Word-internal positions, e.g. [mati] ~ [motiin] (die:EP, 'turn off'). More detailed analysis of such phenomena is required before their relevance to Word structure can be adequately determined. Nevertheless, given the close contact between these language varieties

4. Riau Indonesian in typological perspective

The case study of Riau Indonesian conducted in Section 3 was emic in nature, taking the facts as the point of departure, and then accounting for them by positing language-specific descriptive categories of Core Foot, G-Word and P-Word. The above account now provides the necessary foundation enabling us to return, in a more empirically-enlightened fashion, to the general etic cross-linguistic typological questions posed in Section 2: Does Riau Indonesian instantiate a comparative concept of word? And is Riau Indonesian an isolating language?

With regard to the first question, the answer is a somewhat ambivalent and qualified no. In Section 2.2 it was suggested that a language instantiates a comparative concept of word to the extent that there exists a set of properties converging, more or less, on a structural domain, the language's own specific descriptive category of Word, that constitutes a cut-off point between two types of structure differentiated by the kinds of principles that are generally associated with morphology and syntax respectively. Clearly, Riau Indonesian G-Words and P-Words are the best possible candidates for such a comparative concept; however, the set of criteria that converge on them is meagre in comparison to that in evidence in many other languages. At the very most, one might conclude that the G-Words and P-Words of Riau Indonesian constitute marginal instantiations of a comparative concept of word.

By the same token, Riau Indonesian G-Words and P-Words would constitute at best relatively weak instances of more specific putative comparative concepts of grammatical and phonological words respectively. Not only is their support quantitatively relatively modest, but the relevant sources of evidence do not map straightforwardly onto a viable distinction between grammatical and phonological process. Setting aside the para-linguistic sources of evidence, which in principle could make reference to either grammatical or phonological structures and are therefore neutral to the distinction between the two, we are left with just a single source of evidence, the grammatical process of Reduplication, making reference to G-Words – slim pickings for a proposed distinction between comparative concepts of grammatical and phonological word.[71]

----

and Riau Indonesian, it is quite plausible that whatever Word structures are supported by such phenomena may easily be carried over, in situations of diglossia, from the respective language varieties into Riau Indonesian. More generally, examples such as these underscore the general claim being made in this chapter to the effect that the comparative concept of word is an abstract notion that is not tied to a particular menu of concrete linguistic features.

71. A possible additional argument supporting the identification of the Riau Indonesian descriptive category of G-Word with a potential comparative concept of grammatical word would derive from the interface of the admittedly very limited morphological structure of Riau Indonesian with

In general, a number of features characteristic of morphological systems, underlying descriptive categories of Word in other languages, are completely lacking from Riau Indonesian. One such feature, discussed in Section 2.1, is non-linearity; another, also discussed there, is paradigmatic syncretism – indeed, Riau Indonesian is completely lacking in agreement and associated paradigms, which are one of the most common sources of such paradigmatic syncretism. A third feature almost completely absent from Riau Indonesian is the phonological erasure of morpheme boundaries by morphophonemic rules, the only case of this being the nasal assimilation and replacement associated with some of the allomorphs of *N-* . For these and other reasons, Riau Indonesian is the kind of language that does not provide strong motivation for an Item and Process approach to morphology founded on a clear cut distinction between morphological and syntactic structures. Conversely, Riau Indonesian does lend itself readily to an Item and Arrangement approach in which similar or identical principles of a syntactic nature govern the concatenation of meaningful expressions all the way down to the smallest of individual morphemes.

A striking instantiation of such vertical homogeneity and corresponding absence of a clear cut distinction between morphology and syntax in Riau Indonesian comes from the domain of Reduplication and its relationship to Repetition. The problem is that Reduplication and Repetition in Riau Indonesian appear to constitute two ends of a cline with no clear cut-off point between the two. Gil (2005a) proposes 6 criteria distinguishing Reduplication from Repetition.[72] These criteria are etic in the sense that they are based on generalisations derived from clear cases of reduplication and repetition in other languages, and they are predicated

---

its more productive syntax. Specifically, in those cases where G-Word and P-Word constituency conflict, it would seem to be the case that it is the G-Word rather than the P-Word, that is relevant to syntactic structure. Consider, for example, constructions consisting of a Weakly Bound Light morpheme together with a Compound or Reduplication, e.g. *diSelat Baru* (LOC Selat Baru). In cases such as these, the syntactic structure, as well as its semantic interpretation, are clearly consonant with the G-Word constituency *di-* [*Selat Baru*] rather than the P-Word constituency [*diSelat*] Baru.

72. These criteria are summarised in Table (i) below:

Table i. Criteria for Reduplication and Repetition

Criterion	Reduplication	Repetition
unit of output	equal to or smaller than word	greater than word
communicative reinforcement	absent	present or absent
interpretation	arbitrary or iconic	iconic or absent
tonal domain of output	within one tone group	within one or more tone groups
contiguity of copies	contiguous	contiguous or disjoint
number of copies	usually two	two or more

on the assumption that reduplication is a morphological process while repetition is a syntactic one. However, Gil (2005a) shows that these criteria fail to converge towards a clear cut distinction in Riau Indonesian: in addition to many unambiguous cases of Reduplication, and many other unambiguous cases of Repetition, there are also a large number of cases where the criteria do not provide a clear characterisation of the construction in question as involving either Reduplication or Repetition.[73] Thus, the lack of a clear cut distinction between Reduplication and Repetition in Riau Indonesian underscores the absence of a clear cut distinction in Riau Indonesian between morphology and syntax, and the concomitant absence of a well-defined comparative concept of word as the cut off point between the two.

As a language failing to robustly instantiate the comparative category of word, Riau Indonesian may thus be characterised as an isolating language, in accordance with the definition presented in Section 2.3. Indeed, Riau Indonesian fits the bill of an isolating language in both of the ways proposed in Section 2.3: its Words are both typically small, containing few morphemes and endowed with minimal complexity, and in addition they are of relatively little import with regard to the grammatical organisation of the language, since the properties that define them are both few and divergent.

The characterisation of Riau Indonesian as an isolating language presented in this chapter forms the third and final piece in a triad of investigations leading towards a holistic typological profile of Riau Indonesian as an *Isolating-Monocategorial-Associational (IMA)* language – see Gil (2017). Monocategoriality, argued for in Gil (1994, 2000, 2005b, 2013), refers to the syntactic property of having but a single open syntactic category, that is to say, no distinction between lexical categories such as noun, verb, adjective and preposition, and no distinction between lexical categories and their phrasal projections. Associationality, argued for in Gil (1994, 2005b, 2012), pertains to the compositional semantics, and represents the state of affairs in which, given a juxtaposition of expressions, the meaning of the whole is associated with the meanings of the constituent parts in a minimally differentiated way, without recourse to additional more specific rules of semantic interpretation making reference to particular grammatical configurations – for example, *Ayam makan* (chicken eat) can mean anything to do with chicken and eating, without further specification of features such as the thematic role of the chicken, and whether

---

73. For example, in (12), *mandi mandi* is neutral with regard to all six criteria presented in Table (i) in the preceding footnote: (i) the unit of output is not clearly definable in terms of wordhood (this being the whole point of this chapter); (ii) communicative reinforcement is absent; (iii) the interpretation is iconic (iterative or frequentative); (iv) the tonal domain of output is contained within a single tone group; (v) the copies are contiguous; and (vi) the number of copies is two.

the meaning of the whole is predicative or attributive. Elsewhere it is argued that each of the three properties of an IMA language constitutes a default setting within its respective domain, representing the case of maximal simplicity, a foundation on which more complex structures may subsequently be built. For this reason, it is also argued that an IMA language may provide a model for earlier stages in the evolution of language itself; see Gil (2006a, 2009b, 2017).

Nevertheless, the three defining properties of an IMA language are not just logically independent of each other but also empirically so; languages seem to be able to vary freely along each of the three dimensions independently of their values with respect to the other two. In the case at hand, namely, that of isolating languages, this raises the question whether the property of being an isolating language bears any further consequences beyond itself, that is to say, whether it correlates with any other features of languages, in a way that would make it of interest to typologists.

Two negative observations are worthy of mention here. First, being isolating has nothing to do with monosyllabicity. As noted in Section 2.3 earlier, there is a common practice of anointing a particular language as representative of a linguistic type, following which unrelated properties of the language are lumped together as being representative or even criterial of the type in question. In the case at hand, Chinese was the prototypical isolating language, and since Chinese is monosyllabic, the two properties became indelibly connected in many people's minds. However, as Riau Indonesian clearly shows, a language need not be monosyllabic in order to be isolating.[74]

Secondly, being isolating does not necessarily entail having more complex syntax. It is commonly assumed that when languages lose their morphology, often due to sound changes, they compensate by tightening up their syntax. For example, in the development from Latin to modern Romance languages, loss of case marking is said to be made up for by the development of rigid word order – the assumption being that this was necessary in order for the language to continue to maintain the ability to distinguish, in one way or another, between thematic roles. However, as argued in Gil (2008), the so-called compensation hypothesis is belied by the existence of languages, Riau Indonesian among them, that use neither flagging (case marking or adpositions) nor word order to distinguish between thematic roles.

---

74. In fact, a reasonable case could even be made for Riau Indonesian being "more" isolating than many or most Sinitic languages. The crucial factor would be compounding, which appears to be much more common in Sinitic than in Malay/Indonesian, thereby significantly adding to the morpheme-to-word ratio of Sinitic. Indeed, Yip (1992) and others have argued for the existence of "word-minimality" constraints in Chinese, suggesting that the canonical word in Chinese may actually be a disyllabic, bimorphemic compound.

On a more positive note, however, the property of being isolating does in fact correlate with another, logically independent grammatical property, that of *low grammatical-morpheme density*. In accordance with Gil (2015: 317–320), a language is said to have low grammatical morpheme density to the extent that its sentences consist largely of "contentives", with a relative paucity of "grammatical" or "functional" markers of various kinds. Syntagmatically, low grammatical-morpheme density is evident in a low ratio of grammatical as opposed to contentive morphemes in texts, while paradigmatically, low grammatical morpheme density manifests itself in the predominance of grammatical markers that are optional as opposed to obligatory. Logically, it would be possible for an isolating language to exhibit high grammatical morpheme density, through an abundance of grammatical or functional morphemes that are separate words. However, this does not seem to happen, and for an obvious reason: the distinction between contentives and grammatical markers correlates positively with that between separate words and bound forms. This correlation in turn has a plausible diachronic explanation in the paths of grammaticalisation transforming contentives into grammatical markers, which combine twin processes of semantic and phonological erosion – the former changing more concrete meanings into more abstract ones, the latter reducing free standing words to bound forms. In this respect, at least, the notion of isolating language reveals itself to be of broader relevance to linguistic typology.

A final correlate of the isolating language type is of an expressly diachronic nature, and constitutes one of the central themes of this volume. Isolating language structure is often the outcome of processes of simplification associated with language contact, as evident in pidgins, creoles, Non-Hybridised Conventionalised Second Languages and other such language varieties. Indeed, as argued in Gil (this volume, Chapter 3), language contact may play a significant role in the history of Riau Indonesian, accounting, in part, for its nature as an isolating language.

The twin take-home messages of this chapter constitute a synthesis of positive and negative conclusions. On the negative side, Riau Indonesian is revealed to be a language whose grammar would appear to be lacking a robust descriptive category of Word. But on the positive side, the characterisation of Riau Indonesian as an isolating language underscores the relevance of the morphology/syntax distinction to linguistic typology, and the concomitant viability of the cross-linguistic comparative concept of word.

Acknowledgements

References

Anderson, Stephen R. 1982. Where's morphology? *Linguistic Inquiry* 13: 571–612.

Anderson, Stephen R. 1985. Typological distinctions in word formation. In *Language Typology and Syntactic Description, Volume III: Grammatical Categories and the Lexicon*, T. Shopen (ed.), 3–56. Cambridge: CUP.

Anderson, Stephen R. 1992. *A-morphous Morphology*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511586262

Arkadiev, Peter M. 2005. *Grammaticalization of polysynthesis (with special reference to spoken French)*. 4th Typological School, Tsakhkadzor, Armenia, 26 September 2005.

Aronoff, Mark. 1994. *Morphology by Itself, Stems and Inflectional Classes*. Cambridge MA: The MIT Press.

Azizul Rahman Abdul Rahman. 1995. *Bahasa Rahsia Bahasa Melayu* (Secret Languages in Malay). Bangi: Penerbit Universiti Kebangsaan Malaysia.

Baerman, Matthew, Brown, Dunstan & Corbett, Greville G. 2017. *Morphological Complexity*. Cambridge: CUP. https://doi.org/10.1017/9781316343074

Bagemihl, Bruce. 1988. Alternate Phonologies and Morphologies. PhD dissertation, University of British Columbia, Vancouver.

Bagemihl, Bruce. 1989. The crossing constraint and backwards languages. *Natural Language and Linguistic Theory* 7: 481–549. https://doi.org/10.1007/BF00205156

Blust, Robert A. 1988. *Austronesian Root Theory: An Essay on the Limits of Morphology* [Studies in Language Companion Series 19]. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.19

Blust, Robert A. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Bresnan, Joan & Mchombo, Sam M. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13: 181–254. https://doi.org/10.1007/BF00992782

Bruening, Benjamin. 2018. The lexicalist hypothesis: Both wrong and superfluous. *Language* 94:1–42. https://doi.org/10.1353/lan.2018.0000

Brugman, Claudia & Conners, Thomas. 2018. Register properties of SMS and Twitter in Indonesian: A contrastive study. *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqy028

Chambert-Loir, Henri. 1998. Slang and secret tongues. In *Indonesian Heritage, Language and Literature*, John H. McGlynn (ed.), 132–133. Singapore: Archipelago Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.

Conners, Thomas, Bowden, John & Gil, David. 2015. Valency classes in Jakarta Indonesian. In *Valency Classes in the World's Languages*, A. Malchukov & B. Comrie (eds), 941–986. Berlin: De Gruyter Mouton.

Dardjowidjojo, Soenjono. 1978. *Sentence Patterns of Indonesian*. Honolulu HI: University of Hawaii Press.

Dixon, Robert M. W. 2010–2012. *Basic Linguistic Theory*. Oxford: OUP.

Dixon, Robert M. W. & Aikhenvald, Alexandra Y. 2002. Word: A typological framework. In *Word. A Cross-Linguistic Typology*, Robert M. W. Dixon & Alexandra Y. Aikhenvald (eds), 1–41. Cambridge: CUP.

Donohue, Mark & Sawaki, Yusuf. 2007. Papuan Malay pronominals: Forms and functions. *Oceanic Linguistics* 47: 253–276. https://doi.org/10.1353/ol.2007.0017

Dryer, Matthew S. 2005. Order of subject, object, and verb. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds), 386–9. Oxford: OUP.

Embick, David. 2015. *The Morpheme: A Theoretical Introduction*. Berlin: De Gruyter. https://doi.org/10.1515/9781501502569

Fabre d'Olivet, Antoine. 1816. *La langue hébraïque restituée et le vétitable sens des mots hébreux rétabli et prouvé par leur analyse radicale*. Paris: J.-M. Eberhart.

Geertzen, Jeroen, Blevins, James P. & Milin, Petar. 2016. The informativeness of linguistic unit boundaries. *Italian Journal of Linguistics* 28(2):1–24.

Gil, David. 1985. What does grammar include? *Theoretical Linguistics* 12: 165–172. https://doi.org/10.1515/thli.1985.12.2-3.165

Gil, David. 1990. הבטים בתאוריה פרוזודית: הבית היאמבי-על-יאמבי (hebetim beteoria prozodit: habayit hayambi-ʕal-yambi 'Aspects of Prosodic Theory; The Iambic-over-Iambic Quatrain'). *Hebrew Linguistics* 28–30:111–122. (In Hebrew)

Gil, David. 1991. The *Muwaššah*: Artistic convention or cognitive universal. In *Israel Oriental Studies XI, Studies in Medieval Arabic and Hebrew Poetics*, Sasson Somekh (ed.), 137–159. Leiden: Brill.

Gil, David. 1993. 'Il pleut doucement sur la ville': The rhythm of a metaphor. *Poetics Today* 14: 49–82. https://doi.org/10.2307/1773140

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17: 179–200. https://doi.org/10.1017/S0332586500003000

Gil, David. 1999a. Riau Indonesian: A language without words. Fifth International Conference on the Languages of Far East, South-East Asia and West Africa, St. Petersburg, Russia, 9 September 1999.

Gil, David. 1999b. Why words? Criteria for bond strength between morphemes in Riau Indonesian. Third Meeting of the Association for Linguistic Typology, Amsterdam, The Netherlands, 27 August 1999.

Gil, David. 2000. Syntactic categories, cross-linguistic variation and universal grammar. In *Approaches to the Typology of Word Classes*, Petra M. Vogel & Bernard Comrie (eds), 173–216. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110806120.173

Gil, David. 2001. Creoles, complexity and Riau Indonesian. *Linguistic Typology* 5: 325–371.

Gil, David. 2002a. Ludlings in Malayic languages: An introduction. In *PELBBA 15, Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya: Kelima Belas*, (Fifteenth Linguistics Conference of the Atma Jaya language and Culture Research Centre), Bambang Kaswanti Purwo (ed.), 125–180. Jakarta: Unika Atma Jaya.

Gil, David. 2002b. The prefixes *di-* and *N-* in Malay / Indonesian dialects. In *The History and Typology of Western Austronesian Voice Systems*, Fay Wouk & Malcolm Ross (eds), 241–283. Canberra: Pacific Linguistics.

Gil, David. 2004a. Learning about language from your handphone; *dan, and* and *&* in SMSs from the Siak River Basin. In *Kolita 2, Konferensi Linguistik Tahunan Atma Jaya*, Katharina Endriati Sukatmo (ed.), 57–61. Jakarta: Pusat Kajian Bahasa dan Budaya, Unika Atma Jaya.

Gil, David. 2004b. Riau Indonesian *sama*: Explorations in macrofunctionality. In *Coordinating Constructions* [Typological Studies in Language 58], Martin Haspelmath (ed.), 371–424. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.58.20gil

Gil, David. 2005a. From repetition to reduplication in Riau Indonesian. In *Studies on Reduplication*, Bernhard Hurch (ed.), 31–64. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110911466.31

Gil, David. 2005b. Word order without syntactic categories: How Riau Indonesian does it. In *Verb First: On the Syntax of Verb-Initial Languages*, [Linguistik Aktuell/Linguistics Today, 73], Andrew Carnie, Heidi Harley & Sheila A. Dooley (eds), 243–263. Amsterdam: John Benjamins. https://doi.org/10.1075/la.73.17gil

Gil, David. 2006a. Early human language was Isolating-Monocategorial-Associational. In *The Evolution of Language, Proceedings of the 6th International Conference (EVOLANG6)*, Angelo Cangelosi, Andrew D. M. Smith & Kenny Smith (eds), 91–98. Singapore: World Scientific.

Gil, David. 2006b. Intonation and thematic roles in Riau Indonesian. In *Topic and Focus, Cross-Linguistic Perspectives on Meaning and Intonation*, Chungmin Lee, Matthew Gordon & Daniel Büring (eds), 41–68. Dordrecht: Springer.

Gil, David. 2008. How complex are isolating languages?" In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94], Matti. Miestamo, Kaius Sinnemäki & Fred Karlsson, (eds), 109–131. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.08gil

Gil, David. 2009a. Austronesian nominalism and the thinginess illusion. *Theoretical Linguistics* 35: 95–114. https://doi.org/10.1515/THLI.2009.005

Gil, David. 2009b. How much grammar does it take to sail a boat? In *Language Complexity as an Evolving Variable*, Geoffrey Sampson, David Gil & Peter Trudgill (eds), 19–33. Oxford: OUP.

Gil, David. 2009c. Riau Indonesian: What kind of a language is it? *Kongres Internasional Masyarakat Linguistik Indonesia* (International Conference of the Linguistic Society of Indonesia) *KIMLI 2009*, 29–61. Malang: Masyarakat Linguistik Indonesia and Universitas Negeri Malang.

Gil, David. 2012. Where does predication come from? *Canadian Journal of Linguistics* 57: 303–333. https://doi.org/10.1017/S0008413100004795

Gil, David. 2013. Riau Indonesian: A language without nouns and verbs. In *Flexible Word Classes: Typological Studies of Underspecified Parts of Speech*, Jan Rijkhoff & Eva van Lier (eds), 89–130. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199668441.003.0004

Gil, David. 2015. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia, The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: De Gruyter Mouton.

Gil, David. 2016. Describing languoids: When incommensurability meets the language-dialect continuum. *Linguistic Typology* 20: 439–462. https://doi.org/10.1515/lingty-2016-0017

Gil, David. 2017. Isolating-Monocategorial-Associational language. In *Handbook of Categorization in Cognitive Science*, 2nd edn, Henri Cohen & Claire Lefebvre (eds), 471–510. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-08-101107-2.00020-8

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Grammar*, Joseph H. Greenberg (ed.), 73–113. Cambridge MA: The MIT Press.

Halle, Morris & Marantz, Alec. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Kenneth Hale & Samuel J. Keyser (eds), 111–176. Cambridge MA: The MIT Press.

Halle, Morris & Vergnaud, Jean-Roger. 1987. *An Essay on Stress*. Cambridge MA: The MIT Press.

Haspelmath, Martin. 2009. An empirical test of the agglutination hypothesis. In *Universals of Language Today* [Studies in Natural Language and Linguistic Theory 76], Sergio Scalise, Elisabetta Magni & Antonietta Bisetto (eds), 13–29. Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8825-4_2

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistics studies. *Language* 86: 663–687. https://doi.org/10.1353/lan.2010.0021

Haspelmath, Martin. 2011a. The gradual coalescence into 'Words' in grammaticalization. In *The Oxford Handbook of Grammaticalization*, Heiko Narrog & Bernd Heine (eds), 342–355. Oxford: OUP.

Haspelmath, Martin. 2011b. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45: 31–80. https://doi.org/10.1515/flin.2011.002

Haspelmath, Martin. 2015. Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In *How Categorical Are Categories? New Approaches to the Old Questions of Noun, Verb, and Adjective* [Studies in Generative Grammar 122], Joanna Błaszczak, Dorota Klimek-Jankowska & Krysztoff Migdalski (eds), 273–304. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614514510-009

Haspelmath, Martin. 2016. The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics* 17: 291–319.

Haspelmath, Martin. 2018. The last word on polysynthesis: A review article. *Linguistic Typology* 22: 307–326. https://doi.org/10.1515/lingty-2018-0011

Haspelmath, Martin. 2019a. How comparative concepts and descriptive linguistic categories are different. In *Aspects of Linguistic Variation*, Daniël van Olmen, Tanja Mortelmans & Frank Brisard (eds), 83–114. Berlin: De Gruyter.

Haspelmath, Martin. 2019b. Indexing and flagging, and head and dependent marking. *Te reo, the Journal of the Linguistic Society of New Zealand* 62(1): 93–115.

Hockett, Charles F. 1944. Review of *Linguistic Interludes* and *Morphology: The Descriptive Analysis of Words* (1944 edition) both by Eugene A. Nida. *Language* 20: 252–255. https://doi.org/10.2307/410127

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10: 210–234. https://doi.org/10.1080/00437956.1954.11659524

Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203: 88–111. https://doi.org/10.1038/scientificamerican0960-88

Hordofa, Kebbede & Unseth, Peter. 1986. 'BirdTalk' in Oromo. *Quaderni di Studi Etiopici* 6–7: 74–83.

Jackendoff, Ray S. 1977. *X-Bar Syntax: A Study of Phrase Structure*. Cambridge MA: The MIT Press.

Jensen, John. 1980. X̄ Morphology. In *Proceedings of the Eleventh Annual Meeting of the North Eastern Linguistic Society*, Victoria Burke & James Pustejovsky (eds), 155–172. Amherst MA: University of Massachusetts.

Jukes, Anthony. 2015. Asserting peripherality in Sulawesi: Local varieties and the rejection of Jakarta-based norms. 19th International Symposium on Malay/Indonesian Linguistics (ISMIL19), Jambi, Indonesia, 14 June 2015.

Katamba, Francis. 1993. *Morphology*. London: MacMillan.

Keenan, Edward L. 1976. Towards a universal definition of subject. In *Subject and Topic*, Charles N. Li (ed.), 303–333. New York NY: Academic Press.

Kluge, Angela. 2014. *A Grammar of Papuan Malay*. Utrecht: LOT.

Lefkowitz, Natalie. 1991. *Talking Backwards, Looking Forwards: The French Language Game Verlan*. Tübingen: Gunter Narr.

Lerdahl, Fred. 1992. Cognitive constraints on compositional systems. *Contemporary Music Review* 6: 97–121. https://doi.org/10.1080/07494469200640161

Lerdahl, Fred & Jackendoff, Ray. 1983. *A Generative Theory of Tonal Music*. Cambridge MA: The MIT Press.

Levin, Juliette. 1985. A Metrical Theory of Syllabicity. PhD dissertation, MIT.

Lieber, Rochelle. 1980. On the Organization of the Lexicon. PhD dissertation, MIT.

Manns, Howard. 2010. Indonesian slang in internet chatting. In *Studies in Slang and Slogans*, Sola Babatunde, Akin Odebunmi, Akin Adetunji & Mahfouz Adedimeji (eds), 71–99. Munich: Lincom.

Mithun, Marianne. 2005. On the assumption of the sentence as the basic unit of syntactic structure. In *Linguistic Diversity and Language Theories* [Studies in Language Companion Series 72] Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds), 169–183. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.72.09mit

Nevins, Andrew & Endress, Ansgar. 2007. The edge of order: Analytic bias in ludlings. *Harvard Working Papers in Linguistics* 12: 43–53.

Peterson, John. 2008. Kharia: A South Munda Language, Vol. 1: Grammatical Analysis. Habilitation thesis, University of Osnabrück.

Rizzolo, Olivier. 2007. Utrovački and Šatrovački: Description and theoretical perspectives of two Serbo-Croatian language games. In *Linguistic Investigations into Formal Description of Slavic Languages*, Hans-Werner Eroms, Ludwig M. Eichinger, Vilmos Agel & Peter Hellwig (eds), 27–40. Frankfurt: Peter Lang.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York NY: Harcourt, Brace and Co.

Shen, Yeshayahu. 1985. Action Structure in the Narrative Text. PhD dissertation, Tel Aviv University.

Shen, Yeshayahu. 1988. Schema Theory and the processing of narrative texts: The X-Bar story grammar and the notion of discourse topic. *Journal of Pragmatics* 12: 639–676. https://doi.org/10.1016/0378-2166(88)90051-3

Shen, Yeshayahu. 1989. The X-Bar grammar for stories: Story grammar revisited. *Text* 9: 415–467. https://doi.org/10.1515/text.1.1989.9.4.415

Shoshany, Ronit. 1986. Prosodic structures in Jeremiah's poetry. *Folia Linguistica Historica* 7: 167–206.

Sneddon, James Neil. 1996. *Indonesian: A Comprehensive Grammar*. Canberra: Allen & Unwin.

Sneddon, James Neil. 2006. *Colloquial Jakartan Indonesian*. Canberra: Pacific Linguistics.

Tadmor, Uri. 2004. The morpheme structure of prehistoric Malay: An internal reconstruction. *Eighth Symposium on Malay/Indonesian Linguistics* (ISMIL8), Penang, Malaysia, 1 August 2004.

Tadmor, Uri. 2009. Indonesian. In *World Loanword Database*, Martin Haspelmath & Uri Tadmor (eds). Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wold.clld.org> (11 December 2018).

Tallman, Adam J. R. To appear. Constituency and coincidence in Chácobo (Pano). *Studies in Language*.

Tallman, Adam J. R. & Epps, Patience. 2020. Morphological complexity, autonomy, and areality in Amazonia. In *Morphological Complexity*, Peter Arkadiev & Francesco Gardani (eds). Oxford: OUP.

Tallman, Adam J. R., Wylie, Dennis, Adell, Eric, Bermudez, Natalia, Camacho, Gladys, Epps, Patience, Everdell, Michael, Gutierrez, Ambrocio, Juarez, Cristian & Woodbury, Anthony C. 2018. Constituency and the morphology-syntax divide in the languages of the Americas: Towards a distributional typology. *21st Annual Workshop on American Indigenous Languages*. University of California, Santa Barbara.

van Minde, Don. 1997. *Malayu Ambong: Phonology, Morphology, Syntax*, Leiden: CNWS.

Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12: 245–274.

Yip, Moira. 1992. Prosodic morphology in four Chinese dialects. *Journal of East Asian Linguistics* 1(1): 1–35. https://doi.org/10.1007/BF00129572

The loss of affixation in Cham

Contact, internal drift and the limits of linguistic history

Marc Brunelle

University of Ottawa

Chamic languages have been spoken in Central Vietnam since about 600 AD. While Classical Cham (9th–15th centuries), had already lost a significant proportion of its Austronesian affixation, it also borrowed new affixes from Mon-Khmer. Modern Cham (16th–19th centuries) underwent another wave of reduction that led to a largely monosyllabic and affixless Colloquial Eastern Cham (20th–21st centuries).

In this paper I first look at representative Classical Cham inscriptions, establishing the extent to which they exhibit a reduction of affixation, and discussing possible contact scenarios that may have led to this reduction. I then assess the prevalence of affixation in Modern Cham manuscripts and in Colloquial Eastern Cham, and argue that the role Vietnamese played in Cham monosyllabisation must have been more indirect than previously assumed.

Keywords: Cham, loss of affixation, monosyllabisation, contact, Mon-Khmer, inscriptions

Chamic languages (Rade, Jarai, Bih, Haroi, Cham, Raglai, Chru, Tsat) are Malayo-Polynesian languages spoken in southern Vietnam, Cambodia and Hainan. Acehnese, a language spoken in northern Sumatra, is also closely related to Chamic, although it is still controversial if it is a Chamic language itself (Cowan 1991; Thurgood 1999), or a sister to Chamic (Shorto 1975; Durie 1990; Sidwell 2006; Brunelle 2019).[1] Chamic languages are closely related to Malayic languages, with which they form the Malayo-Chamic subgrouping. While the exact route followed by the Chams to reach Mainland Southeast Asia is mostly a matter of speculation,

---

1. The exact subgrouping of Acehnese has little bearing on the arguments made in this paper. My views on the matter are articulated in Brunelle (2019).

scholars agree that modern Chamic speakers are the descendants of the Sa Huỳnh archaeological culture that is attested in central Vietnam from 600 BC (Bellwood 1985; Bronson & White 1992).

The lexicon and grammatical structures of Chamic languages have clearly been influenced by neighbouring Mon-Khmer languages, but due to limited archaeological and historical evidence, the circumstances and actual social and linguistic mechanisms underlying the linguistic contact between the two language families are still unclear. In this chapter, I look at a specific type of possible contact-induced change in Chamic: the simplification of its Austronesian affixation. I revisit its chronology, try to determine the socio-demographic conditions under which it took place and assess the role of contact in its development.

In Section 1, I give the readers some background about morphological simplification in Chamic, discussing previous accounts of how it took place. In Section 2, I determine the prevalence of affixation in Cham inscriptions and manuscripts at different time periods and establish a chronology for its reduction, showing that this occurred in two stages: (1) Ancient Cham lost most of its affixes before the first Cham language stone inscriptions were carved in the 9th century (at the latest), and (2) colloquial Eastern Cham lost its remaining morphology because of a more general process of monosyllabisation that took place in the 19th–20th centuries.

In Section 3, I review evidence for contact between Chamic and Mon-Khmer populations and try to assess the likelihood that it had an impact on the reduction of affixation in Ancient Cham. This evidence leads me to argue that language shift and the accompanying simplification of grammatical structures, which imply intensive contact with Mon-Khmer languages, might have been less important than usually assumed, and probably occurred early in Cham prehistory.

In Section 4, I account for the monosyllabisation of colloquial Eastern Cham by invoking internal linguistic factors such as phonetic reduction and constraints on word shapes. Finally, in Section 5, I argue that Vietnamese may very well have played a role in monosyllabisation and the ensuing loss of remaining affixation in colloquial Eastern Cham, but that this role may have been more indirect than previously assumed, and I summarise the paper in Section 6.

Two general points I try to make in this chapter are that it is possible to use historical, epigraphic, archaeological and genetic evidence to fine-tune our representations of Chamic linguistic history; and that we need to achieve a finer-grained understanding of the causal chains that lead to contact-induced change.

1. The loss of affixation in Cham languages: Overview and previous proposals

Although Chamic languages belong to the Western-Malayo-Polynesian family, they have very little productive morphology. The languages of the family only preserve a small number of prefixes and infixes, and the best described Chamic language, Eastern Cham, has gone a step further, losing all affixation in its modern colloquial form.

As will be illustrated in more detail in Section 2, the loss of affixation in Cham seems to have taken place in two major stages, separated by several centuries of apparent stability. The first wave of morphological simplification must have happened early; the oldest attested Cham stone inscriptions, dating from the first millennium AD, show that the language had already lost most of the affixation typical of Malayo-Polynesian. Stone inscriptions only provide clear evidence for three productive affixes: the causatives *pa-* and *ma-* and the nominaliser *-an-*. While *ma-* and *-an-* can be traced back to Austronesian and Mon-Khmer, respectively, *pa-* is attested in both Austronesian and Mon-Khmer (Aymonier & Cabaton 1906; Thurgood 1999).[2] Unfortunately, it is difficult, if not impossible, to date the exact time at which most of the rich Malayo-Polynesian morphology was lost and affixes were borrowed from Mon-Khmer, but we can safely claim that this happened before the first inscriptions were carved in 9th century (at the latest).

After this initial reduction of the affixation system, there was a long period of stability stretching from the 9th to the 19th century, during which there is little evidence of further simplification.[3] In fact, most Chamic languages still remain at this stage, preserving a handful of productive prefixes and infixes. However, a second stage of morphological reduction has affected Eastern Cham since the 19th century: a process of monosyllabisation has eliminated affixes in its colloquial variety (Aymonier 1889; Lee 1966; Alieva 1991, 1994; Bùi 1996; Brunelle 2008b, 2009) and the grammatical functions of former affixes have been taken over by periphrastic devices.

The first stage of morphological reduction, the initial simplification of affixation in the entire Chamic branch, is sometimes attributed to contact with Mon-Khmer

---

2. Proponents of the Austric family, a putative macro-phylum that would include Austronesian and Austroasiatic, have attempted to link the Mon-Khmer nominalising infix *-an-* to the Austronesian perfective infix *-in-* (Reid 1994, 2005). However, since the functions of the two affixes do not coincide and since *-in-* is not attested in Malayic, there is no evidence for an Austronesian origin of Chamic *-an-*.

3. Some Austronesian affixes rarely found in stone inscriptions even resurface in Modern Cham manuscripts, like the frequentative/inadvertent prefix *ta-* (Aymonier 1989; Aymonier & Cabaton 1906; Moussay 2006), which is only attested in the 15th century C.215 Trà Kiêu inscription (Arlo Griffiths, p.c.)

languages (Alieva 1984; Thurgood 1999). Thurgood (2000) proposes that a large-scale and long-term language shift from Mon-Khmer to Chamic languages would have made the grammatical structures of Cham analytic and semantically transparent. In Section 2, I revisit this proposal and argue that while a massive shift to Mon-Khmer could very well have occurred, it would likely be circumscribed to an early stage of Chamic (pre)history and is not a necessary condition for morphological simplification.

The recent loss of all remaining affixation in modern Eastern Cham is not due to a population shift and seems tied to a more general process of monosyllabisation. The reduction of sesquisyllables to monosyllables is often attributed to contact with monosyllabic languages in Mainland Southeast Asian linguistic literature (Alieva 1991, 1994; Thạch 1999; Thurgood 1999; Grant 2005, 2007). As summarised by Thurgood: "…the subsequent reduction to monosyllables seems to be due in large part to subsequent Phan Rang Cham [Eastern Cham] contact with the monosyllabic Vietnamese and Utsat contact with the monosyllabic languages of Hainan" (Thurgood 1999: 66). I will argue that while language contact could possibly facilitate monosyllabisation, speakers are unlikely to make their first language monosyllabic to imitate a neighbouring language. I propose (following Brunelle & Pittayaporn 2012) a finer-grained scenario according to which language-internal processes set in motion by changes in the prosodic system of early Chamic languages are largely responsible for monosyllabisation of Eastern Cham.

In the following sections, I first treat the set of affixes attested in Cham inscriptions (9th–15th centuries) and Modern Cham manuscripts (16th–19th centuries) (Section 2), discussing the possible role of contact and learnability in its distribution (Section 3). I then describe the recent loss of affixation (Section 4), evaluating evidence that it was triggered by contact (Section 5).

2.   Affixation in Cham inscriptions[4] and manuscripts

The two stone inscriptions assumed to be the oldest documents written in Cham were found at Đông Yên Châu (C.174) and Samo (C.199) – both in Quảng Nam province – in the 1930s.[5] They were dated based on paleographic evidence: the Đông Yên Châu inscription was dated from the middle of the 4th century based

---

4.   The Corpus of the Inscriptions of Champa is available online: <http://isaw.nyu.edu/publications/inscriptions/campa/>.

5.   I follow the classification scheme established by George Coedès (1908) and still followed by Cham epigraphists (Griffiths et al. 2008). C refers to a Cham inscription. The following number was originally meant to indicate geographical location (from South to North), but now indicates the order of discovery.

on similarities with Sanskrit inscriptions found 2.5 kilometres away (Cœdès 1939), while the Samo inscription was dated from the 7th century based on the use of the virāma, a superscript diacritic that suppresses inherent vowels, to mark final consonants (Wittarayat 2004–2005). While plausible, these contextual datings should be considered with care for two reasons. First, the next earliest inscriptions in Cham date from the 9th century (Schweyer 1999). Second, there are no obvious linguistic or epigraphic differences between the Đông Yên Châu inscription and these 9th century inscriptions, despite a purported 500-year gap. There is therefore a distinct possibility that the Đông Yên Châu and Samo inscriptions are in fact several centuries posterior to the dates proposed by Cœdès and Wittarayat.

In any case, since the Đông Yên Châu and Samo inscriptions contain neither affixation nor Mon-Khmer loanwords, possibly because of their brevity (respectively 32 and 39 words, including Sanskrit loanwords), I will not dwell on them and will immediately turn to the next earliest inscription to which I have access, the inscription of Bakul-Yang Kur (C.23; Ninh Thuận province), dated 829 AD. This inscription contains a single affix, the Mon-Khmer nominalising -*an*- infix, found in the word *vanuh*. Aymonier interprets this word as a nominal derivation of *vuh*, 'to give' (Aymonier 1891).[6]

(1) *humā pralauŋ, humā padaiŋ, ney ʃaka* <u>*vanuh*</u> *humā dvā nan 751.*
field Pralon, field Paden DEM Śaka.era donation field two DEM 751

yāŋ maŋdara di parvvata. Vihāra devarakṣa di krauŋ. yāŋ
god Mandara at mountain monastery Devaraksa at river God

praŋaveʃvara di mandauh. Vihāra ney avista nan sā pu
Pranavesvara at ??? monastery DEM all DEM one master

pov puŋya.
lord good.deeds

'The fields of Pralong and Padeng. Shaka 751 (829AD) is the year of donation of these two fields. God Mandara at the mountain, Devaraksa Monastery at the river, God Pranaveçvara at [???]. All these monasteries [are given] to the God of good deeds.'

The presence of a single affix in this inscription, which is a mere list of land donations to temples and monasteries, may be due to its choppy syntax and to the fact that it contains no verbs. The Bàn Lãnh stela inscription (C.106; Quảng Nam province), dated 898 AD, is longer and contains more affixes, as illustrated by this fragment (2).

---

6. The transliteration of the Cham inscriptions presented here is based on Arlo Griffiths and Amandine Lepoutre's revisions of the original French colonial sources, with minor adaptations to IPA transcription for velars, palatals and retroflexes. The translations are as conservative as possible, leaving many unglossed words.

(2) *yāŋ pov ku   siniy <u>ranakṣā</u> yāŋ guru   pu pov ku   di yāŋ   taml*
    god lord king here protection god teacher god lord king at god(?) until

    thun pāt slauv vrāh sā   vlah   kain kamn yāŋ guru   pu pov ku
    year four ?   ?   one piece[7] cloth give(?) god teacher god lord king

    man   jmāy   top   urāŋ yāŋ niy   vāra yo tavun driy
    again/indeed(?) jewel(?) wealth(?) person god DEM ?   ?   ?   CLF

    niy   <u>kanadhā</u> <u>sanupah</u> ntrā mad(d)an ājna yāŋ pu pov ku   ʃrī
    DEM discourse oath   also there.is   order god god lord king saint

    Jaya Siŋhavarmadeva maddan urāŋ   tapah pāt driy siy vṛliy
    Jaya Singhavarmadeva with   person ascetic four CLF who whoever

    urāŋ   yamŋ <u>marakṣā</u> sarvvadravya yāŋ pov ku   ʃrī
    person god   protect all.properties god lord king saint

    Rudramadhyeʃvara tra   yāŋ pov ku   ʃrī   ʃivaliŋeʃvara   tra
    Rudramadhyeshvara also(?) god lord king saint Shivalingeshvara also(?)

    hulun lamvov kravāv humā māh pirak <u>pasrauŋ</u> ya nan   avista karaŋa
    slave cow   buffalo field gold silver support ?   DEM all   activity

    <u>panūjā</u> yāŋ pu pov ku   tra   udakāna tra   urā   nan nau
    cult   god god lord king also(?) water.food also(?) person DEM go

    svargga maddan inā   amā   aviskāla.
    heavens with   mother father forever

    'The king protected the teacher God (Shiva) in the temple for four years… a piece of valuable clothing to the teacher God… wealth and temple people… This oath is also the order of king Jaya Singhavarmmadeva to four monks. Whatever temple people guard the belongings of the divine Rudramaddhyeshvara and Shivalingeshvara (slaves, cows, buffaloes, fields, gold, silver), support the cult of the god king and provide water and food, these people will go to heavens with their mothers and fathers forever.'

                     (Loose translation adapted from Finot 1904)

In this excerpt, there are four tokens of the nominal derivative -*an*-, all in Sanskrit roots: *ranaksā* 'protection', from *raksā* 'to protect'; *kanadhā* 'discourse', from *kadhā* 'story'; *sanupah* 'oath', from *supah* 'to swear'; *panūjā* 'worship', from *pūjā* 'ceremony'. We also find instances of the causatives *ma-* and *pa-* prefixes: *maraksā* 'to guard', from *raksā* 'to keep' and *pasrauŋ* 'to provide for, to support', from *srauŋ* 'to live, to survive, be present'.

    Based on the Bàn Lãnh inscription and on the Phú Thuận stela (C.139), which is from the same period and geographical area, it seems that affixation was common and productive in the 9th century. However, its prevalence tends to decrease

---

7.   Vickery (1992)

in following centuries. The Po Klong Garai inscription (C.13; 1050 AD), located in Ninh Thuận province, only contains the *pa-* prefix, and we find the same impoverished morphology in later inscriptions, like that of the Royal Gate of Bình Định (C.47; 1401 AD). The late Biên Hoà stela (C.1; 1421 AD), found further south in Đồng Nai province, contains at most one affix with unclear semantics (Cabaton 1904; Finot 1915; Marrison 1975; Griffiths 2019).

The Vietnamese conquest of the political centre of Vijaya (Bình Định) in 1471, and the disorganisation that ensued, led to the interruption of the production of stone inscriptions. However, Pāṇḍuraṅga, a Cham polity located further south in modern day Ninh Thuận and Bình Thuận, took over and maintained a certain amount of independence until 1832 (Po 1987, 1991, 1994). Cham chronicles and manuscripts that were produced during that period allow us to follow the evolution of the language, despite a gap of more than a century between the last inscription and the first manuscripts. *Modern Cham* is the medium of these 17–19th century manuscripts and constitutes the basis of the contemporary written language. It is impossible to determine if it faithfully reflects the spoken language of that period or if it corresponds to a more conservative variety, but it is definitely different from the language of the inscriptions. Crucially, it does preserve a set of affixes similar to what is attested in stone inscriptions.

The most common Modern Cham affix is the causative *pa-*. It was mostly used with verbs (both active and stative) but could also be used with other lexical categories (Aymonier 1889; Aymonier & Cabaton 1906; Moussay 2006).

(3) Active verbs[8]

plĕj	'to buy'	*paplĕj*	'to sell'
th̆w	'to know'	*path̆w*	'to inform'
mɨtai	'to die'	*pamɨtaj*	'to kill'

Stative verbs

mɨpuʔ	'drunk'	*pamɨpuʔ*	'to get someone drunk'
ƙăm	'stuck'	*paƙăm*	'to stick'
sjam	'beautiful'	*pasjam*	'to embellish'

Others

pu	'rice porridge'	*papu*	'to make rice porridge'
j̆w	'like'	*aj̆w*	'to compare'
pier	'low'	*papier*	'to lower'

---

8. For convenience and homogeneity, Examples (3) to (8) are given in an IPA transcription of the formal contemporary reading of Modern Written Cham. The subscript circle under onset consonants represents the low register, following Moussay (1971). These transcription decisions have no bearing on the realisation of affixation.

Interestingly, *pa-* also seems to have a possible inflectional use, turning intransitive verbs into transitives (Aymonier 1889; Moussay 2006). Clear examples like *mɨm* 'to cover oneself', *pamɨm* 'to cover something or someone' are rare but note that the prefixed stative verbs listed above indirectly acquire this transitive meaning.

The second, less productive, prefix is *ma-* [mɨ]. It has a less predictable meaning ranging from causative to adjectival and can be concatenated to nouns or verbs (Aymonier 1889; Aymonier & Cabaton 1906; Moussay 2006).

(4) *pɔh* 'egg, fruit' *mɨpɔh* 'to lay eggs'
     jŭt 'friend' *mɨjŭt* 'to befriend'
     jaŋ 'god' *mɨjaŋ* 'god-like'
     tjan 'belly' *mɨtjan* 'pregnant'
     kɛ̆ʔ 'to bite' *mɨkɛ̆ʔ* 'to get angry'

A third derivational prefix, *ta-*, originally described as a frequentative, seems to be more accurately characterised as a reflex of the Proto-Malayo-Chamic inadvertent prefix *tAr (Aymonier 1889; Aymonier & Cabaton 1906; Adelaar 1992; Thurgood 1999; Moussay 2006).

(5) *k̥alŭŋ* 'to roll' *takalŭŋ* 'to roll around'
     katwă̆ʔ 'surprised' *takatwă̆ʔ* 'to be startled'
     lapuh 'to fall' *talapuh* 'to abort'

Modern Cham also preserved the productive Mon-Khmer nominalisation infix *-an-* already common in Cham inscriptions.

(6) *poc* 'to talk' *panoc* 'word'
     ʈ̥ɔʔ 'to live in' *ʈ̥anɔʔ* 'abode'
     ap 'proper' *kanap* 'good manners'
     prŏ̆ŋ 'large' *panrŏ̆ŋ* 'notable'

Two additional infixes, *-pa-* and *-ma-* [-mɨ-] are also listed in Aymonier & Cabaton (1906). The small number of forms in which they are used suggests fossilisation (note that 'food, provisions' seems to contain more than a simple infix).

(7) *ranăm* 'love' *rapanăm* 'cherish'
     raça 'ceremony' *rapaça* 'high-ranking priest'
     ɓăŋ 'to eat' *ɓamɨnɨŋ* 'food, provisions'
     karaŋ 'less' *kamɨraŋ* 'deficit'
     pa-jɔ̃w 'to compare' *pamɨjɔ̃w* 'to compare'

Finally, there is a large number of fossilised affixes that are only attested in a few forms and have more or less predictable meanings (Aymonier & Cabaton 1906:

xxiv). I will only mention a possible passive infix, described as a future tense infix in Aymonier (1889).

(8) *akʰan*   'to inform'   *anikʰan*   'to be informed'
     apăn   'to hold'   *anipăn*   'to be held'

The evidence presented in this section suggests that a small set of derivational affixes has been in use since the first stone inscriptions. Their apparent rarity in late stone inscriptions might be coincidental, as they resurface in Modern Cham manuscripts, but could also be explained by dialectal or register variation. Rather than speculating based on a limited corpus,[9] I conservatively conclude that:

1. We know that Malayo-Polynesian inflectional affixation had been lost, and that the Mon-Khmer nominalising affix had been adopted, by the time of the first inscriptions. Unfortunately, there is no evidence about the linguistic history of Cham before the 4th century at the earliest (the conventional date of the Đông Yên Châu inscription).
2. There is no reliable evidence of change in the prevalence or productivity of Cham affixation from the 9th to the 19th century.

3. Contact, learnability and the initial reduction of affixation

If we accept the apparently uncontroversial claim that the Chams are the descendants of the Sa Huỳnh archaeological culture, Chamic speakers would have settled in central Vietnam by 600 BC (Bellwood 1985; Bronson & White 1992). We also know that at some point in their early history, Chamic populations were in contact and mixed with Mon-Khmer languages. The evidence for this early contact is both linguistic and genetic. First of all, the language of the first firmly dated inscriptions, in the 9th century, already contains a significant proportion of Mon-Khmer words (the Bakul-Yang Kur inscription in (1) above contains the MK word *krong* 'river' and the MK *-an-* infix). Second, mitochondrial and Y-chromosome DNA suggests that a significant proportion of both the maternal and paternal gene pools of the Chams is Mainland Southeast Asian rather than insular (Peng et al. 2010; He et al. 2012). Note that this does not necessarily suggest intensive contact: a constant 0.5% rate of intermarriage over 2500 years would have an enormous genetic impact.

Unfortunately, the exact nature and extent of contact between the Cham settlement on the Mainland and the first stone inscriptions is largely open to speculation.

---

9. There is a total of 237 Cham inscriptions, more than half of which are in Cham (Arlo Griffiths, p.c.). However, most of these inscriptions are very short.

One could imagine that the newcomers were in contact with Mon-Khmer populations from their very arrival or that contact occurred much later. Chamic speakers could have landed in a sparsely populated area or could have been in a minority position from the start. Periods of isolation and contact could have alternated, or contact could have been constant, but relatively moderate.

We barely have more evidence about the classical period. While it is now accepted that classical Champa was a multi-ethnic polity, including Coastal Chamic, Highlands Chamic and Mon-Khmer groups (Hickey 1981; Gay 1994; Népote 2004; Shine 2007; Hardy & Nguyễn 2019), we do not really know if contact was significant enough to trigger language change. As Mon-Khmer groups were mostly living in the highlands of the Annamite Cordillera, which are peripheral and cannot support the same population as the coastal lowlands, one could wonder if they had the demographic weight necessary to trigger major structural change in Lowland Chamic during that period. It is possible that there were Mon-Khmer speaking populations in coastal Chamic polities (as Hardy & Nguyễn 2019 have suggested for Quảng Ngãi), but even if this was the case, it would probably have been limited to specific areas; there are very few speakers of Mon-Khmer languages near the current Cham-speaking areas of coastal south-central Vietnam (except Vietnamese, of course). We can probably assume more intense contact between Highland Chamic languages and Mon-Khmer, but this does not necessarily entail language shift. In fact, reconstruction of Bahnaric and Katuic suggests that borrowing might have gone from Chamic to Mon-Khmer much more than in the opposite direction (Sidwell 2007, 2008).

In light of this limited evidence, let us now examine possible scenarios for the loss of affixation in Chamic. The first plausible account is that an initial simplification of affixation was caused by an early influx of Mon-Khmer second-language speakers. Thurgood (2000) casts this in terms of *learnability*: the language became simpler in order to accommodate new speakers. In order to avoid the teleological implications of the concept of learnability, we could reframe this by proposing that at some point in its history, Cham or its ancestor was a Non-Hybridised Conventionalised Second Language (McWhorter 2008), a language whose grammatical structures were severely simplified by an influx of untutored adult learners.

Obviously, one cannot claim that the overall structures of modern Chamic languages are simpler than those of Austronesian languages in some absolute sense. Some aspects of Chamic grammars are objectively more complex than their Malayo-Polynesian counterparts. The integration of the Mon-Khmer nominalising -*an*- infix, that is even generalised to Sanskrit loanwords in stone inscriptions, is certainly not a simplification. Moreover, the phonemic inventory of Proto-Chamic, that includes central vowels, diphthongs, implosives and aspirated stops (Lee 1966; Thurgood 1999), is far more complex than that of its ancestor,

suggesting that contact did not only lead to simplification. Nonetheless, it is possible that Malayo-Polynesian affixation was lost because it was difficult to acquire for Mon-Khmer speakers. If a massive number of Mon-Khmer speakers shifted to Ancient Cham (or Chamic languages), a large proportion of the language community might have failed to master some affixes, preventing them from being passed on to the following generation.

Since the reduction of affixation occurred before the first stone inscriptions, this scenario would entail intensive contact and a massive shift of Mon-Khmer speakers to Cham at some point between the Chamic settlement of Mainland Southeast Asia and the 9th century. That said, a long period of sustained contact is not required (there is in fact no positive evidence for it) and we may want to qualify the received idea that "Cham has undergone two thousand years of unending, unrelenting language contact" (Thurgood 2000).

However, there is also a less circumstantial problem with the learnability and the Non-Hybridised Conventionalised Second Language scenarios: simplification of grammatical structures is not necessarily caused by intensive contact (Gil 2001, 2007, 2008 on Malayic). For example, Middle Vietnamese shared with Cham all the learnable grammatical structures listed in Thurgood (2000), although there is no evidence of large-scale language shift towards Vietnamese before the 17th century, except from closely related varieties (Phan 2012). Furthermore, most Eastern Mainland Southeast Asian languages exhibit the highly learnable structures attributed by Thurgood to an influx of new adult speakers, even if, in most cases, there is no evidence that they underwent massive language shifts. Language contact and shift are therefore possible, but not necessary conditions for the development of simple (or learnable) structures. We thus cannot be certain that the simplification of Chamic affixation was due to contact. While intensive contact with Mon-Khmer and a massive language shift from Mon-Khmer to Chamic is a perfectly reasonable story, we must also consider the possibility that morphological simplification happened before contact with Mon-Khmer languages, or during a period of limited contact.

4. Iambicity, monosyllabisation and the loss of affixation

The second stage in the loss of Cham affixation seems to have taken place since the 19th century. It is a side effect of a process of monosyllabisation that is quasi-complete in colloquial Eastern Cham but is also more marginally attested in the Western Cham dialects spoken in Cambodia and the Vietnamese Mekong Delta. We can consider monosyllabisation as the last step in a sequence of phonological processes that started with a shift to iambic stress in Proto-Chamic (Lee 1966; Thurgood 1999). As Proto-Chamic was largely disyllabic, iambisation led to

a lexicon mostly composed of monosyllables and disyllables with a weak-strong rhythm. The subsequent phonetic reduction of initial unstressed syllables in turn led to a neutralisation of their phonological contrasts and changed disyllables into *sesquisyllables*, i.e. disyllables composed of a weak and reduced initial syllable, the *presyllable*, and of a heavy unreduced final syllable, the *main syllable* (Matisoff 1973). It is difficult to date iambisation and sesquisyllabisation precisely, but a certain instability in the spelling of the vowels of non-final syllables in inscriptions suggests that they may have occurred early (Arlo Griffiths p.c.).[10]

We know that a small subset of Chamic sesquisyllables became monosyllabic early on. Thurgood (1999) reconstructs the formation of aspirated stops, stop-liquid clusters and implosive stops as the result of the loss of presyllabic vowels in Proto-Chamic. Moreover, the tendency to monosyllabisation seems to have lingered and continued to manifest itself in Cham inscriptions and manuscripts: an example is the reduction of the form *marai* 'to come', common in inscriptions, to *mai* in Modern Cham. Nevertheless, most presyllables were still preserved in Modern Cham manuscripts. The first signs of a more radical monosyllabisation were noticed in Aymonier (1889: 39):

> Même lorsqu'il n'y a pas à craindre la confusion, non seulement la première syllabe varie, mais encore elle est supprimée. On peut lire dans certains cas, par exemple: kok pour akok, tête, rau pour arau, laver le linge, nẽi pour moenẽi, se baigner, vẽi pour havẽi, rotin, etc.

> Even when there is no possible confusion, not only does the first syllable vary, but it is also deleted. We can read in some cases, for example: kok for akok, head, rau for arau, wash clothes, nẽi for moenẽi, bathe, vẽi for havẽi, rattan, etc.

> (My translation)

Aymonier's observation is confirmed in post-colonial sources (Lee 1966; Alieva 1991, 1994; Bùi 1996; Brunelle 2005, 2008b, 2009). An Eastern Cham publication (in Vietnamese) from the early 1970s is quite explicit about its progress:

> Xưa kia người Chàm nói đầy-đủ cả hai vần trong mỗi tiếng, nhưng ngày này thường bớt vần-phụ mà chỉ nói vần-chính, khiến nhiều khi sinh ra sự lẫn nghĩa. Vì không được nói đến, vần-phụ thường bị quên hay bị nói sai đi. (Trung-tâm Văn-hoá Chàm 197?: 10)

> In the past, the Cham pronounced both syllables of each word, but nowadays, they usually reduce the presyllable and only pronounce the main syllable, which often causes semantic confusion. Because they are not pronounced, presyllables are usually omitted or rendered incorrectly. (My translation)

---

10. More crosstalk between proto-Chamic reconstruction and Cham philology, two disciplines that have so far conspicuously ignored each other, should provide straightforward answers.

Contemporary Eastern Cham (including Phan Rang Cham) has developed a situation of symbolic diglossia (Brunelle 2008a, 2009). The colloquial variety of the language (L) has become almost entirely monosyllabic, coexisting with a formal variety (H), which preserves sesquisyllables (and is generally more conservative), but is never used in daily conversation. In the following examples, monosyllabic words are compared with their more conservative, sesquisyllabic variants.

(9) H variety[11] Gloss L variety
    akʰăr 'word, script' *kʰăn*
    tapa 'to cross' *pa*
    rilo 'many, a lot' *lo ~ klo*
    palăj 'village' *plĕj ~ mlĕj*
    caḷan 'road' *ḳlan*

The exact outcome of monosyllabisation (loss of presyllable, creation of an onset cluster) largely depends on syllabic structure and the sonority of onset consonants (Brunelle 2008b, 2009), but there is also variation in the community, some of it along gender lines (Blood 1961). What matters here is that monosyllabisation erodes prefixes and infixes in the same way as tautomorphemic sesquisyllables.

(10) H variety Morphology Gloss L variety
    pa-plăj CAUS+buy 'to sell' *plĕj* (homoph. with 'to buy')
    pa-nɨʔ CAUS+child 'to give birth' *mnɨʔ*
    pa-rapʰa CAUS+part 'to split, divide' *pʰa*
    mɨ-ḳru CAUS+teacher 'to teach' *ḳru*
    p-an-oc talk+NMLZ 'word, speech' *mnojʔ ~ nojʔ*

As a result of monosyllabisation, affixation is either entirely lost or no longer transparent in colloquial Eastern Cham (L). Despite occasional phonological and semantic similarities between the reflexes of the roots and affixed forms (*nɨʔ* 'child' ~ *mnɨʔ* 'to give birth'; *pojʔ* 'to talk'~ *mnojʔ* 'word'), it is unlikely that speakers unfamiliar with the formal variety (H), who make up the majority of the community, are aware of the fossilised remnants of affixes. Derivational morphology has therefore mostly been replaced with periphrastic devices:

---

11. Modern Eastern Cham is given in IPA. Following Moussay (1971), the subscript dot under consonants represents the low register, which is a combination of low pitch, breathy voice and longer duration phonologically associated to the onset, but phonetically realised on the entire syllable.

(11)	Anc. Cham	Morphology	Gloss	L variety	Literal rendering
	pa-băŋ	CAUS+eat	'feed'	*prĕj băŋ*	'give eat'
	pa-siam	CAUS+beautiful	'embellish'	*ŋăʔ sam*	'make beautiful'
	r-an-aksa	protect+NMLZ	'protection'	*prŭʔ khĭʔ*	'task guard'

In effect, the colloquial variety of Modern Eastern Cham has lost its remaining derivational affixation and has become entirely analytic.

5.   The loss of affixation: Contact versus internal restructuring

Can the loss of affixation in Modern Eastern Cham be attributed to contact? The nature of relations between Cham and neighbouring languages has changed dramatically in the 19th century. While Cham seems to have been in a dominant position for most of its history, the Vietnamese conquest and the gradual marginalisation of the vassal state of Pāṇḍuraṅga confined it to a minority language status (Po 1987, 1994). During that period, relations with other minorities became more limited than before. While there are mentions of Raglai and Koho groups in Modern Cham manuscripts (Po 1987; Shine 2007) and while regular contact with Raglai speakers continues to this day, these interactions are unlikely to have had a significant effect on Cham because these groups have little linguistic and cultural prestige and are not shifting to Cham. Monosyllabisation and the loss of affixation thus cannot be attributed to learnability or to a simplification caused by an influx of second language learners.

The major change in the linguistic landscape of Cham communities since the 19th century is that Vietnamese gradually became their main language for communicating with the outside world and government institutions (Brunelle 2008a). We can infer that intensive contact with Vietnamese started at the very latest in the 1940s, from the fact that almost all Eastern Chams now speak Vietnamese fluently (only a handful of elderly women are still monolingual), but it is likely that widespread bilingualism started much earlier, at least in some social groups. It is therefore tempting to view bilingualism in monosyllabic Vietnamese as the motivation for monosyllabisation (Thurgood 1999; Grant 2005). Unfortunately, the claim that Vietnamese is monosyllabic is a serious oversimplification: about half of the Vietnamese lexicon is composed of disyllabic (or even trisyllabic) words, including opaque Sino-Vietnamese compounds and monomorphemic ideophones and loanwords (Nguyễn 1997; Trần & Vallée 2009; Brunelle 2015, 2017). Moreover, no causal mechanism easily explains how contact with Vietnamese could have led to monosyllabisation. An influx of second language speakers cannot be invoked as there is no evidence of any significant shift from Vietnamese into Cham since

the 19th century. Semilingualism is not a possible explanation either, as the vast majority of Eastern Cham speakers are dominant in their native language; a few individuals who grew up in towns and cities are dominant in Vietnamese, but typically, they only maintain occasional contact with their parents' villages and therefore have little impact on the rest of the community. Even from a strictly linguistic point of view, attributing monosyllabisation to bilingualism in Vietnamese poses a problem: while it is possible for bilinguals to add new features (lexical items, phonemes, morphemes) to their L1 under L2 influence, it would be rather peculiar for Cham speakers to drop L1 grammatical features (like sesquisyllables and affixes) in an attempt to make their L1 more similar to their L2 (i.e. monosyllabic).

The solution seems to lie in the fact that Cham was already moving along a continuum ranging from iambicity to monosyllabicity before contact with Vietnamese. The idea of such a continuum is not new (Matisoff 1973), but it can be formalised by proposing that disyllabic iambs have a natural tendency towards monosyllabic reduction for two reasons: a phonetic pressure to reduce unstressed syllables and some basic properties of the Iambic-Trochaic Law (Brunelle & Pittayaporn 2012).

The Iambic-Trochaic Law (Hayes 1985) captures an important asymmetry in the relation between types of prominence and their alignment. Based on cross-linguistic evidence, it states that two types of natural rhythmic groupings, grounded in either psycho-acoustic or cognitive universal properties (Hayes 1985; Hay & Diehl 2007 for instance), rule the metrical properties of languages.

(12)   Iambic-Trochaic Law
    a.   Elements contrasting in intensity naturally form groupings with initial prominence (trochees).
    b.   Elements contrasting in duration naturally form groupings with final prominence (iambs).

As Chamic is iambic, I will concentrate on the second part of the Law. In accordance with (12b), the hierarchy in (13) should govern iambic well-formedness (where L = light syllable, H = heavy syllable):

(13)   L.H, H > L.L

In plain words, feet composed of a light and a heavy syllable or of one heavy monosyllable are better iambic feet than feet composed of two light syllables, that do not exhibit final prominence (single light syllables are dispreferred because they are not binary and would form degenerate feet). Since the two word shapes L.H and H are equally acceptable, Brunelle and Pittayaporn (2012) propose that the natural phonetic tendency to erode unstressed syllables will not be barred by the iambic well-formedness hierarchy in (13). Importantly, phonetic erosion is not

obligatory, and even if it applies, it can be active over short periods of time and then be dormant for centuries. What is crucial is that once a disyllable becomes iambic, it has to follow the diachronic path in (14) in which there are only two options: status quo or reduction of the initial syllable (note that the language could also lose stress altogether).

(14)

Disyllabic Iamb	Sesquisyllabic	Monosyllabic
$\sigma \, \sigma \rightarrow$	$\partial \, \sigma \rightarrow$	σ
$\mu.\mu\mu$	$^{?}.\mu\mu$	$\mu\mu$

A disyllabic iamb can neither reduce its stressed syllable, which would make it a worse iamb (13), nor turn into a trochee, as L.H is not a possible trochee.[12] Similarly, a sesquisyllable can only stall or further reduce its presyllable: it has no phonetic or phonological motivation for weight augmentation. Thus, unless phonological material is added by compounding or by a new wave of affixation, sesquisyllables cannot expand. Rather than seeing the shift from iambs to monosyllables as the manifestation of a constant pressure towards reduction in the language, we can thus conceive it as the only possible, albeit contingent, path of change.

Practically, how does this help us to characterise monosyllabisation in Eastern Cham? By the time of the first inscriptions, an early process of monosyllabisation had already affected a small subset of words (leading to the development of aspirated stops, stop-liquid clusters and implosive stops as described in Thurgood 1999). Phonetic reduction then paused for at least a millennium (9th–19th centuries), possibly because of the functional role of the affixes that had survived the initial wave of morphological reduction, that would have been a necessary or preferred way of expressing causation and nominalisation and would have been wiped out by monosyllabisation. The problem here is to explain why monosyllabisation resumed in the 19th century. I propose that contact in Vietnamese may have played a role, but in an indirect manner.

As mentioned above, it is unlikely that Eastern Cham speakers started making their native language monosyllabic in order to make it more similar to Vietnamese. Yet, a more complex contact-based explanation for the loss of affixation is possible. Bilingualism in Vietnamese may have led the Chams to calque periphrastic Vietnamese forms to express the grammatical functions previously expressed by affixes in their language (see 11) or may simply have reinforced pre-existing

---

12. An anonymous reviewer pointed out that Munda languages are claimed to have shifted from iambic to trochaic rhythm (Donegan & Stampe 2002, 2004; Sidwell & Rau 2015). I would like to direct the reader to recent work that challenges this claim and provides experimental evidence that Munda languages are mostly iambic (Horo & Sarmah 2015; Ring & Anderson 2018). I am therefore unaware of clear cases of a direct shift from iambs to trochees.

periphrastic grammatical devices. At the turn of the century, these new syntactic strategies would have gradually started to spread and to compete with derivational affixes. If, as suggested above, the preservation of affixation was holding the phonetic reduction of unstressed presyllables at bay, this would have allowed radical monosyllabisation. There would thus be no direct relation between the alleged monosyllabicity of Vietnamese and monosyllabisation of Eastern Cham. It is the periphrastic strategies of Vietnamese that would have opened the door to the loss of affixation rather than its monosyllabic structures.

This contact scenario is admittedly speculative: as we have limited data on late 19th century-early 20th century Eastern Cham, it is unlikely that evidence allowing us to determine whether this scenario actually took place will ever surface. Nonetheless, it provides better causal mechanisms than an account merely relying on superficial similarities between languages: the calquing of periphrastic devices in a socially dominant L2 (or the reinforcement of pre-existing periphrastic strategies by parallel structures in the L2) could open the door to morphological simplification. This scenario would also explain why monosyllabisation is less widespread in the Western Cham dialect continuum of the Vietnamese Mekong delta and Cambodia. Western Cham communities are much less integrated in the Vietnamese and Cambodian states than their Eastern Cham cousins (Collins 1996; Nakamura 1999; De Féo 2004; Taylor 2007; Bredenberg 2008) and tend to be less highly bilingual. As such, they are less likely to calque the periphrastic structures of Vietnamese and Khmer. In the end, regardless of the accuracy of this account, the important point is that contact-induced change is not limited to mere imitation; changes in a given area of grammar can easily end up affecting another. Contact-based explanations for language change should therefore take entire grammars into account rather than merely comparing parallel sub-systems.

6.   Conclusion: Revisiting contact and learnability

After re-examining linguistic evidence for the loss of affixation in Chamic, I propose to revise previous accounts of contact with Mon-Khmer and its effect on Austronesian affixation (Alieva 1984, 1991; Thurgood 1999, 2000; Grant 2005, 2007).

Although contact with Mon-Khmer clearly affected early Chamic, there is little historical, genetic, linguistic or archaeological evidence of long-term intensive contact between Mon-Khmer and Chamic languages or populations. This does not mean that there was no such contact, but rather that other forms of contact (intensive and short, constant but diffuse) are also possible. If the loss of Western-Malayo-Polynesian affixation was caused by a massive influx of L2 speakers as proposed by Thurgood (2000), this influx must have occurred before 9th century inscriptions.

However, we must also consider the possibility that this simplification may have occurred before contact with Mon-Khmer, or even without a massive influx of untutored adult L2 learners.

The second wave of morphological simplification, caused by the monosyllabisation of Eastern Cham, likely occurred since the 19th century, at a time where intensive contact with Vietnamese is well attested. Nevertheless, a mechanistic scenario in which Eastern Cham speakers would have made their L1 monosyllabic through direct imitation with Vietnamese is unlikely. Instead I propose that monosyllabisation was driven by internal phonetic and phonological pressures, and speculate that if contact with Vietnamese favoured Eastern Cham monosyllabisation, it happened in indirect ways.

Acknowledgements

References

Adelaar, K. Alexander. 1992. *Proto-Malayic, the Reconstruction of its Phonology and Parts of its Lexicon and Morphology* [Pacific Linguistics 119]. Canberra: ANU.

Alieva, Natalia F. 1984. A language-union in Indo-China. *Asian and African Studies* XX: 11–22.

Alieva, Natalia F. 1991. Morphemes in contemporary spoken Cham: Qualitative and quantitative alternations. *Cahiers de Linguistique Asie Orientale* 20: 219–229.

Alieva, Natalia F. 1994. The progress of monosyllabization in Cham as testified by field materials. In *Proceedings of the Seventh International Conference on Austronesian Linguistics (ICAL)*, Cecilia Ode & Wim Stokhof (eds), 541–549. Amsterdam: Rodopi.

Aymonier, Étienne. 1891. Première étude sur les inscriptions chames. *Journal Asiatique* 17: 5–86.

Aymonier, Étienne François. 1889. *Grammaire de la langue chame*. Saigon: Imprimerie Coloniale.

Aymonier, Étienne François & Cabaton, Antoine. 1906. *Dictionnaire cam-francais*. Paris: E. Leroux.

Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian Archipelago*. Orlando FL: Academic Press.

Blood, Doris W. 1961. Women's speech characteristics in Cham. *Asian Culture* 3: 139–143.

Bredenberg, Kurt. 2008. Educational marginalization of Cham Muslim populations: A report from Cambodia. *Journal of Education for International Development* 3(3): 1–26.

Bronson, Bennet & White, Joyce. 1992. Radiocarbon and chronology in Southeast Asia. In *Chronologies in Old World Archaeology*, Vols. 1 & 2. Robert W. Ehrich (ed.), 491–503. Chicago IL: The University of Chicago Press.

Brunelle, Marc. 2005. Register in Eastern Cham: Phonological, Phonetic and Sociolinguistic Approaches. PhD dissertation, Cornell University.

Brunelle, Marc. 2008a. Diglossia, bilingualism, and the revitalization of written Eastern Cham. *Language Documentation and Conservation* 2: 28–46.

Brunelle, Marc. 2008b. Monosyllabicization in Eastern Cham. In *Papers from the 14th annual meeting of the Southeast Asian Linguistics Society 2004*, Wilaiwan Khanittanan & Paul Sidwell (eds), 43–58. Bangkok: Pacific Linguistics.

Brunelle, Marc. 2009. Diglossia and monosyllabization in Eastern Cham: A sociolinguistic study. In *Variation in Indigenous Minority Languages* [Impact: Studies in Language and Society 25], James Stanford & Dennis Preston (eds), 47–75. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.25.04bru

Brunelle, Marc. 2015. Vietnamese (Tiếng Việt). In *The Handbook of Austroasiatic Languages*, Mathias Jenny & Paul Sidwell (eds), 909–954. Leiden: Brill.

Brunelle, Marc. 2017. Stress and phrasal prominence in tone languages: The case of Southern Vietnamese. *Journal of the International Phonetic Association* 47: 283–320. https://doi.org/10.1017/S0025100316000402

Brunelle, Marc. 2019. Revisiting the expansion of the Chamic language family. In *Champa: Territories and Networks of a Southeast Asian Kingdom*. Arlo Griffiths, Andrew Hardy & Geoff Wade (eds), 287–302. Paris: Presses de l' École française d'Extrême-Orient.

Brunelle, Marc & Pittayaporn, Pittayawat. 2012. Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia. *Diachronica* 29: 411–33. https://doi.org/10.1075/dia.29.4.01bru

Bùi, Khánh Thế. 1996. *Ngữ Pháp Tiếng Chăm*. Hà Nội: Nhà Xuất Bản Giáo Dục.

Cabaton, Antoine. 1904. L'inscription chame de Bien Hoa. *Bulletin de l'École Française d'Extrême-Orient* 4: 116–19.

Coedès, George. 1908. Inventaire des inscriptions du Champa et du Cambodge. *Bulletin de l'École Française d'Extrême-Orient* 8: 37–92. https://doi.org/10.3406/befeo.1908.4180

Coedès, Georges. 1939. La plus ancienne inscription en langue cham (Inscription rupestre de Đông Yên Châu, province de Quảng Nam, Annam). In *New Indian Antiquary: A Volume of Eastern and Indian Studies in Honour of F.W. Thomas, C.I.E.*, Sadashiva M. Katre & Parashuram K. Gode (eds), 46–49. Bombay: Karnatak Publishing House.

Collins, William. 1996. *The Chams of Cambodia*. Phnom Penh: Center for Advanced Studies.

Cowan, Hendrik Karel Jan. 1991. Achehnese dialects in connection with Chamic migrations. In *Western Austronesian and Contact Languages: Papers from the Fifth International Conference on Austronesian Linguistics*, Ray Harlow (ed.), 53–83. Auckland: Linguistic Society of New Zealand.

De Féo, Agnès. 2004. *Les Chams, l'Islam et la revendication identitaire: Des origines de l'islamisation au radicalisme islamique actuel*. Mémoire de DEA. Paris: École Pratique des Hautes Études.

Donegan, Patricia & Stampe, David. 2002. South-East Asian features in the Munda languages: Evidence for the analytic-to-synthetic drift of Munda. In *Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistic Society*, Patrick Chew (ed.), 111–120. Berkeley CA: BLS.

Donegan, Patricia & Stampe, David. 2004. Rhythm and the synthetic drift of Munda. In *The Yearbook of South Asian Languages and Linguistics*, Rajendra Singh (ed.), 3–36. Thousand Oaks CA: Sage. https://doi.org/10.1515/9783110179897.3

Durie, Mark. 1990. *Proto-Chamic and Achenese mid vowels: Towards Proto-Aceh-Chamic*. Ms, University of Melbourne.

Finot, Louis. 1904. Notes d'épigraphie VI: Inscriptions du Quảng Nam. *Bulletin de l'École Française d'Extrême-Orient* 4: 83–115. https://doi.org/10.3406/befeo.1904.1296

Finot, Louis. 1915. Notes d'épigraphie XIV: Les inscriptions du musée de Hanoi. *Bulletin de l'École Française d'Extrême-Orient* 15: 1–38. https://doi.org/10.3406/befeo.1915.5229

Gay, Bernard. 1994. New pperspectives on the ethnic composition of Champa. In *Proceedings of the Seminar on Champa*, Đinh T. Huỳnh (ed.), 43–52. Rancho Cordova CA: Southeast Asia Community Resource Center.

Gil, David. 2001. Creoles, complexity, and Riau Indonesian. *Linguistic Typology* 2–3: 325–371.

Gil, David. 2007. Creoles, complexity and associational semantics. In *Deconstructing Creoles* [Typological Studies in Language 73], Umberto Ansaldo, Stephen Matthews & Lisa Lim (eds), 67–97. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.73.06gil

Gil, David. 2008. Why Malay/Indonesian undressed: Contact, geography, and the roll of the dice. Paper presented at the International Symposium on Malay/Indonesian Linguistics, Leiden.

Grant, Anthony. 2005. The effects of intimate multidirectional linguistic contact: The case(s) of Chamic languages. In *Chamic and Beyond: Studies in Mainloand Austronesian Languages*, Anthony Grant & Paul Sidwell (eds). 37–104. Canberra: Pacific Linguistics.

Grant, Anthony. 2007. Admixture and after: The Chamic languages and the Creole Prototype. In *Deconstructing Creoles* [Typological Studies in Language 73], Umberto Ansaldo, Stephen Matthews & Lisa Lim (eds), 109–39. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.73.07gra

Griffiths, Arlo, Lepoutre, Amandine, Southworth, William A. & Phân, Thành. 2008. Études du corpus des inscriptions du Campa III Épigraphie du Campa 2009–2010: Prospection sur le terrain, production d'estampages, supplément à l'inventaire. *Bulletin de l'École Française d'Extrême-Orient* 95: 435–97. https://doi.org/10.3406/befeo.2008.6118

Griffiths, Arlo. 2019. Epigraphical texts and sculptural stelae produced under the Virabhadra-varmadevas of fifteen century Campā. In *Champa: Territories and Networks of a Southeast Asian Kingdom*, Arlo Griffiths, Andrew Hardy & Geoff Wade (eds), 193–220. Paris: Presses de l'École Française d'Extrême-Orient.

Hardy, Andrew & Nguyễn, Tiến Đông. 2019. Multi-disciplinary evidence for a hypothesis on the Champa territory and population of Quảng Ngãi. In *Champa: Territories and Networks of a Southeast Asian Kingdom*, Arlo Griffiths, Andrew Hardy & Geoff Wade (eds), 121–143. Paris: Presses de l'École française d'Extrême-Orient.

Hay, Jessica S. F. & Diehl, Randy. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics* 69: 113–22. https://doi.org/10.3758/BF03194458

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. In *Proceedings of the XIth Annual Meeting of the Berkeley Linguistics Society*, Mary Niepokuj, Mary VanClay, Vassiliki Nikiforidou & Deborah Feder (eds), 429–446. Berkeley CA: BLS.

He, Jun-Dong, Peng, Min-Sheng, Quang, Huy Ho, Dang, Khoa Pham, Trieu, An Vu, Wu, Shi-Fang, Jin, Jie-Qiong, Murphy, Robert W., Yao, Yong-Gang & Zhang, Ya-Ping. 2012. Patrilineal perspective on the Austronesian diffusion in mainland Southeast Asia. *PLOS ONE* 7.e36437.

Hickey, Gerald. 1981. *Sons of the Mountains: Ethnohistory of the Vietnamese Central Highlands to 1954*. New Haven CT: Yale University Press.

Horo, Luke & Sarmah, Priyankoo. 2015. Acoustic analysis of vowels in Assam Sora. *North East Indian Linguistics* 7: 69–88.

Lee, Ernest W. 1966. Proto-Chamic Phonologic Word and Vocabulary. PhD dissertation. University of Indiana, Bloomington.

Marrison, Geoffrey E. 1975. The early Cham language, and its relationship to Malay. *Journal of the Malaysian Branch of the Royal Asiatic Society* 48: 52–59.

Matisoff, James. 1973. Tonogenesis in Southeast Asia. In *Consonant Types and Tone*, Larry M. Hyman (ed.), 71–96. Los Angeles CA: USC.

McWhorter, John. 2008. Why does a language undress? Strange cases in Indonesia. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94], Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 167–190. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.12mcw

Moussay, Gérard. 1971. *Dictionnaire cam-vietnamien-français*. Phan Rang: Trung-tâm Văn-hoá Chăm.

Moussay, Gérard. 2006. *Grammaire de la langue Cam*. Paris: Les Indes Savantes.

Nakamura, Rie. 1999. Cham in Vietnam: Dynamics of Ethnicity. PhD dissertation, University of Washington, Seattle.

Népote, Jacques. 2004. *Entre hinterland et réseaux maritimes: Élements d'une histoire du Champa de la fin de la Préhistoire à nos jours avec un développement particulier pour la période comprise entre le XVe et le XVIIIe siècles*. Ms.

Nguyễn, Đình-Hoà. 1997. *Vietnamese* [London Oriental and African Language Library 9]. Amsterdam: John Benjamins. https://doi.org/10.1075/loall.9

Peng, Min-Sheng, Ho, Quang Huy, Pham, Dang Khoa, Trieu, An Vu, Wang, Hua-Wei, Yao, Yong-Gang, Kong, Qing-Peng & Zhang, Ya-Ping. 2010. Tracing the Austronesian footprint in mainland Southeast Asia: A perspective from mitochondrial DNA. *Molecular Biology and Evoulution* 27: 2417–2430. https://doi.org/10.1093/molbev/msq131

Phan, John. 2012. Mường is not a subgroup: Phonological evidence for a paraphyletic taxon in the Viet-Muong sub-family. *Mon-Khmer Studies* 40: 1–18.

Po, Dharma. 1987. *Le Panduranga (Campa) 1802–1835*. Paris: Ecole Française d'Extrême-Orient.

Po, Dharma. 1991. Le déclin du Campa entre le XVIe et le XIXe siècle. In *Le Campa et le Monde Malais*, 47–64. Paris: Publications du Centre d'Histoire et Civilisations de la Péninsule Indochinoise.

Po, Dharma. 1994. Status of the research on the data of absorption of Champa by Vietnam. In *Proceedings of the Seminar on Champa*, Đinh T. Huỳnh (ed.), 53–64. Rancho Cordova CA: Southeast Asia Community Research Center.

Reid, Lawrence. 1994. Morphological evidence for Austric. *Oceanic Linguistics* 33: 323–44. https://doi.org/10.2307/3623132

Reid, Lawrence. 2005. The current status of Austric. In *The Peopling of East Asia: Putting Together Archeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 132–60. London: Routledge Curzon. https://doi.org/10.4324/9780203343685_chapter_8

Ring, Hiram & Anderson, Gregory. 2018. On prosodic structures in Austroasiatic diachrony: 'Rhythmic holism' revisited in light of preliminary acoustic studies. In *Papers from the Seventh International Conference on Austroasiatic Linguistics*, Hiram Ring & Felix Rau (eds), 1–35. Honolulu HI: University of Hawai'i Press.

Schweyer, Anne-Valérie. 1999. Chronologie des inscriptions publiées du Campā: Études d'épigraphie cam I. *Bulletin de l'École Française d'Extrême-Orient* 86: 321–344. https://doi.org/10.3406/befeo.1999.3415

Shine, Toshihiko. 2007. Montagnards and the Cham kings: Labor and land administration as seen in the documentary and oral Archives. Paper presented at the conference *Modernities and dynamics of tradition in Vietnam: Anthropological Approaches*, Bình Châu.

Shorto, Harry. 1975. Achinese and mainland Austronesian. *Bulletin of the School of Oriental and African Studies* 38: 81–102. https://doi.org/10.1017/S0041977X00047042

Sidwell, Paul. 2006. Dating the separation of Acehnese and Chamic by etymological analysis of the Aceh-Chamic lexicon *Mon-Khmer Studies* 36: 187–206.

Sidwell, Paul. 2007. The Mon-Khmer substrate in Chamic: Chamic, Bahnaric and Katuic contact. In *Papers from the 12th Annual Meeting of the Southeast Asian Linguistics Society 2002*, Ratree Wayland, John Hartmann & Paul Sidwell (eds), 113–118. Canberra: Pacific Linguistics.

Sidwell, Paul. 2008. The origins of the Chamic lexicon: Lexification and back-borrowing into Mon-Khmer languages. In *Papers from the 14th Annual Meeting of the Southeast Asian Linguistics Society 2004*, Wilaiwan Khanittanan & Paul Sidwell (eds), 261–267. Canberra: Pacific Linguistics.

Sidwell, Paul & Rau, Felix. 2015. Austroasiatic comparative-historical reconstruction: An overview. In *The Handbook of Austroasiatic Languages*, Vol. 1, Paul Sidwell & Mathias Jenny (eds), 221–363. Leiden: Brill. https://doi.org/10.1163/9789004283572

Taylor, Phillip. 2007. *Cham Muslims of the Mekong Delta*. Honolulu HI: University of Hawai'i' Press.

Thạch, Ngọc Minh. 1999. Monosyllabization in Kiengiang Khmer. *Mon-Khmer Studies* 29: 81–95.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*. Honolulu HI: University of Hawai'i Press.

Thurgood, Graham. 2000. Learnability and direction of convergence in Cham: The effects of long-term contact on linguistic structures. In *Proceedings of the Western Conference On Linguistics, WECOL 2000*, 507–527. Fresno CA: California State University.

Trần, Thị Thúy Hiên & Vallée, Nathalie. 2009. An acoustic study of interword consonant sequences in Vietnamese. *Journal of the Southeast Asian Linguistics Society* 1: 231–249.

Trung-tâm Văn-hoá Chàm (Cham Cultural Center). 197? Số Đặc-biệt về Ngôn-ngữ Chàm (Special issue on the Cham language). *Roh twah sưu-tầm* (Collections), 28 pp. Phan Rang: Trung-tâm Văn-hoá Chàm (Cham Cultural Center).

Vickery, Michael. 1992. Evidence for prehistoric Austronesian-Khmer contact and linguistic borrowing. *Mon-Khmer Studies* 21: 185–89.

Wittayarat, Daoruang. 2004–2005. Les inscriptions rupestres de Samo (inédites): Une tentative de déchiffrement, de traduction et de datation. *Lettre de la Société des Amis du Champa Ancien* 11: 14–17.

Dual heritage

The story of Riau Indonesian and its relatives

David Gil

Max Planck Institute for the Science of Human History

How and why did Riau Indonesian acquire its isolating profile? Its isolating structure may traced back continuously through a series of increasingly large language networks – Malay/Indonesian koinés, Malay/Indonesian in general, Malayic, Western Nusantara, and the Mekong-Mamberamo linguistic area – thereby refuting claims that it is the product of a recent event of creolisation. Riau Indonesian and its relatives exhibit dual heritage: Austronesian and Mekong-Mamberamo. From an Austronesian perspective, the isolating profile developed when Austronesian languages first spread into Nusantara, as a result of contact with the languages that were already present in the region. However, from a Mekong-Mamberamo point of view, the isolating structure may be viewed as the outcome of vertical inheritance dating back as far as we can see.

Keywords: isolating, Riau Indonesian, creolisation, dual heritage, language contact, complexity, Mekong-Mamberamo

1. Introduction

The Riau dialect of Indonesian, as shown in Gil (this volume) represents a relatively clear case of an isolating language, with little morphological structure, and a paucity of evidence for a distinct structural level of word. This chapter, a sequel to that earlier chapter, addresses the question: How and why did Riau Indonesian become isolating?

This chapter argues that Riau Indonesian owes its isolating profile in large part to a constitutive event that took place some 3500 to 4000 years ago: the coming together of two distinct linguistic heritages. The first heritage, *Austronesian*, is of a genealogical nature, reflected primarily in an influx of vocabulary which, around that time, spread into the region from the north. The second heritage, *Mekong-Mamberamo*, is an ancient sprachbund – a linguistic area which, as argued in Gil (2015), extends

from Mainland Southeast Asia through the Indonesian archipelago, referred to here as Nusantara, and into western New Guinea, and is associated with a specific array of grammatical properties, among which is isolating structure.

The dual heritage of Riau Indonesian points towards two distinct but equally valid perspectives on the historical processes that gave rise to the isolating profile of Riau Indonesian. In accordance with the first, genealogically-oriented perspective, isolating structure entered into the Austronesian lineage at the original spread of Austronesian languages into the Western Nusantara region, as a result of simplification due to language contact of one form or another. This perspective follows the traditional method of viewing language histories in terms of tree structures based largely on the phonological substance of morphemes and words. However, in accordance with an alternative areally-oriented perspective, isolating structure can be traced back in direct and uninterrupted fashion to an ancient Mekong-Mamberamo linguistic area, as far back in time as we can see using the traditional methods of historical linguistics; it thus constitutes an instance of vertical inheritance. This perspective is based on an alternative historical approach focusing more on abstract patterns of grammatical features. As argued in this chapter, these two perspectives represent complementary and equally valid takes on the story of Riau Indonesian and its isolating structure.

The profile that a language presents today is the outcome of myriad diverse processes and developments playing out over various time spans ranging from decades through centuries to millennia. And of course, no contemporary language has been in existence as a distinct entity for anywhere near that length of time. Thus, the story of Riau Indonesian is inevitably one shared with other languages connected to it either genealogically or geographically – its relatives, both close and distant.

Figure 1 below shows Riau Indonesian embedded within a sequence of language networks of increasing size, each network wholly contained within the network (or networks) immediately above it in the diagram, each network thereby comprising an aggregation of increasingly distant relatives of Riau Indonesian.

Working upwards through Figure 1, Riau Indonesian is one of a set of Malay/Indonesian koinés, which includes also varieties such as Sabah Malay, Jakarta Indonesian, Kupang Malay, Papuan Malay and others. These in turn are a subset of all Malay/Indonesian varieties, including also the likes of Siak Malay, Patani Malay, Brunei Malay as well as Standard Malay/Indonesian. Malay/Indonesian itself is but one member of the Malayic language family, including among others languages such as Minangkabau, Besemah, Mualang and Iban. Malayic in turn may be considered to be part of a larger set of Western Nusantara languages, spoken in Sumatra, Java, Borneo and their assorted satellite islands; some other examples of Western Nusantara languages might include Mentawai, Acehnese, Javanese and Kenyah. Here, though, the diagram splits into two: On the one hand, Western Nusantara

Austronesian   Mekong-Mamberamo

Western Nusantara

Malayic

Malay/Indonesian

Malay/Indonesian koinés
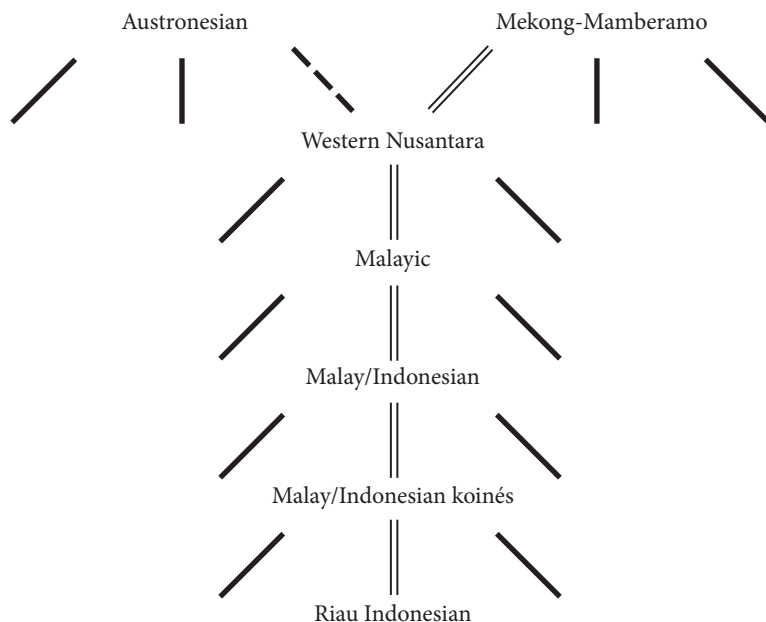
Riau Indonesian

Figure 1. The story of Riau Indonesian

languages are all members of the Austronesian language family, which includes also languages of Taiwan, the Philippines, Madagascar, the Pacific and elsewhere; on the other hand, Western Nusantara languages are part of the Mekong-Mamberamo linguistic area, which, in addition, includes also non-Austronesian languages such as Karen, Vietnamese, Bunaq and Hatam.

Of the language networks represented in Figure 1, only Malayic and Austronesian constitute genealogical groupings. Malay/Indonesian koinés and Malay/Indonesian are defined in terms of a combination of linguistic and sociolinguistic features, Western Nusantara is a purely geographical grouping even though all of its languages are Austronesian, while Mekong-Mamberamo is a sprachbund, or linguistic area, containing genealogically unrelated languages sharing a set of linguistic properties.

While Figure 1 makes reference to contemporary networks of languages, each such network is associated with a set of shared features which may accordingly be attributed to a point in the past. Figure 1 may thus be understood as representing a time-line for Riau Indonesian, reaching upwards from the present, at bottom, to the distant past, at top. The mixture of genealogical and other kinds of language networks in Figure 1 reflects the fact that the story of Riau Indonesian can only be properly understood in terms of a complex interplay of genealogical and areal factors.

Figure 1 provides a schematic representation of the story of Riau Indonesian proposed in this chapter. The branching at the top of the diagram offers a graphic portrayal of the dual Austronesian and Mekong-Mamberamo heritage of Riau Indonesian and its relatives. The genealogical perspective is reflected in the dashed line which portrays the adoption of isolating structure as the outcome of language contact that took place during the original spread of Austronesian languages into Nusantara some 3500–4000 years ago. In contrast, the areal perspective is represented by the double line which traces the vertical inheritance of the isolating profile from contemporary Riau Indonesian all the way back to the ancient Mekong-Mamberamo sprachbund.

As implied by Figure 1, the story of Riau Indonesian is shared by an increasingly large sequence of language networks as one goes back in time. Thus, much of what is argued here about Riau Indonesian holds true for other Malay/Indonesian koinés, the entire Malay/Indonesian language, the whole Malayic language family, and the languages of Western Nusantara more generally. Indeed, echoes of the arguments presented here may also be found in Conners' (this volume) account of Javanese, as well as, even further afield, Elias' (this volume) analysis of Lio, and Brunelle's (this volume) story for Cham. There is nothing at all special about Riau Indonesian: its choice as the focal point of this chapter's narrative is for largely practical and rhetorical reasons – one has to start somewhere.

As proposed in Gil (2015), isolating structure may be viewed as one aspect of a more general notion of grammatical simplicity. In particular, as argued in Gil (2017a), the typological profile of Riau Indonesian comes close to instantiating an ideal language type, that of *IMA Language*, characterised by *Isolating* word structure (Gil this volume, Chapter 1), *Monocategorial* syntax, lacking distinctions between major parts of speech such as noun and verb (Gil 1994, 2000, 2005, 2013), and *Associational semantics*, in which the meaning of a complex expression is derived from the meaning of its constituent parts by loose association, without recourse to more construction-specific rules (Gil 1994, 2017a). Stepping back from the Riau dialect, it was a similar IMA-language typological profile that led the renowned novelist Anthony Burgess to describe his encounter with the Malay language as like "diving into a bath of pure logic [in which] everything is pared to a minimum" (Burgess 1975: 183).

Accordingly, this chapter situates the development of the isolating profile of Riau Indonesian and its relatives within the broader question of how and why these languages came to acquire their characteristically simple grammatical profiles. A venerable tradition of linguistic scholarship attributes grammatical simplification to language contact. The present account follows in that tradition; however, it differs from many previously proposed accounts in arguing that the contact in question,

or at least the lion's share of it, took place at a much earlier period, long before the formation of the Malay language, or even the Malayic language family.

Many of the ideas presented in this chapter are the outcome of an ongoing dialogue with John McWhorter, spanning the better part of two decades, debating the role of language contact in the shaping of Riau Indonesian and other similar language varieties. In Gil (2001) it is argued that in the recent past, at least, language contact plays a relatively minor role in the simplicity characteristic of Riau Indonesian, and that, in particular, there is no justification in characterising it as a creole language. In response, McWhorter (2001c, 2007, 2011a, 2019 and elsewhere) insists that Riau Indonesian is indeed a creole, and that, together with other related language varieties, it is the product of two waves of contact-induced simplification: the first leading from proto-Malayic to the Malay language, which he characterises as a Non-Hybrid Conventionalised Second Language, and the second leading from Malay to Riau Indonesian and other Malay/Indonesian koinés, which he views as involving creolisation. Much of this chapter is a continuation of this dialogue, and a response to McWhorter's above arguments. In particular, in the story of Riau Indonesian presented in this chapter, the role of language contact is afforded greater recognition, reflecting, inter alia, the force of McWhorter's arguments. Nevertheless, the present account still differs from that offered by McWhorter in two substantial respects. First, the vast majority of the contact-induced simplification is argued to have taken place much earlier than supposed by McWhorter, namely at or soon after the original intrusion of Austronesian languages into Nusantara, that is to say long before the formation of proto-Malayic. Secondly, it is suggested that the simple grammatical profile reconstructed for Mekong-Mamberamo languages prior to the Austronesian intrusion, subsequently inherited by Riau Indonesian and many of its relatives, must, for the time being at least, remain unexplained; at such great time depth, there is simply insufficient evidence to support a contact-induced simplification account.

The remainder of this chapter is organised as follows. Section 2 presents a critical review of previous accounts for the isolating profile of Riau Indonesian and its relatives. Section 3 surveys the sociohistorical landscape of the region, while Section 4 provides an overview of the linguistic landscape with a focus on isolating structure. Finally, Section 5 brings together the sociohistorical and the linguistic facts, proposing a historical account for the isolating profile of Riau Indonesian and its relatives in terms of a dual Austronesian / Mekong-Mamberamo heritage.[1]

1.　Most of the data cited in this chapter is from the author's own fieldwork, supplemented by the various corpora in Gil et al. (2015).

2.   Previous accounts

The isolating structure and apparent grammatical simplicity of Riau Indonesian and its relatives have prompted discussion and debate not just amongst Malay/Indonesian specialists but in the linguistic community at large.

As is typically the case more generally, diachronic accounts of Malay and Indonesian are often couched in terms of labels denoting language types whose definitions make reference to external sociohistorical circumstances: *Pidgin-Malay Derived, Creole, Non-Hybrid Conventionalised Second Language*. This section surveys, in turn, the discussion and debate around each of these three terms.

As is suggested below, while all of the above types may be of some relevance to the formation of Riau Indonesian and its relatives, none come close to constituting the whole story of how these languages came to be the way they are. More generally, such labels, in at least some of their usages, reflect an unwarranted exoticisation of Riau Indonesian and its relatives, founded on an unjustified conception of what an "ordinary" language should look like – a relic of a hopefully bygone era in which the point of departure for linguistic theory was still standardised national languages, Standard Average European, and the languages of mostly WEIRD (White, European, Industrialised, Rich and Democratic) speakers.

2.1   Pidgin Malay Derived

The term *Pidgin Malay Derived (PMD)* is introduced by Adelaar and Prentice (1996), where it represents one of a triad of "categories of sociolects" of Malay, alongside inherited Malay dialects and classical literary Malay. In principle, the category of PMD seems as though it ought to correspond to that of the Malay/Indonesian koinés in Figure 1 above. However, while including eastern varieties such as Kupang, Manado and Papuan Malay, only one western variety is included in their category of PMD, namely Jakarta Indonesian. Although Adelaar and Prentice purport to provide an exhaustive survey of Malay/Indonesian varieties, no mention is made of any of the other varieties of colloquial Indonesian spoken in Sumatra, Java and Borneo. (A plausible explanation for this lacuna can be found in the fact that they rely largely on bibliographical sources, and at that time, very little material was available on western varieties of Indonesian, such as Riau Indonesian.)

Adelaar and Prentice provide no explicit justification for their characterisation of these varieties as being derived from pidgins. Indeed, they seem to be hedging their bets somewhat, as is evident in the following passage:

> … this lingua franca Malay variety […] was apparently the result of a pidginisation process. Some scholars (Lim 1988, Prentice, Adelaar) believe that this process was reinforced by contact between Malay and Chinese traders, as this lingua franca Malay shares with southern Chinese dialects a number of features which distinguish it from other forms of Malay […] Other scholars (Collins, Nothofer) argue that these phenomena are the result of common pidginisation processes. Moreover, some of them are also recorded in vernacular Malay dialects and may be regularly inherited features which were lost in literary Malay. (pp. 674–5)

Even today, pidgins are woefully underdescribed in the linguistic literature, so it is hardly surprising that there is little material available on a potential Malay pidgin that might have been spoken in the region in the past, from which today's Malay/Indonesian koinés would potentially be derived. As they point out, Malay-based pidgins are spoken today; the so-called Bazaar Malay of Malaysia is cited as one such example. So, it is quite likely that such pidgins were around also in the past, and not beyond the realm of possibility that some of those pidgins might have undergone creolisation, resulting in some of the Malay/Indonesian koinés spoken today. However, no direct evidence is provided for this scenario. And more generally, as argued by McWhorter (2007: 235–236), it is not necessarily the case that the kind of radical simplification exhibited by PMDs is necessarily preceded by a prior stage of pidginisation.

Adelaar and Prentice propose a list of eight grammatical features common to PMDs (p. 675):

(1) a. Possessive constructions consisting of possessor + *puɲa + possessed item
   b. Plural pronouns derived from singular pronouns + *oraŋ ('human being')
   c. Retention of *tar- and *bar as the only productive original Malayic affixes
   d. *ada, the Malay existential marker, indicating progressive aspect
   e. Reduced forms of the demonstratives *ini and *itu preceding a noun and functioning as determiners
   f. The use of a reduced form of *pərgi 'to go' as a verb as well as a preposition meaning 'towards'
   g. Causative constructions consisting of the auxiliaries *kasi/*bəri ('to give') or *bikin/*buat ('to make') + the head verb
   h. The use of *sama or another word as a multifunctional preposition (also for direct and indirect objects)

Adelaar and Prentice emphasise that "[t]he above features are not diagnostic in themselves for being PMD-derived, but large configurations of them are". However, since all of these features involve Malay-specific forms, they cannot be considered as supportive in principle of a pidgin origin for PMDs, but only as contingent features potentially reconstructible back to an ancestral dialect of Malay regardless of its

putative status as a pidgin. In any case, their applicability to the western Indonesian koinés is at best limited. Thus, as shown in Gil (2001: 335–337), Riau Indonesian has only (a), somewhat marginally, and (g); similarly, Jakarta Indonesian has just (g), even though Adelaar and Prentice explicitly characterise it as a PMD (p. 676)[2] Accordingly, whatever the status, pidgin or otherwise, of a potential proto-dialect of Malay defined by the features in (1), it is not a plausible ancestral dialect for Riau and other western koiné varieties of Indonesian. In other words, such varieties are not PMDs (pace McWhorter 2007: 222–229 and elsewhere, who uses the term PMD with reference to Riau Indonesian and other such koinés).

In the absence of explicit evidence, one gets the feeling that the characterisation of koiné varieties of Malay/Indonesian as pidgin-derived is more of a reflection of their perceived aberrant nature when viewed against a more normative notion of what a Malay variety "should" look like – a notion that stems largely from the privileged status afforded to standardised varieties of Malay within historical investigations. Thus, when describing such varieties, Adelaar and Prentice take Standard Malay as their yardstick of comparison, writing that "[t]hey differ from literary Malay most notably in a reduced morphology…" (p. 674). They are hardly alone

---

2. (a) Prenominal possessives with *punya* are marginal in Riau and absent from Jakarta. (b) Plural pronouns with *orang* are absent from both Riau and Jakarta (in Gil 2001 they are characterised as marginal, but this is probably more appropriately attributed to code-mixing with other dialects). (c) Both Riau and Jakarta have productive use of an agent-oriented generalised voice prefix commonly represented as *N-*, a patient-oriented generalised voice prefix *di-*, a causative/ applicative suffix *-kan* or *-in*, and one or two other productive affixes. (d) Both Riau and Jakarta use *lagi*, rather than *ada*, to mark progressive aspect. (e) In both Riau and Jakarta, demonstratives in attributive function follow, rather than precede, their hosts. (f) In both Riau and Jakarta, the reflex of **pərgi* lacks the [r] but remains disyllabic; there are no reduced forms such as *gi* or *pi* attested elsewhere. Also, it has no prepositional usages. (g) Both Riau and Jakarta form periphrastic causatives with *bikin* or *buat*. However, unlike most eastern Malay varieties, they also mark causatives with suffixation of *-kan* or *-in* (mentioned above). (h) The use of *sama* as a multifunctional preposition marking, among others, indirect objects, is widespread across Malay/Indonesian, including Riau and Jakarta in the west, as well as several eastern Malay varieties. However, the overt marking of direct objects is absent from Riau and Jakarta, as well as, for that matter, from most eastern varieties; to the best of my knowledge it is attested only in Manado and North Maluku Malay, most commonly with *pa*, though occasionally also with *sama*.

(With regard to the latter feature, I am inclined to suspect that Adelaar and Prentice must have been familiar with these facts, and simply erred in their wording. If, in (1h), "direct and indirect" is replaced with "direct and/or indirect", or. better still, just with "indirect", the resulting characterisation would hold true of most of their PMDs, as well as, for that matter Riau and Jakarta Indonesian. Indeed, in Gil (2001: 337) I also "misread" them in this way, and characterised Riau Indonesian as satisfying this feature. Adopting such an alternative formulation would thus slightly "jack up" the score of Riau and Jakarta Indonesian with respect to the features in (8), though it would still remain very low.)

in this respect; an even more striking example of such a perspective is offered by Sneddon's (2006) reference grammar of Jakarta Indonesian, which chooses to describe it in terms of how it differs from Standard Indonesian.[3]

In reality, though, it is the colloquial varieties of the language that are ontologically prior; they are the ones that are acquired automatically and at an early age by children and are mastered by everybody in the speech community. In contrast, the standardised versions are acquired at a later age, largely through schooling, and are fully mastered by only some members of the speech community; they are thus derivative from the colloquial varieties. This is clearly visible from an areal perspective, where it is the Malay/Indonesian koinés such as Riau Indonesian that represent the norm with respect to the grammatical profile of an Austronesian language of the region. Some features of Riau Indonesian common to the languages of the region but absent from Standard Indonesian include (a) coexpression of 'and' and 'with'; (b) distributive numerals formed by reduplication; and (c) free word order in the absence of argument flagging for the expression of thematic roles – see Gil (2009b) for more discussion. With respect to features such as these, then, Riau Indonesian, like other Malay/Indonesian koinés, represents the regional norm, while Standard Indonesian is the aberrant language variety in need of explanation. In particular, as argued in Gil (2015) and again in Section 4 below, the strongly isolating nature of the Malay/Indonesian koinés is in fact a characteristic feature of the languages of the region, and not something that arose out of circumstances peculiar to the history of Malay/Indonesian.

2.2 Creole

More recently, the history of Malay and Indonesian has featured in a more general debate on the nature of *creole* languages. At issue is whether the class of creole languages as a whole exhibits structural features distinguishing them systematically from other older non-creole languages; that is to say, can one tell that a language is a creole just by looking at it, without familiarity with its history and external circumstances? On one side of the debate is the uniformitarian position, which holds that, when viewed synchronically, creole languages are just like other languages in all

---

3. Thus, in introducing his grammar, Sneddon writes "At first I expected to produce a work which paralleled my grammar of formal Indonesian, intending it to be a comprehensive statement of the colloquial speech of educated people living in Jakarta. But as I began to analyze data it soon became apparent that this would not be practical." (p. ix). "This work, therefore, is not a comprehensive nor systematic study of grammar. Instead it deals largely with aspects of [Colloquial Jakarta Indonesian] grammar which are noticeably different from corresponding structures in [Formal Indonesian]." (p. 9).

respects; some advocates of this position are Mufwene (1986, 2001, 2008), DeGraff (2001, 2003, 2005), Ansaldo (2004), Aboh (2016), Blasi, Michaelis & Haspelmath (2017) and others. On the other side is a common conviction that creole languages do look different from other languages, a view that is supported by Bakker et al. (2011) and others. In particular, it is argued that creole languages tend to be associated with simpler grammatical structures; see, for example, Parkvall (2008). The most forceful proponent of this latter position is John McWhorter, who, in a series of publications – McWhorter (1998, 2000, 2005, 2011a, 2018) and elsewhere – musters an array of empirical evidence arguing that creole languages can indeed be distinguished from older languages on purely synchronic grounds, with reference to the notion of complexity. As McWhorter puts it in the title of one of his articles (2001b), "the world's simplest grammars are creole grammars".

Enter Riau Indonesian. Gil (2001) presents a contrastive analysis of Riau Indonesian and Saramaccan, the latter being one of McWhorter's stock examples of a creole language, and argues that, according to McWhorter's own criteria, Riau Indonesian is every bit as simple as Saramaccan. On the basis of this analysis, it is then argued that McWhorter's two-way implication, simple if and only if a creole, needs to be weakened to a one-way implication, simple if a creole – but allowing for the possibility, instantiated by Riau Indonesian, that a non-creole language could also exhibit a comparable degree of simplicity. Or, to put it differently: As McWhorter argues, you can tell that some languages are *not* creoles by just looking at them; however, contrary to McWhorter, you cannot tell that a language *is* a creole by just looking at it.[4] However, in his response to Gil (2001), McWhorter (2001c) concurs with the characterisation of Riau Indonesian as being as simple as Saramaccan, but defends his proposed two-way implication by suggesting, pace Gil (2001), that Riau Indonesian is in fact a creole – a position that he has consistently maintained in numerous subsequent publications, including McWhorter (2011a, 2011b: 110, 2018: 28, 2019).

In Gil (2001), it is argued that there is no independent external sociohistorical evidence to suggest that Riau Indonesian is a creole language, that is to say, the product of a recent process of simplification due to massive imperfect

---

4.   The proposed unidirectionality of the relationship between creoles and complexity argued for in Gil (2001) on the basis of Riau Indonesian and Saramaccan was subsequently supported in an extensive cross-linguistic study by Parkvall (2008). In this study, 155 languages from the *World Atlas of Language Structures* (Haspelmath et al. 2005) are ranked with respect to their overall complexity. Of these languages, two, Ndyuka and Sango, are creoles, and as expected they fall at the lower end of Parkvall's scale of complexity. Crucially, however, several non-creole languages, including, among others, Pirahã, Hmong Njua, Kobon and Maybrat, also emerge as being of simplicity comparable to that of the two creole languages, thereby suggesting that simplicity of structure does not necessarily entail being a creole language.

second-language acquisition by adults. These arguments are further developed and refined in Section 3 below. But it's not just Riau Indonesian. A major rhetorical point of Gil (2001) was to dispel the notion that Riau Indonesian is some kind of wacky outlier dialect in comparison to other more run of the mill colloquial varieties of Malay/Indonesian. For this purpose, Riau Indonesian, the major lingua franca of the eponymous province, was compared to one of the traditional dialects of the region, used by ethnic Malays for intra-ethnic communication, namely Siak Malay, showing that in terms of complexity, Siak Malay was only very slightly less simple than Riau Indonesian, while at the same time remaining well within the ballpark of simplicity that McWhorter associates with creole languages. Again, McWhorter's response is to assert that Siak Malay, too, is a creole language, in fact, not just Siak Malay but more generally: "Riau Indonesian and similar dialects of Malay and Indonesian *are* creoles – born of the widespread adult acquisition of Malay/Indonesian" (2018: 28). While McWhorter does not enumerate which dialects of Malay/Indonesian are sufficiently "similar" to Riau Indonesian to qualify, according to his purely structural criterion, as creoles, it is obviously quite a lot of them – as is suggested by the discussion in Sections 4.1 and 4.2 below. Moreover, as evidenced by Siak Malay, it is not just koiné varieties such as Riau Indonesian that exhibit the simple grammatical profile, but also local village dialects whose current sociolinguistic profiles provide no prima facie external reason whatsoever to posit recent events of contact-induced simplification.

In reinforcement of the very general notions of simplicity and complexity, McWhorter (1998, 2005, 2011b) offers a more specific characterisation of creole languages, in terms of what he calls the *Creole Prototype*. In its latest version (2011b: 111) it reads as follows:

(2) A natural language is a creole (i.e. born recently from a pidgin and thus emerged from broken transmission) if it has:
   i. *morphologically:* little or no inflectional affixation, and among unbound inflectional markers, none of contextual inflection, or of inherent inflection of the paradigmatically complex sort,
   ii. *phonologically:* little or no distinction of monosyllabic lexical items or morphosyntactic distinctions via tone or register, and no typologically unusual proliferation of vowels, and
   iii. *semantically:* little or no non-compositional combination of nonreduplicative derivational morphemes with roots.

Clearly and uncontroversially, Riau Indonesian and many other varieties of Malay/Indonesian uphold the first two properties of the Creole Prototype; what is at issue is the extent to which they also uphold the third. With good reason, McWhorter (2007, 2019) grumbles that written sources often do not provide information on

the extent to which languages have non-compositional or semantically opaque derivational morphology, and it is true that my own descriptions of Riau Indonesian have hitherto not addressed this point. So, this is the time and place to do so. As is shown below, Riau Indonesian, as well as other similar Indonesian koinés, do have some non-compositional derivational morphology, but not much, and probably not enough to prevent them from satisfying McWhorter's criteria for meeting the Creole Prototype.

Of course, there isn't much derivational morphology at all, so in our search for non-compositional derivational morphology we are, by necessity, feeding on crumbs. In principle, the kind of opaque morphological semantics that we are looking for can have any or all of the following three diachronic sources: (a) recent innovation; (b) recent borrowing from the standard language; and (c) inheritance from a much earlier stage of the language.[5] As McWhorter reasonably points out, it is only the third of these, namely inheritance, that constitutes a serious challenge to the Creole Prototype; the first two processes may clearly have taken place subsequent to the hypothesised creolisation event.

Perhaps the most well-known affixes in colloquial Malay/Indonesian varieties are the generalised voice markers (Gil 2002; Conners, Bowden and Gil 2015), whose function is primarily semantic, namely, to mark the host word as having an argument bearing the relevant thematic role in its semantic frame. One is the patient-oriented marker *di-*, present in Riau Indonesian and several other varieties; as far as I know, its semantics is always transparent, and I am not familiar with any instances of it being associated with non-compositional semantics. A somewhat more complex picture, however, is painted by the agent-oriented marker *(me-)(N-)*.[6] While its semantics is largely transparent, simply marking the existence of an agent

---

5.   For completeness, a fourth diachronic source needs to be recognised, namely, recent borrowing from other colloquial Malay/Indonesian dialects. Given, however, that our goal here is a general evaluation of Malay/Indonesian dialects with respect to the Creole Prototype, instances of cross-dialectal borrowing do not shed any new light on the question, but merely kick the can further down the road. If, say, an instance of non-compositional derivational morphology in Riau Indonesian is borrowed from Jakarta Indonesian, as indeed a few are, then this borrowing absolves Riau Indonesian of a putative violation of (2-iii) but at the cost of needing to come up with an account for the corresponding form in Jakarta Indonesian.

6.   In this chapter, *(me-)(N-)* is used to refer generically to combinations of two optional elements (a) *me-* (or its cognates), plus (b) *N-*, denoting prenasalisation, a complex morphophonemic process usually involving a homorganic nasal either occurring in front of the host word or replacing its first segment, the details differing from dialect to dialect. For present purposes, no stand is taken on whether forms such as *meng-*, instantiating *(me-)(N-)*, consist of two distinct morphemes, as is suggested by the notation, or a single coalesced one, as is commonly assumed; the answer likely differs from one variety to another.

or actor, there are several classes of exceptions, whose effect is to change the meaning of the host word in ways that are at least partly arbitrary and unpredictable; such exceptions would appear to instantiate each of the three diachronic sources proposed in the preceding paragraph.

A first class of exceptions to (2-iii) is illustrated by the alternation between *simpan* 'deposit', 'put away' and *menyimpan* 'tidy', 'pack' in Papuan Malay. In Papuan Malay, as in most other eastern varieties, the agent-oriented generalised voice marker is absent, except in a handful of words denoting activities in which it is obligatory, such as *menyebrang* 'cross' and *menyanyi* 'sing'. In these cases, its semantics is transparent, asserting the presence of an agent, and thereby assigning the word to the ontological category of activity. In contrast, in the case of the *simpan~menyimpan* alternation, the semantics is partly non-compositional, involving an arbitrary change in lexical meaning: as suggested by the glosses, addition of *meN-* would seem to add an element of pluractionality or plurality of patients. However, such a semantic effect is unattested elsewhere for *meN-*. Thus, the *simpan~menyimpan* alternation is in violation of the third property of the Creole Prototype. Still, it would be unreasonable to invoke it to argue against a possible creole history for Papuan Malay. While occurring in some other eastern varieties of Malay, to the best of my knowledge it is unattested in other varieties of Malay/Indonesian. Assuming, as is plausibly suggested by Paauw (2009), that eastern Malay varieties are the descendants of a common ancestral dialect most probably spoken a few hundred years ago in Maluku, the *simpan~menyimpan* alternation with its non-compositional semantics is thus most likely a recent innovation that occurred subsequent to the break-up of Proto-Eastern Malay. Thus, it would not be inconsistent with the possibility that Proto-Eastern Malay was a creole language entirely lacking in non-compositional derivational morphology.

A second class of exceptions to (2-iii) involves loans from the standard language, such as Riau Indonesian *meninggal* 'pass away' from *tinggal* 'stay' / 'leave', and *mendarat* 'land' (said of an airplane) from *darat* 'land' / 'ground', both instantiating non-compositional semantic relationships. Their status as loans is clear from the form of the prefix, *meN-*, rather than the usual Riau Indonesian *N-*: the expected Riau Indonesian forms would have been *ninggal* for the former, and for the latter either *ndarat*, or, more likely, just *darat*, with no prefixation at all. On the other hand, their pronunciation as [meniŋgal] and [mendarat] shows that they are phonologically integrated loans in Riau Indonesian: if their occurrences in naturalistic Riau Indonesian speech were instances of code-switching rather than borrowing, they would be more likely to be pronounced as [məniŋgal] and [məndarat]. All colloquial varieties of Malay/Indonesian are replete with such examples, an obvious consequence of their sociolinguistic ecology as basilectal varieties in a diglossic relationship to the more prestigious standard language. In general, loan words from

the standard language tend to belong to semantic fields associated with technology, nation-wide culture, and globalised civilisation – all brought in from the outside. And of course, in many such cases, the recency of the borrowing is evident from the referent of the loan word, which can easily be dated – such as, for *mendarat*, to the introduction of aviation. Thus, examples such as these should, also, not be taken as evidence against the characterisation of Riau Indonesian and other Malay/Indonesian varieties as creoles; indeed, they present a close parallel to similar examples in Haitian Creole argued by McWhorter (2005: 26–29) to be recent loans from Standard French.

It is the third class of exceptions to (2-iii) that pose a more serious challenge to the claim that Riau Indonesian and other related varieties satisfy the Creole Prototype; however, at present, I am familiar with just a couple of cases of semantically opaque *(me-)(N-)* that are neither an innovation nor a borrowing from the standard language.

One example is the alternation between *cari* 'look for' and *mencari* 'seek livelihood'/'gather natural foodstuffs (e.g. shells on tidal flat)'. Although the meanings of the two forms are clearly connected, the addition of *meN-* perhaps introducing durativity, pluractionality or plurality of patients (reminiscent of *simpan~menyimpan* previously), the semantic relationship between them is nontransparent, and needs to be specified on an ad hoc basis in the lexicon. While *cari*, and its regular prenasalised forms, meaning 'look for', are common across Malay/Indonesian, the distribution of *mencari* (and its cognates) with the above-mentioned more specific meaning is rather less widespread. Among the Malay/Indonesian koinés, it is present in Papuan, North Maluku and Ambonese Malay, and also, somewhat less commonly, in Riau Indonesian; however, it is absent from Jakarta Indonesian and many other varieties. Crucially, it is present also in Minangkabau and Besemah, both spoken in Sumatra, the latter in particular being relatively distantly related to Malay and Indonesian within the Malayic subgroup. However, it is absent from Standard Indonesian, in which the prenasalised form *mencari* has the same word meaning as simple *cari* – so it is clearly not a borrowing from the standard language. Thus, it would seem to be a case of non-compositional derivational morphology that is a plausible candidate for reconstruction to proto-Malayic, or at least to a stage in the diversification of Malayic that is prior to whatever recent events of creolisation might have putatively taken place in the history of one or more of the Malay/Indonesian koinés.

A second rather different instance of non-compositional derivational morphology involving *(me-)(N-)* is that of its application to the question word *apa* 'what', which tends, in general, to result in two semantic outcomes. The first is purely transparent: *(me-)(N-)* plus *apa* means, simply, 'do what'. This usage is common across the western Indonesian koinés, such as Riau and Jakarta Indonesian,

typically with *ngapain* 'do what'.[7] However, the second is at least partly opaque: *(me-)(N-)* plus *apa* constitutes one of several strategies for asking about a reason or intention, that is to say, to express the meaning 'why'.[8] Some examples of this include Riau Indonesian *ngapa*, Siak Malay *ngapo*, Minangkabau *manga*, Besemah *ngape*, and Mualang *ngapa* (Tjia 2007: 251), as well as Standard Malay/Indonesian *mengapa* – all meaning 'why'. The diversity of these forms would seem to preclude the possibility that they are all borrowings from Standard Indonesian. And the widespread distribution of these forms, especially in Mualang of West Borneo, a member of the Ibanic subgroup, a possible first-order branching of Malayic, suggests that the non-compositional meaning of *(me-)(N-)* plus *apa* is reconstructible to proto-Malayic, and thus also prior to any recent possible events of creolisation that may have potentially given rise to some of the Malay/Indonesian koinés.

Other affixes in Riau Indonesian and its relatives present a similar picture. For the most part, cases of non-compositionality appear to be either relatively recent innovations, or else borrowings. For example, Conners, Bowden and Gil (2015) cite examples of semantic opacity involving the end-point-oriented generalised voice marker *-in* in Jakarta Indonesian; whereas in most cases its associated semantics is causative or benefactive, in a few cases it is unpredictable, such as, for example in *tutup* 'close' / *tutupin* 'cover'. However, to the best of my knowledge, this particular idiosyncrasy dos not occur in any other variety; thus, for example, in Riau Indonesian, *tutupin*, or its more common variant *tutupkan*, are associated with the usual benefactive interpretation 'close for'. Thus, the non-compositional semantics of *tutupin* would appear to be a relatively recent innovation in Jakarta Indonesian. In contrast, there would seem to be a relatively small number of cases of semantic opacity that are reconstructible to earlier stages of the Malayic subgroup.

The prefix *ber-*, sometimes referred to as a marker of "medial" voice, is associated with a variety of usages many of which may be characterised in terms of a depatientive generalised voice marker whose function is to demote or reduce the salience of a patient argument. However, in a few cases, its function is clearly

---

7. In Riau Indonesian, *ngapain* is probably a borrowing from Jakarta Indonesian. Support for this claim is that whereas for most words, the end-point-oriented generalised voice markers *-in* and *-kan* occur in free variation (with *-kan* the more common and more basilectal of the two), in the case of *ngapain*, the alternative *ngapakan* is rare. Further support is provided by the naturalistic spelling facts described in Gil (this volume, Chapter 1, Section 3.3.14): whereas *-in* is occasionally written separately from its host, when occurring in *ngapain* it almost never is, thereby suggesting that *ngapain* was borrowed into Riau Indonesian as a single item.

8. Some other forms based on *apa* meaning 'why' in koiné varieties of Malay/Indonesian include *kenapa*, historically derived from *kena apa* (undergo what), in Riau Indonesian and many other varieties; *apasal*, from *apa pasal* (what cause), in Kuala Lumpur Malay; and *buat apa* (do what) in Jakarta Indonesian.

non-transparent. One example is the alternation between *angkat* 'lift' and *berangkat* 'depart', the latter presumably being derived from the former in the context of a boat raising its anchor. This alternation is evident in forms such as Riau Indonesian *berangkat*, Jakarta Indonesian *berangkat,* Papuan Malay *brangkat*, Minangkabau *barangkek,* Besemah *berangkat*, Sambas Malay *berangkat* and Standard Indonesian *berangkat*, all meaning 'depart'. A second example is the alternation between *diri* 'self' / 'body' and *berdiri* 'stand', observable in forms such as Riau Indonesian *bediri*, Papuan Malay *badiri*, Sambas Malay *bediri*, Meliau Malay *bediri,* Belangin Kendayan *badiri*, Mahap *badiri*, and Standard Indonesian *berdiri*, all meaning 'stand'. A third example is the alternation between *apa* 'what' (cf. preceding paragraph) and *berapa* 'how much', evident in forms such as Riau Indonesian *berapa,* Jakarta Indonesian *berapa*, Papuan Malay *berapa*, Minangkabau *bara*, Besemah *berape*, Iban *berapa*, Mualang *berapa* (Tjia 2007: 390), and Standard Indonesian *berapa*, all meaning 'how much'. In all three cases, the genealogically and geographically widespread distribution of the alternations suggests that they may be reconstructed to proto-Malayic, and hence that they predate any more recent event of creolisation that might have taken place since.

Perhaps the most productive source of non-compositional derivational morphology in Riau Indonesian and its relatives is provided by the suffix *-an*. Descriptions of Standard Malay/Indonesian typically describe the primary function of *-an* as that of deriving nouns, for example Sneddon (1996: 30–34); however, such an analysis is at odds with the characterisation of Riau Indonesian and at least some of its relatives as monocategorial, lacking in a noun-verb distinction. And indeed, examining the effect of adding *-an* in terms of semantic categories of thing- and activity-denoting expressions, one finds cases of all four logical possibilities, as illustrated by the following examples from Riau Indonesian: *jaring* 'net' / *jaringan* 'network' (thing > thing); *gaji* 'salary' / *gajian* 'receive salary' (thing > activity); *jual* 'sell' / *jualan* 'sell habitually' (activity > activity); *main* 'play' / *mainan* 'toy' (activity > thing) – under an analysis identifying semantic categories with putative distinct parts of speech, only the last of the examples would qualify as a nominalisation. The preceding examples offer some feel for the semantic diversity and irregularity of the usages of the suffix *-an*; indeed, for the most part it is hard to identify any systematic semantic effects associated with the addition of *-an*, it could thus be said that virtually all of its usages are semantically opaque.[9]

---

9.   An exception to the complete lack of semantic transparency of *-an* can be observed in Jakarta Indonesian, in which, for property words, adding *-an* systematically induces a comparative interpretation, e.g. *gede* 'big' / *gedean* 'bigger', *bagus* 'good' / *bagusan* 'better', and so forth (see also Section 4.1 below). However, to the best of my knowledge, no other variety of Malay/Indonesian has this usage; moreover, even in Jakarta Indonesian, it is an island of regularity

Worthy of note is that in many of the Malay/Indonesian koinés, such as Riau Indonesian, Papuan Malay and Sabah Malay, the suffix -*an* would seem to be totally non-productive, with its many occurrences falling into two main classes. The first, exemplified by Riau Indonesian alternations such as *bundar* 'round' / *bundaran* 'roundabout', *takbir* 'profession of Islamic faith via the pronunciation of *Allāhu akbar*' / *takbiran* 'participate in procession during which the *takbir* is chanted', *internet* 'internet' / *internetan* 'use the internet', involve innovations, as evidenced by the meanings, which typically refer to concepts associated with relatively recent times, and/or the forms, which are often based on stems borrowed from languages such as English or Arabic, with which Malay/Indonesian was in contact in the last few centuries. Indeed, it is probable that the forms containing -*an* were themselves borrowed into Riau Indonesian whole hog from some other variety, such as Jakarta Indonesian, Standard Indonesian, or some earlier variety of classical Malay.

The second class, however, contains forms with -*an* for which there is no evidence to support that the claim that they represent recent developments or borrowing. One such example is the alternation, in Riau Indonesian and other koiné varieties, between *duri* 'thorn' and *durian* 'durian' (a kind of thorny fruit). The word *durian* is pretty much pan-Malayic, occurring in varieties such as Minangkabau *durian*, Besemah *durian*, Sambas Malay *durian,* Iban *rian*, Mualang *rian* (Tjia 2007: 403), as well as Standard Malay/Indonesian *durian*; in fact, it is also borrowed into other languages as in Thai *thúrian*, and even globalised English. In all likelihood, then, the form *durian* dates back to proto-Malayic and even earlier. A large number of other words resemble *durian* in that they consist of a native word plus -*an*, exhibit opaque semantics, and refer to concepts that are common and not of recent provenance; some examples of such alternations in Riau Indonesian include *masak* 'cook' / *masakan* 'cuisine', *tutup* 'close' / *tutupan* 'lid', *kotor* 'dirty' / *kotoran* 'dirt', *bulan* 'moon', 'month' / *bulanan* 'monthly'. All of these, plus many others, are widespread throughout Malay/Indonesian, and occur also in the standard language. Each requires its own individual analysis in order to determine how far back it can be reconstructed, and it may indeed be the case that some of these forms represent borrowings from the standard language or some earlier version of classical Malay. Unfortunately there is currently insufficient data available for precise determinations to be made in most cases. Still, it is likely that many, like *duri/durian*, can

---

within an otherwise largely irregular system – thus, all of the four examples cited above from Riau Indonesian are present also in Jakarta Indonesian. Elsewhere I have glossed -*an* as AUG, suggesting the presence of some kind of shared *gesamtbedeutung* involving a notion of "augmentative"; however, this should not be taken too literally, and in any case, even if such a common general meaning were shown to be present, its realisation in individual forms would still contain a significant component that remained semantically unpredictable.

be reconstructed a significant way back, perhaps as far as proto-Malayic or even further, in which case they would have had to have passed unscathed through the hypothesised event of creolisation that occurred in the more recent past of Riau Indonesian and other Malay/Indonesian.

In summary, then, forms like Riau Indonesian *mencari* 'seek livelihood' / 'gather natural foodstuffs', *ngapa* 'why', *berangkat* 'depart', *bediri* 'stand', *berapa* 'how much', and *durian* 'durian' represent instances of non-compositional derivational morphology of the kind counterindicated by item (2-iii) of the Creole Prototype. Moreover, this is hardly an exhaustive list; further investigations would presumably reveal numerous other such cases, involving the same as well as other affixes, in Riau Indonesian and also other Malay/Indonesian koinés. Possibly though, Riau Indonesian and its relatives could be squeezed into the Creole Prototype by invoking the "little or no …" hedge conveniently provided in (2-iii). In fact, it is not clear to me why even a large number of cases of non-compositional derivational morphology could not be expected to survive an event of radical restructuring and creolisation; after all, when the relationship between pairs such as, say, *cari* and *mencari* is opaque, the would-be imperfect adult second-language learner could simply acquire them as separate words from the lexifier language.

But my aim here is not to question McWhorter's claim that all creoles satisfy the Creole Prototype. Rather, it is to argue against his claim that *only* creoles satisfy the Creole Prototype. Just as in Gil (2001), the relationship between general simplicity and creoles is argued to be unidirectional rather than bidirectional, so, in the present case, the relationship between the Creole Prototype and creoles is suggested to be unidirectional rather than bidirectional: while being a creole entails upholding the Creole Prototype, upholding the Creole Prototype does not necessarily mean being a creole. Thus, notwithstanding the occasional instances of non-compositional derivational semantics, Riau Indonesian and its relatives do uphold the Creole Prototype pretty well. But this should not be taken to mean that they *are* in fact creoles.

Indeed, in spite of the influence of the proposals by Adelaar and Prentice and then McWhorter, the view that Riau Indonesian and other similar contact varieties of Malay/Indonesian are creole languages remains a minority position within the field; numerous other scholars have expressly rejected such claims. For example, Wolff (1988: 86–87) writes that "I have seen absolutely no proof that any of the living dialects of Indonesian/Malay are indeed creoles, despite the uncritical repetition of this notion in article after article and textbook after textbook." Similarly, Paauw (2009: 27) claims that "[…] it seems unlikely that Low Malay is itself the result of creolisation, as there is no evidence for large-scale language contact in the Malay homeland, unless it happened over 2000 years ago, in the original migrations of Malay speakers from the Malay homeland in Borneo". Other rejections of a creole

characterisation for specific varieties of Malay/Indonesian include Collins (1980) for Ambonese Malay, Steinhauer (1991) for Larantuka Malay, and Kluge (2014) for Papuan Malay.

In line with the conclusions of the above-cited scholars, Section 3 below argues that there is simply no evidence that Riau Indonesian and its relatives are the product of a recent event of creolisation, and in fact a number of good reasons to believe that they are not. With reference to an arbitrarily chosen creole language Angolar, McWhorter (2001c: 408) writes:

> I thoroughly understand that at this point, some readers will be inclined to grouse: 'If you give him a language that does look like a creole, he'll just say it's a creole too!' But actually, my conception is eminently refutable. If I were presented with a language whose history did NOT involve acquisition being more often by adults outside of a school setting than by children, and this language were nevertheless as underspecified as Riau Indonesian, then I would readily concede that even an older language can attain a level of relative simplicity akin to Angolar's.

However, as argued in Sections 3 and 4 below, Riau Indonesian is in fact the language that McWhorter is asking for, as indeed are many of its close and not so close relatives.

2.3 Non-hybrid Conventionalised Second Language

As portrayed above, an unfortunate property of the creole debate is that it tends to be couched in categorical terms: either a language is a creole or it isn't. Early attempts to break free of the binary straightjacket are reflected in terms such as "creoloid" – see Platt (1975) on Singlish. A more systematic move towards redressing this problem is that of McWhorter (2007), offering a conceptual framework making it possible to speak, in continuous rather than discrete terms, of more contact being associated with more simplification – with creoles representing the extreme point on the cline, involving the greatest amount of simplification.

McWhorter's framework is summarised in Table 1 below, adapted from McWhorter (2007: 254). Although represented in tabular form with discrete cells, McWhorter emphasises that the categories in question are idealisations, and that the reality is actually more fuzzy and continuous involving two orthogonal scales. The horizontal axis represents the grammatical domains in which contact effects are observed, ranging from lexical though lexicon and syntax to lexicon, syntax, morphology and phonology, while the vertical axis represents the degree of contact-induced simplification, ranging from none through moderate to extreme.

The bottom row of the table, involving extreme simplification, represents various types of creole languages. An implication of the table is that creoles, as the

Table 1. McWhorter's typology of language contact

	Lexicon only	Lexicon and syntax	Lexicon, syntax, morphology, phonology
No Simplification	languages with lexical borrowing	languages with sprachbund effects	intertwined languages and language areas
Moderate Simplification	Non-hybrid Conventionalised Second Languages	semi-creoles	semi-creoles based on typologically close languages
Extreme Simplification	mesolectal creoles without substrate influence	mesolectal creoles with moderate substrate influence	deeper/radical creoles

term is commonly understood, are not qualitatively different from other kinds of contact languages but only quantitatively so, in that they exhibit contact effects to a greater degree than do other contact languages. A putative qualitative difference between creoles and non-creole languages would be that the former are derived via the nativisation and subsequent initial complexification of even simpler pidgins, whereas other cases of simplified contact languages do not pass through a prior pidgin stage; however, McWhorter expressly denies that creoles must necessarily be derived from an earlier pidgin stage.[10]

McWhorter's focus is on the middle row, representing language types involving moderate contact, and, in particular, on the leftmost cell in the row, representing a hitherto unrecognised type for which McWhorter introduces the term *Non-hybrid Conventionalised Second Language*, or *NCSL*. According to McWhorter, NCSLs are typically "large" languages, which, by dint of their sociohistorical dominance, have acquired substantial numbers of second-language speakers, whose imperfect mastery of the language results in simplification, which then spreads from them to the language as a whole.[11] Structurally, NCSLs are typically recognisable by being of lesser complexity than their smaller non-NCSL near relatives. McWhorter (2007)

---

10. McWhorter's views on this particular issue would appear to have evolved over the years, albeit somewhat inconsistently. In earlier writings, e.g. McWhorter (2001b, 2011b: 83,111) he seems to adhere to the position that creoles must necessarily be derived from pidgins, criticising other scholars who do not share this view. However, in other publications, e.g. McWhorter (2007: 17–18, 2011a: 221, 2019), he explicitly embraces the possibility that at least some creoles may have not passed through a prior pidgin stage.

11. McWhorter's notion of NCSL is in line with a number of recent proposals to the effect that larger languages tend to favour simplification while smaller ones tend to be more conducive to complexification; see, for example, Dahl (2004, 2009), Nichols (2009), Sinnemäki (2009), and Trudgill (2009, 2011).

presents five case studies of large languages, arguing that in each case, the language in question is less complex than its closest relatives and hence worthy of characterisation as an NCSL. The first four languages are English, Mandarin Chinese, Persian and Colloquial Arabic, while the fifth is "Malay" – which, here, refers to Standard Malay/Indonesian.

As a point of departure, McWhorter (2007: 197) approvingly cites Dalby's (1998: 391) statement to the effect that "A lingua franca needs to be easy to grasp, and Malay has a more approachable structure than its relatives". McWhorter supports his characterisation of Malay as an NCSL by providing evidence to the effect that Malay is indeed simpler than its relatives; this evidence is summarised in the following table (adapted from McWhorter 2007: 208) comparing the relative complexity of Malay and 12 of its relatives with respect to 9 selected features indicative of grammatical complexity. In Table 2, white cells represent cases of low complexity, and grey cells instances of higher complexity.

Table 2. McWhorter's analysis of Malay as an NCSL

As suggested by Table 2, Malay is simpler, overall, than its 12 selected relatives; its relationship to them is thus similar to that of say English in comparison to other Germanic languages such as Dutch, German, Danish and so forth. Prima facie, McWhorter's characterisation of Malay as an NCSL would appear to be plausible. However, Table 2 and the study underlying it are problematical in several ways.

The first is empirical. All broad typological studies have an error rate, which should not necessarily be taken as damning: in the quest for the big picture, a small number of factual mistakes should be tolerated, with the understanding that they will come out in the statistical wash. Nevertheless, the apparent error rate of Table 2

is a bit disconcerting.[12] Looking just at the language that I am most familiar with, Minangkabau, there are two apparent errors: for both demonstratives and classifiers, I am not familiar with any evidence to the effect that the Minangkabau system is more complex than the Malay one.[13] Similarly, for Javanese, the demonstrative system is also not significantly more complex than that of Malay.[14] One can only hope that this error rate does not carry over into other languages in the table. Still, the fault is not entirely with McWhorter, as a study of this kind is only as good as the descriptions on which it is based, and these are often lacking in many respects. And it is reasonable to suspect that, once the errors are fixed, the overall picture presented in Table 2 would not change beyond recognition.

A second problem is with the choice of languages. As McWhorter (2007: 198) points out, Malay differs from the other NCSLs in his study in that comparative historical linguistics has not yet come up with a generally agreed upon classification of Malayic within the larger group of Malayo-Polynesian languages, that is to say, a set of nearest relatives that would correspond to, say, Germanic in the case of English, and to which Malay would be compared. McWhorter's solution was thus to delimit the set of languages geographically, to the western and central parts of the Indonesian archipelago. It would seem, however, that the net was cast rather too widely: a more plausible set of languages to compare Malay to would have excluded the four languages of more distant Sulawesi – Makassarese, Buginese, Tukang Besi

---

12. A first ominous sign is the misrepresentation of Karo Batak as "Daro Batak", not a mere typo as it is repeated throughout the chapter. (I have chosen to correct the spelling in Table 2 above.)

13. For demonstratives, I have not been able to substantiate McWhorter's claim that Malay has a two-way distinction while Minangkabau has a three-way one. In both languages, the norm is for a two-way proximal vs. distal distinction, such as *ini* vs. *itu* in Standard Malay, *iko* vs. *itu* in Standard Minangkabau, and corresponding forms in most colloquial varieties of the two languages. Indeed, the one case that I am familiar with of greater complexity (though there could easily be others that I am not aware of) is actually not in Minangkabau but in Jakarta Indonesian, which has innovated a three-valued paradigm consisting of *ini*, *itu* and *ono*. As for classifiers, McWhorter's figure of 3 for Malay but 20 for Minangkabau has little basis in reality. In general there is a tendency for classifier systems to be more elaborate in formal registers but less so in their more colloquial counterparts, however I am not aware of any sociolinguistically like-for-like comparative study that characterises Minangkabau as having a more complex classifier system than Malay, and impressionistically, at least, this does not seem to be the case. Thus, for example, the figure of 20 classifiers for Minangkabau is from Moussay (1981: 134–36), but most or all of Moussay's examples would carry over *mutatis mutandis* into a correspondingly high register of Malay.

14. Again, the basic demonstrative system of Javanese is two-way, though the actual forms differ considerably from one dialect to the other; the reported three-way distinction, with *iki*, *iku* and *ika*, occurs only in *wayang* (a genre of drama) and in some literature (Thomas Conners pers. comm.).

and Muna. Genealogically, the languages of Sulawesi have been argued to form one or more distinct subgroups within Malayo-Polynesian; thus Makassarese and Buginese are included in Adelaar's (1994) Greater South Sulawesi family, while Tukang Besi and Muna form part of Mead's (2003) Celebic family. Moreover, from a typological point of view, at least some of the languages of Sulawesi tend to bear a closer resemblance to the languages of the Philippines than do their counterparts in Sumatra, Java and Borneo.[15] Excluding Tukang Besi and Muna would have reduced the overall level of complexity of the languages in the table, and hence the contrast in complexity between Malay and its select sample of relatives. By the same token, it might have been more appropriate to exclude Nias, spoken off the west coast of Sumatra, in view of its proposed classification as belonging to a separate group of Barrier Island languages (Nothofer 1986, 1994), and also its apparent typological diversity. Thus, a more plausible geographically-defined set of languages to which Malay might be compared would perhaps have been languages spoken on the the three major islands, Sumatra, Borneo and Java; that is to say, the languages of Western Nusantara, as described in Section 4.4 below.

The most serious problem, however, with the data in Table 2, is that it raises the question of why all the other languages in the sample are also as simple as they are. Of the other 12 languages, two, Minangkabau and Iban, belong to the Malayic language family; as suggested by the data in Table 2, while perhaps not quite as simple as Malay, they are still simpler, on the whole, than the ten other non-Malayic languages in the sample. This poses a challenge to the characterisation of Malay as an NCSL. While we do not know when Minangkabau and Iban split off from Malay, a reasonable guess would be something in the range of 1000–2000 years ago, presumably more recent for Minangkabau, further back for Iban. This clearly precludes the possibility that the NCSL nature of Malay is due to language contact in the colonial era or later, because if that were the case, one would not expect to find that Minangkabau and Iban were also simpler than their non-Malayic counterparts. The possibility remains, however, that the NCSL character of Malay was acquired during the period of the Srivijaya empire, located in Sumatra around the 8th–12th

---

15. In particular, Tukang Besi is somewhat of a regional outlier with regard to a number of important typological features. Thus, for example, in the map of TAM marking in Gil (2015: 355, Map 9), Tukang Besi is clearly visible as almost the only language of the core Mekong-Mamberamo linguistic area with obligatory TAM marking. The only other such language is Inanwatan in the New Guinea Bird's Head, which is suggested by de Vries (2004: 13–16) to be a member of the Marind family originating far to the east. Otherwise, languages with obligatory TAM marking are only found in areas that are borderline Mekong-Mamberamo, such as northern Sulawesi and Borneo, the Bomberai peninsula of New Guinea, and Burma. Indeed, the exceptionality of Tukang Besi is acknowledged by McWhorter (2007: 206), who characterises it as the the region's counterpart to Icelandic.

centuries CE. Unfortunately, though, we just do not know enough about the linguistic landscape of those times, and whether Old Malay or whatever the language of the Srivijaya empire might have been had already split off from the ancestors of contemporary Minangkabau and Iban. The problem with the characterisation of Malay as an NCSL, then, is that it does not account for the simplicity that is evident across the Malayic family.

However, it's not just Minangkabau and Iban that are unaccountably simple. If we were to add English to Table 2, we would find that it is significantly more complex than Malay, exhibiting complexity in 6 out of the 9 features, comparable to languages such as Nias and Tukang Besi.[16] However, English, according to McWhorter (2007), is also an NCSL, which poses the question why it is so much more complex than Malay. This is something about which McWhorter's NCSL analysis has nothing to say. But the answer is obvious: in accordance with McWhorter's NCSL analysis, Malay started off like Nias and Tukang Besi, while English started off like Dutch and German – but Nias and Tukang Besi are themselves significantly less complex than Dutch and German. Thus, in the formation of NCSLs, the different end points of Malay and English are a direct consequence of their different starting points. It's all in the geography: in spite of their differences, English, Dutch and German are typical European languages, while Malay, Nias and Tukang Besi are within the ballpark of what is characteristic of languages of Nusantara. Indeed, as argued in Section 5 below, it is the geography, or more specifically, the various sprachbund effects resulting from multiple contact events spread out over space and time, that are the central factor underlying the simple typological profile of Riau Indonesian and its relatives, including Standard Malay/Indonesian.

Returning to Riau Indonesian and other similar contact varieties of Malay/Indonesian, McWhorter (2007) brings together his creolisation and NCSL stories, arguing that such varieties are the product of two consecutive waves of contact-induced simplification, the first producing the NCSL that is Malay, the second involving creolisation, resulting in Riau Indonesian and other Malay/Indonesian koinés. In McWhorter's words:

---

16. Specifically, English is more complex than Malay with respect to (a) its inventory of phonemes; (b) a (somewhat marginally) more complex demonstrative system, containing a singular/plural distinction; (c) the presence of copulas; (d) a complex system of negation, of the sort characterised by Miestamo (2008) as asymmetric; (e) the presence of an obligatory and elaborate system of tense/aspect marking; and (f) the presence of grammatical agreement, in subject-verb and noun-demonstrative environments. On the other hand, it is of equivalent complexity to Malay with regard to (a) the absence of infixes; and (b) the absence of overt imperative marking, and is actually somewhat simpler than Malay in that it lacks numeral classifiers entirely.

> Malay underwent two "passes" of reduction. The literary Malay of the courts and literature, while slightly more complex than modern Malay, was still unusually reduced compared to earlier [Indonesian-Type] languages. Presumably this resulted from the language's widespread and long-term use by non-native speakers simplifying it before it was committed to the written medium. Then, the colloquial Malays were reduced even further, as the result of continued heavy impact from non-native acquisition. (2007: 235)

> Riau Indonesian is a streamlining of a streamlining, resulting from two passes of non-native acquisition. (2019)

However, as we shall see below, there are two serious problems with McWhorter's scenario. The first is that there is no external evidence whatsoever for any recent event of creolisation in the history of Riau Indonesian and most of its relatives; in fact, the sociohistorical circumstances suggest that such an event is rather unlikely to have taken place. The second is that, as foreshadowed in the preceding paragraph, the simple typological profile and isolating structure of Riau Indonesian and other Malay/Indonesian koinés is hardly exceptional; rather, it is shared by a wide range of languages in the region, for which there is no reason whatsoever to associate their recent histories with either the formation of NCSLs or with processes of creolisation. These two problems are addressed in Sections 3 and 4 respectively.

2.4 Ethnicity, register and geography

In an attempt to go beyond the confines of categories such as PMDs, creoles and NCSLs, Gil (2001: 358–367) provided a preliminary attempt to account for the overall grammatical simplicity of Riau Indonesian, in terms of the following three factors:

(3) *Determinants of the Structure of Riau Indonesian*
    a. *Ethnicity*
       Riau Indonesian is so simple because it is a contact language
    b. *Register*
       Riau Indonesian is so simple because it is basilectal
    c. *Geography*
       Riau Indonesian is so simple because it is spoken in Southeast Asia

The above three factors are presented in increasing order of influence: ethnicity matters less than register, register less than geography. Table 3 below, from Gil (2001: 360), represents the above three factors in terms of three orthogonal binary distinctions: basilectal vs. acrolectal (horizontally within each 2 × 2 matrix), contact vs. non-contact (vertically within each 2 × 2 matrix), and Southeast Asia vs.

other regions (horizontally across the two 2 × 2 matrices). Each cell in the table is exemplified with a language, with boldface marking Riau Indonesian and the three languages differing from it minimally, with respect to just one of the three factors.

Table 3. Determinants of the structure of Riau Indonesian

Southeast Asia	basilectal	acrolectal	Other regions	basilectal	acrolectal
contact	**Riau Indonesian**	**Standard Indonesian**	*contact*	**Daghestani Russian**	Standard Russian
non-contact	**Siak Malay**	Standard Minangkabau	*non-contact*	Novgorod Russian	Standard Sorbian

As argued in Gil (2001), comparison of these three minimally-differing languages in boldface provides evidence for the relative importance of the three factors in (3). The fact that Standard Indonesian is more complex than Siak Malay suggests that the basilectal nature of Riau Indonesian is more significant than its nature as a contact variety; similarly, the fact that Daghestani Russian is more complex than Standard Indonesian suggests that the geographical location of Riau Indonesian is more significant than its basilectal nature.

McWhorter (2007, 2019) takes issue with all three factors proposed in (3). With regard to (3a) and the nature of Riau Indonesian as a contact variety, this is of course, according to McWhorter, the primary factor accounting for the simplified grammatical profile of Riau Indonesian, and not, as suggested in (3), the least important of the three factors. However, as pointed out above, the account proposed in Gil (2001) pertained only to the contrast between Riau Indonesian and other varieties of Malay/Indonesian, and as shown in detail in Section 4.2 below, Riau Indonesian is of roughly the same level of complexity as most other traditional varieties of Malay/Indonesian, and hence the significance of it being a contact language is at best minimal.

As for (3b) and the status of Riau Indonesian as an oral, basilectal language, McWhorter (2007: 225) points out that "most of the world's languages are used almost exclusively orally. And yet these include languages of awesome complexity in morphology and beyond: that is, all Native American and Australian languages and most Caucasian ones. Thus oral usage alone cannot explain why Riau Indonesian is so telegraphic." But of course it cannot, and it was never claimed that it could; on the contrary, as represented in (3), register plays second fiddle to geography: Amerindian, Australian and Caucasian languages may be as complex as they are precisely because they are spoken in other parts of the world, not in the Mekong-Mamberamo area. The point of the comparison to Standard Indonesian is to hold everything else equal and to factor out the effects of geography; once this is done, the effect of register and the basilect/acrolect distinction emerges clearly.

The differences between Riau Indonesian and Standard Indonesian are anything but fortuitous; in terms of morphological structure, examples abound of colloquial varieties being of lesser morphological complexity than their standard counterparts. Thus, in languages as diverse as Arabic, Sinhalese and Japanese, case marking is obligatory in the standard language but either optional or absent in the colloquial varieties. Work in progress, some preliminary results of which are presented in Gil and Shen (2019), points to a systematic and significant tendency for languages associated with greater polity complexity to also be endowed with greater grammatical complexity in the domain of compositional semantics and thematic role assignment; the proposed scale of polity complexity, ranging from small remote and egalitarian hunter-gatherer societies to world-language English subsumes, inter alia, the distinction between the polities associated with regional colloquial varieties of national languages and their corresponding nationwide standardised counterparts. This in turn dovetails with recent experimental work by Raviv, Meyer and Lev-Ari (2019) suggesting that larger societies tend to favour more systematic and complex grammatical structures than their smaller counterparts.[17]

Finally, with regard to (3c), McWhorter systematically downplays the relevance of geography and "sprachbund effects" to the formation of Riau Indonesian and its relatives. It should be acknowledged, though, that the sprachbund referred to in Gil (2001) and which McWhorter takes issue with is that of Mainland Southeast Asia, with respect to which Riau Indonesian occupies a rather peripheral position. In contrast, the present account makes reference to the Mekong-Mamberamo area, introduced in Gil (2015), with respect to which Riau Indonesian and its relatives occupy a privileged position, smack in the middle.

More generally, though, McWhorter's reticence with regard to sprachbund effects seems to be part and parcel of a more general agenda of his, arguing against language-mixing approaches to creole formation, such as the "feature pool" hypothesis proposed by Mufwene (2001), and echoed more recently in Blasi, Michaelis and Haspelmath (2017). I have no dog in this fight; if anything, my sympathies here tend towards the stance espoused by McWhorter. However, McWhorter appears to systematically misrepresent my position on Riau Indonesian, suggesting that I have characterised it as the product of some kind of language mixing. Thus, for example, he writes that "I sense that Gil means that […] language mixture was the main factor

---

17. Admittedly, such results run counter to an increasingly large body of literature, alluded to in footnote 11 above, including Dahl (2004, 2009), Nichols (2009), Sinnemäki (2009), and Trudgill (2009, 2011) as well as McWhorter himself, suggesting an opposite correlation, namely that larger languages tend to favour simplification rather than complexification. However, the two proposed correlations pertain to different grammatical features and are products of quite different causal mechanisms; hence there is no contradiction between the two.

distinguishing Riau Indonesian from the standard" (2007: 229–230). On the contrary, I have never claimed that Riau Indonesian is a mixed language; in fact, in Gil (2009b) I argue explicitly against such a claim. Once again, the issue is one of timing: McWhorter seems to be interpreting my appeal to geography and sprachbund effects as pertaining to the relatively recent history of Riau Indonesian, as though perhaps it were a recent arrival to the region such as, say, Singlish, which has abruptly "gone Asian" (see Gil 2003). However, as argued in Gil (2015) and in this chapter, the relevant sprachbund is the Mekong-Mamberamo area, which is of great antiquity; moreover, the presence within it of Austronesian languages is known to date back roughly 3500–4000 years. As argued in Donohue and Denham (this volume), different contact scenarios may have played out at different times and locations, in the course of the Austronesian expansion into the region. Thus, the sprachbund effects that underlie the isolating profile of Riau Indonesian are themselves ancient; moreover, there is no particular reason to attribute them to language mixing.

The account proposed in Gil (2001) and summarised in (3) provides a preliminary basis for the more fleshed-out story of Riau Indonesian and its relatives developed in Section 5 below. However, the story developed below differs from that of Gil (2001) in two respects. First, whereas the focus of Gil (2001) is on the recent history of Riau Indonesian within Malay/Indonesian, the present story goes back much further in time and casts a wider net over a substantially larger number of languages. Secondly, whereas the sprachbund referred to in Gil (2001) is that of Mainland Southeast Asia, the current version of the story is couched in terms of the significantly larger Mekong-Mamberamo linguistic area.

In order to account for the development of the isolating profile in Riau Indonesian and its relatives, we must first lay the necessary empirical foundations. Accordingly, Sections 3 and 4 survey the sociohistorical and linguistic landscapes that are of relevance to the history of Riau Indonesian and other related languages. As a point of logic, these two surveys proceed independently of one another: the sociohistorical survey in Section 3 makes no reference to particular linguistic features, while the linguistic discussion in Section 4 is expressly a-historical. It is only once the sociohistorical and linguistic facts are independently established that one can attempt to tie them together in order to seek correlations between the two, and propose hypotheses accounting for the linguistic facts in terms of the sociohistorical ones.[18]

---

18. In principle the above methodological point should be obvious and in no need of belabouring; however, in practice it is often violated. As captured in McWhorter's (2001c) title "What people ask David Gil and Why", when people encountering a description of Riau Indonesian wonder whether it's a creole, they are assuming that if a language exhibits a grammatical profile of a sufficient degree of simplicity, then it is likely to be, or indeed must necessarily be, a creole language. However, this assumption is promoted by McWhorter from an empirical hypothesis to an article

3.   The sociohistorical landscape

As we go back in time, our knowledge of the sociohistorical facts falls off rather dramatically, especially with regard to the milieus that are of greatest relevance to us here, namely those associated with colloquial speech varieties. Given that the very existence of Riau Indonesian and other western Indonesian koinés has only come to be acknowledged over the course of the last few decades, it is hardly surprising that we have few if any records of the sociohistorical landscape of the region in the past. What written texts we have, dating back to perhaps the 7th century, are presumably all in an acrolectal variety of Malay, shedding little light on how the different varieties of colloquial Malay were actually spoken across the region. Accordingly, for the most part, the best we can do is to observe the present, and try to make reasonable inferences from what we see now to what must have been the case in the past.

At the beginning of the 21st century, the most salient fact is that one or more varieties of Malay or Indonesian are spoken by an overwhelming majority of the roughly 300 million inhabitants of Malaysia and Indonesia. An obvious if unfortunate measure of this fact is provided by the efforts to which linguists working in the region sometimes have to go in order to find reliable speakers of the over 800 other languages spoken in Malaysia and Indonesia. Of crucial importance, though, is the proportion of Malay/Indonesian speakers who speak it natively, as a first language. While no accurate estimates are available, it would seem that a substantial majority of the populations of both Malaysia and Indonesia do indeed qualify as native speakers of at least one variety of Malay/Indonesian.[19]

---

of faith, repeated constantly throughout his work. In justification, McWhorter (2008: 188) offers an analogy from geology: "As geologists treat cracked quartz as a sign of volcanic eruptions in the past, linguists might treat […] simplicity […] as evidence of social disruption in the past." Appealing as the analogy may be, it rests crucially on the assumption that the natural, default state of quartz is in whole crystals, and that whenever we find a piece that is cracked, or broken, well then something must have happened to it. While this makes sense for quartz, it is less obvious that it carries over also to languages. Indeed, as suggested in Section 5.3 below, simplicity may actually constitute a default state for language, or at the very least, a potential pole of stability around which languages may coalesce.

19. A radically divergent estimate is given by Prentice (1987: 915), who asserts that just 45% of Malaysians and 7% of Indonesians speak Malay/Indonesian as a "mother tongue". Based on my own experience and impressions, I find this figure dubious for the 1980s and pretty much incredible for today's Malaysia and Indonesia – in most parts of which first-language mastery of at least one variety of Malay/Indonesian is the rule rather than the exception. I do not know the reasons behind Prentice's much lower estimate: it may have to do with a more restricted definition of what counts as "Malay", or, alternatively, as a mother tongue. (It is important to keep in mind the "a" in "a mother tongue", given that many speakers in the region have native competence in one or more varieties of Malay/Indonesian alongside one or more other languages.)

In order to be able to look back in time, a primary distinction must be drawn between two geographically-based categories of Malay/Indonesian varieties: those spoken in regions where the indigenous language varieties are themselves Malay/ Indonesian or otherwise Malayic, and those spoken in regions where the indigenous languages are non-Malayic. The first category, which we shall refer to as *heartland* varieties, comprises Malay/Indonesian varieties spoken in a large swathe of central Sumatra, some coastal regions of Borneo, and most of the Malay peninsula. It includes both traditional dialects such Siak Malay, Jambi Malay, Brunei Malay and Ulu Muar Malay, as well as contact varieties and koinés such as Riau Indonesian, Sumatra Barat Indonesian, Kalimantan Barat Indonesian, and Kuala Lumpur Malay. The second category, which we shall refer to as *transplanted* varieties, consists of Malay/Indonesian varieties spoken everywhere else. These are for the most part contact varieties of various kinds, including the likes of Sri Lankan Malay, Nonthaburi Malay, Sabah Malay, Lampung Indonesian, Jakarta Indonesian, Makassarese Malay, Kupang Malay and Papuan Malay.[20] Since the histories of these two categories of Malay/Indonesian varieties are quite different from each other, we shall consider each of these two categories in turn.

3.1   Heartland varieties

Beginning with the heartland varieties, the most salient fact is that the regions in which they are spoken have remained predominantly or exclusively Malay or Malayic speaking for the better part of the last two thousand years. While in central Sumatra the previously present non-Malayic languages were replaced early and without trace, in the Malay peninsula and coastal Borneo, the process is still under way, with Austronesian Dayak and non-Austronesian Aslian languages respectively still present, albeit in retreat. In addition, as far back as we know, there has been migration into the region from other parts of the archipelago, by speakers of Javanese, Madurese, Buginese and many other languages. As a rule, however, such immigrants abandon their original languages and switch to the local variety of Malay or Indonesian.

Of course, the above-mentioned processes give rise to a substantial amount of language contact, and clearly, such language contact is likely to have triggered a certain amount of contact-induced simplification. However, there is no historical evidence whatsoever for any circumstances that might be associated with

---

20. It should be noted that the distinction between heartland and transplanted varieties of Malay/ Indonesian proposed here cross-cuts two other important classificatory distinctions between Malay/Indonesian varieties that are of importance in other contexts: that between traditional and contact varieties, and within the class of Indonesian koinés, that between western and eastern varieties.

substantial imperfect adult second-language acquisition of a variety of Malay that might accordingly merit characterisation as a Malay-based creole. In the context of Riau Indonesian, McWhorter (2019) argues the case for creolisation by citing my description, in Gil (2001), of a study showing that "one in four Riau Indonesian speakers grew up in homes where at least one parent was not a native Malay speaker, that, 'the present-day Riau province was the venue of substantial language contact over much of the last 2000 years,' that 'various contact varieties of Malayic must have arisen during this lengthy period,' and that 'such contact varieties constitute plausible ancestors for what is now Riau Indonesian.'" For sure, there was contact, most likely leading to contact varieties; however, the point being made here is that we have no evidence that such contact varieties involved the kind of widespread imperfect adult second-language acquisition that would lead to radical restructuring of the kind associated with creolisation.[21]

However, what we are faced with here is not just the absence of positive evidence for creolisation; it is also the lack of plausibility for such events. The sociohistorical landscape of the region over the past one or two thousand years is simply not one that is conducive to the development of Malay based creoles, and the reason for this is quite straightforward – the overall predominance of Malay. Admittedly, one can imagine any number of hypothetical situations that might have given rise in the past to a Malay-based creole, and in fact one can even observe analogs in the present. One scenario would involve speakers of an indigenous non-Malayic language switching to Malay; thus, Uri Tadmor (pers. comm) suggests that the Malay of Balai Berkuak and similar downstream varieties in western Kalimantan exhibit a strong (non-Malayic) Land Dayak substrate, manifest mostly in the syntax and lexical semantics, indicative of a significant shift from Land Dayak to Malay in the past. Similar situations can also be observed in the Malay peninsula, as populations switch from Aslian languages to Malay. A second scenario might involve migrants from outside the region developing a creolised version of Malay, as has been suggested to be the case for Baba Malay by Ansaldo and Matthews (1999) and others. Crucially, however, such localised creoles or creole-like varieties stand no chance of "taking over" Malay/Indonesian as a whole; on the contrary, they tend

---

21. McWhorter's citation of the study showing that "one in four Riau Indonesian speakers grew up in homes where at least one parent was not a native Malay speaker" calls for further qualification. The 25% figure cited in Gil (2001: 330) "represents the percentage of households in the Riau of some 50 years ago in which at least one of the core members was an immigrant". What is perhaps not clear from the above description is that in well over half of these cases, the non-Malay immigrant in question is of another Malayic-speaking ethnicity, most commonly Minangkabau. In such cases, because of the similarity between the language varieties, communication within the household would typically take place in a mixture of Malay and the other Malayic language, with all members of the household being fluent in both varieties, as a result of which the contact language, Riau Indonesian, would tend to be kept for communication of a less intimate nature.

to remain localised, threatened, and ultimately replaced by more general forms of Malay/Indonesian. Thus, today's Balai Berkuak Malay is threatened by Kalimantan Barat Indonesian, while Baba Malay is on the verge of being replaced by Kuala Lumpur Malay, which has formed the basis for a koiné spoken throughout the Malay peninsula and beyond. Presumably, similar situations to these might have arisen time and time again in the past; however, whatever local creoles may have arisen as a result, they would have soon been absorbed with little or no trace within the predominant Malay.

In order to appreciate how unlikely it is that any of the heartland Malay varieties, traditional or koiné, have a recent history of massive imperfect adult second-language acquisition of the kind that might give rise to a creole, one may compare the sociohistorical landscape of the Malay-speaking heartland with that of another, perhaps better-known part of the world, namely Germany plus the German- and Dutch-speaking regions of neighboring countries. The similarities, both contemporary and historical, between these two disparate parts of the world are numerous and substantial. Compare the contemporary linguistic landscape of, say, Pekan Baru, the capital of Riau Province, with that of a northern German city such as Bremen. In both cities, the basic situation is multiglossic, involving the standard language, one or more local dialects, and, in-between the two as it were, a regional contact variety – Standard Indonesian corresponding to Standard German, the local varieties of Riau Malay such as Siak Malay to the local varieties of Low German, and Riau Indonesian to the local variety of High German. In addition, both cities are host to large migrant populations, be it economic migrants from West Sumatra, Java, and other parts of Indonesia in Pekan Baru, or recent refugees from Syria, Iraq and Libya in Bremen – with the incoming populations ultimately assimilating to the local linguistic landscape. In contrast, in both cities, there is, or was until recently, an older non-indigenous population which developed its own distinct variety of the local language: in Pekan Baru, the distinct Malay/Indonesian variety of the local Chinese community (referred to in Gil 2001 as Outsider Malay), and in Bremen, the Germanic dialect of the Jewish community, namely, Yiddish. Head out of town and in both cases, more divergent dialects or closely related languages will be encountered: Sakai and Akit to the north of Pekan Baru, Frisian to the north or west of Bremen. Switch on the television and a change in channels will reveal programs in a variety seemingly more closely related to the local indigenous dialect, be it Riau Malay or Low German, except that it will actually be broadcasts in the similar standard language of a neighboring country, Standard Malay or Dutch.[22]

---

22. Of course, like all analogies, it only goes so far, and several differences must be acknowledged. First, the role of Standard Indonesian in everyday communication is much smaller than that of Standard German. Related to this, the local koiné, Riau Indonesian, is substantially more

The above-mentioned analogies characterising the two contemporary linguistic landscapes may be extrapolated back in time. Most importantly, for a substantially large area, be it central Sumatra for Malay or the historical area of the Francs and Saxons in the case of German, the present-day dominant languages, Malay/Indonesian or German, were in place as the dominant languages for the past couple of millennia – ever since they first replaced the other languages that were there first. As such, there was never any context in which it would be likely for a significant population of adults to acquire the local language, either Malay or German as a second language, and, crucially, pass on their imperfectly acquired version to subsequent generations, until it ultimately replaced the earlier variety of Malay or German that was there before. To be sure, communities of migrants might have developed Malay- or German-based pidgins, but these would never have been given the chance to stabilise and undergo creolisation; once the migrants became integrated into the local communities, the pidgins would simply give way to the generally spoken versions of Malay or German. Thus, there was never any situation in which a creolised version of Malay or German could emerge and then gain currency to become a stabilised and widely spoken language variety.

For German this is hardly controversial; nobody to the best of my knowledge has ever argued that some variety of German spoken in Germany or neighboring German-speaking countries is a creole.[23] But as suggested here, on purely sociohistorical grounds there is no more reason to suggest that Riau Indonesian, Siak Malay and their ilk are any different from their counterparts in northern Germany. There is no escaping the conclusion that to the extent that McWhorter and others characterise Riau Indonesian and other heartland Malay/Indonesian varieties as creoles, it is – however much they may insist otherwise – a claim based not on independent sociohistorical evidence, but rather motivated by the simple grammatical profile exhibited by these language varieties, and the assumption that such simplicity can only be reflective of a creole origin.

---

divergent from its acrolect than is its counterpart the High German variety of Bremen. Another difference is that the local variety of Riau Malay is considerably more widely spoken than the local dialect of Low German which is endangered. Yet another difference is that whereas the migrants in Pekan Baru are all from other parts of the same country, sharing similar colloquial varieties of Indonesian, those in Bremen are from all over the world, many starting out with little or no knowledge of the local language. And there are more differences, however, they do not detract significantly from the force of the analogy.

23. Perhaps the closest that anybody has come to such a proposal is Wexler's (2002) characterisation of Eastern Yiddish as being a relexification of a Slavic language, Judaeo-Sorbian. But Eastern Yiddish was not spoken in Germany or neighbouring German-speaking countries, and besides, this proposal has been met with near-universal opprobrium.

3.2   Transplanted varieties

Turning now to the transplanted varieties of Malay/Indonesian, these in turn are of two different types. The first consists of varieties spoken by an ethnic minority in countries whose national language is not Malay/Indonesian; the leading exemplars of this type are Sri Lankan Malay and various Malay dialects spoken in Thailand such as Patani and Nonthaburi Malay. In both cases, substantial contact-induced restructuring has taken place; see Nordhoff (2012) for Sri Lankan Malay and Tadmor (1995) for Nonthaburi Malay. However, the circumstances of their formation are very different from each other as well as from other transplanted varieties, so we shall not have anything more to say about them here.

Our focus, instead, is on the second type of transplanted varieties, those spoken in countries whose national language is Standard Malay/Indonesian, for example Sabah Malay, Lampung Indonesian, Jakarta Indonesian, Makassarese Malay, Kupang Malay and Papuan Malay, to name just a few. Simplifying somewhat, each transplanted variety has a history that can be broadly partitioned into three stages, a first original stage during which its earlier ancestors were spoken in the Malay heartland region, a possibly quite short second transitional stage when one of its ancestors was spoken both within the Malay heartland and outside of it, and a third stage, stretching up to the present, in which its more recent ancestors were spoken in or near the current location of the language.

With regard to the first original stage, during which the ancestors of the transplanted varieties were spoken in the Malay heartland, their sociohistorical circumstances are identical to those of the current heartland varieties considered in Section 3.1 above, and there is therefore no reason to suspect that, during that stage, they underwent the kind of radical restructuring characteristic of creole languages. It is, rather, the second and third stages that merit our attention here.

Of all the sociohistorical situations associated with the history of Malay/Indonesian over the last two millennia, it is perhaps the second stage, that in which the ancestors of transplanted varieties of Malay/Indonesian were spoken over a wide area encompassing both heartland and non-heartland regions, that provides the most plausible sociohistorical background for some kind of radical restructuring or creolisation. The most likely context for such possibly widespread imperfect adult second-language acquisition of Malay would be a Malay trade language, as indeed is commonly argued to constitute an ancestor for eastern Malay varieties such as Kupang Malay, North Maluku Malay and Papuan Malay, see for example Paauw (2009). Still, even if it were such a trade language that brought Malay to what is now eastern Indonesia, it is not necessarily the case that it involved a sufficient degree of imperfect adult second-language acquisition to qualify as a creole – we simply do not know. Moreover, it is not clear to what extent such a trade-language scenario is

at all relevant to other locations in which a transplanted Malay/Indonesian koiné is spoken, such as, to name a random set of examples, Mentawai, Central Java and Sabah. In summary, then, in the case of the stage-two ancestors of transplanted Malay varieties, a creolisation story is perhaps less implausible than in other cases, but once again, there is still no positive evidence to the effect that anything like that actually happened.

It is the third stage, leading up to the present, for which we have more direct information about the relevant sociohistorical landscape. And here, as before, there is no positive evidence that I am familiar with of any kind of recent second-language acquisition by adults of the type that is conducive to creolisation.

Once again, it is helpful to draw parallels with other parts of the world, this time between the transplanted Malay/Indonesian varieties and other major languages spoken outside their homelands. Three cases, all involving a transplanted major language spoken as a lingua franca in a region of high linguistic diversity, are Russian in the Russian republic of Daghestan, Swahili in upcountry central and western Kenya, and Spanish in those regions of Guatemala in which indigenous Mayan languages are still widely spoken. The first case, that of Daghestani Russian, was already discussed in Section 2.4 above, albeit in the context of Riau Indonesian, a heartland variety of Malay/Indonesian. In fact, though, it is the transplanted varieties of Malay/Indonesian that present a closer parallel to Daghestani Russian, to the extent that, in both cases, there is the same amount of evidence for creolisation: none. McWhorter (2001c: 405–6) contests the analogy, arguing that Daghestanis have traditionally been taught Russian in school, which presumably assures that most speakers attain a high level of acquisition. In contrast, he asserts, until a few decades ago, the way most Southeast Asians acquired Malay was "on the fly, through casual, oral contact"; this point is made once again in McWhorter (2007: 269). In actual fact, however, the situation in Daghestan bears a greater resemblance to that in places such as Sabah and Papua than is acknowledged by McWhorter. In a study of three rather remote villages in Daghestan, Daniel, Dobrushina and Knyazev (2010: 67), report that in two of them, children "have some command of Russian even before they go to school". As for the urban centres, they note that many persons are actually monolingual speakers of Daghestani Russian, so obviously it is being acquired in completely natural circumstances, and not dependent on formal schooling. Similarly, pre-school acquisition by children is also reported by Michael Marlo (pers. comm.) for Swahili in upcountry Kenya, and by Pedro Mateo Pedro (pers. comm.) for Spanish for at least some children in the Mayan-speaking regions of Guatemala. What these analogies suggest, then, is that on purely external sociohistorical grounds, there is no less reason to suppose that the Russian of Daghestan, the Swahili of upcountry Kenya, and the Spanish of Mayan-speaking areas of Guatemala are creoles, than there is to suggest that various transplanted

varieties of Malay/Indonesian are creoles. But the obvious fact is that these varieties are of comparable complexity to the corresponding ordinary stay-at-home varieties of Russian, Swahili and Spanish respectively – and it's not schooling that is responsible. Once more, there is no avoiding the conclusion that the proposed characterisation of transplanted Malay/Indonesian varieties such as Sabah Malay and Papuan Malay as creoles is driven solely by their simple grammatical profile and the assumption that such simplicity can only be indicative of a history of creolisation.

Why then are Riau Indonesian and its relatives so much simpler than their sociohistorical analogues from other parts of the world, such as the High German of Bremen, the Russian of Daghestan, the Swahili of upcountry Kenya, and the Spanish of Mayan-speaking areas of Guatemala? If their recent and not so recent sociohistorical circumstances provide no answer, then we need to look even further back in time, which can only be done by adopting a broad areal-linguistic perspective.

4.   The linguistic landscape

With the sociohistorical landscape independently established, we now turn to consider the linguistic facts. In accordance with the approach laid out at the outset of this chapter as represented in Figure 1, we take Riau Indonesian as our starting point, in order to examine increasingly large and mostly nested networks of dialects and languages, whose shared features can be reconstructed to increasingly far back points in time. Evidence is provided for the following 6 factual observations:

(4) a. Riau Indonesian is of roughly the *same* degree of complexity as other Malay/Indonesian koinés;
   b. Malay/Indonesian koinés are at most only *moderately less* complex than other varieties of Malay/Indonesian;
   c. Malay/Indonesian varieties are at most only *moderately less* complex than other Malayic varieties;
   d. Malayic varieties are at most only *moderately less* complex than other languages of Western Nusantara;
   e. Languages of Western Nusantara are at most only *moderately less* complex than other languages of the Mekong-Mamberamo region;
   f. Languages of Western Nusantara are *substantially less* complex than other Austronesian languages, and in particular, those of Taiwan and the Philippines.

The observations in (4) make reference to a general notion of complexity, of which morphological complexity and word structure are but one aspect, albeit a major one. In the following discussion, the focus is on word structure and the extent to

which the dialects and languages in question are isolating, however, in order to come up with an adequate assessment of isolating structure, a broader examination of grammatical patterns may sometimes be necessary.

As argued in Gil (this volume, Chapter 1), in order to determine whether and to what degree a language is isolating, a detailed and painstaking exploration of word structure is required. Unfortunately, many grammatical descriptions simply adopt a transcription parsing their texts into orthographic words, with little or no attempt to provide explicit grammatical arguments in support of their assignments of wordhood. To be sure, such proposed orthographies may be based on reasonable albeit implicit grammatical intuitions with regard to what constitutes a word in the language; however, there are invariably distortions. Suffice it to recall the example cited in the introduction to Gil (this volume, Chapter 1), in which two respectable sources assume different assignments of wordhood for the exact same string in Papuan Malay, with Kluge's (2014: 377) *sa pu bapa* painting the language as isolating, Donohue and Sawaki's (2007: 260) *sa=pu=bapa* making it seem more polysynthetic. What follows, therefore, is a necessarily impressionistic overview of the relative degree of morphosyntactic complexity of the various language networks referred to in (4).[24]

4.1   Riau Indonesian and other Malay/Indonesian koinés

As suggested in (4a), Riau Indonesian exhibits a degree of isolating structure that is more or less on a par with that of other Malay/Indonesian koinés. What this says is that an analysis of word structure of any other such variety, along the lines proposed for Riau Indonesian in Gil (this volume, Chapter 1), would reveal a picture that is similar in its overall shape. Of course, there would be numerous differences of detail, but these would not significantly challenge the overall characterisation of such varieties as sharing the isolating profile of Riau Indonesian.

Given that many of the bound morphemes in Riau Indonesian are shared with other Malay/Indonesian koinés, one may ask whether the degree to which they are bound, for example whether they are weakly or strongly bound in the sense of Riau Indonesian as per the analysis in Gil (this volume, Chapter 1), is the same across the different varieties. One source of variation is evidenced by focus intonation, a pragmatically-conditioned process which assigns suprasegmental prominence to two syllables, together constituting a core disyllabic foot, which, in the prototypical

---

24. The present survey is based on a convenience sample of dialects and languages that I happen to have greater access to and familiarity with. There is no reason, however, to believe that the sample is biased in any particular substantive direction.

case, is coextensive with a disyllabic monomorphemic word – see Gil (this volume, Chapter 1, Section 3.3.2). At issue is what happens when the form in question bears host to the associative marker *-nya* or the end-point-oriented generalised voice marker *-kan*. In Riau Indonesian, the prominence remains on the last two syllables of the host; this is argued to show that *-nya* and *-kan* fall outside of the core foot, supporting their characterisation as weakly bound morphemes. In contrast, in Jakarta Indonesian, when *-nya* or *-in* (the counterpart of *-kan*) are added, focus intonation shifts one syllable to the right, to include the added marker. Prima facie, this would seem to suggest that these two markers are more strongly bound in Jakarta Indonesian than in Riau Indonesian, which in turn would provide one small piece of evidence in favour of Jakarta Indonesian being just a little bit less isolating than Riau Indonesian. However, this can only be safely concluded within the framework of a more extensive investigation of Jakarta Indonesian, one that would consider other available criteria for wordhood.

With regard to the end-point-oriented generalised voice marker at least, Manado Malay would seem to differ from Riau Indonesian in the opposite direction. In Manado Malay, the cognate form *akang* is disyllabic, and is able to bear focus intonation on its own, as a separate word. Again, a more extensive investigation is called for, however, on the face of things, it would appear that Manado Malay *akang* is less strongly bound to its host than its Riau Indonesian counterpart, suggesting that if anything, Manado Malay may be a tad more isolating than Riau Indonesian.[25]

Somewhat easier to analyse are the more straightforward cases in which an apparent bound morpheme is present in some varieties but absent in others, its function instead being expressed lexically or periphrastically, or, as is often the case, remaining unexpressed. For example, the Riau Indonesian agent-oriented generalised voice marker *(me-)(N-)* is present, in one form or another, across the western Indonesian koinés, and indeed, impressionistically at least, its use is even more common in Jakarta Indonesian, though even there it is generally optional. However, it is largely absent from the koinés of Malaysia and eastern Indonesia, except for a few frozen forms (e.g. Papuan Malay *menyimpan* discussed in Section 2.2 above); the words that would otherwise bear host to *(me-)(N-)* instead occur in bare unmarked form. Similarly, the Riau Indonesian patient-oriented generalised voice marker *di-* occurs throughout the western Indonesian koinés but is rare or absent from many of their Malaysian and eastern Indonesian counterparts. Even in Riau Indonesian and other western Indonesian koinés, the use of *di-* is largely optional, its functions expressible instead through other means, such as word order

---

25. Unlike its counterparts in Riau and Jakarta Indonesian, in Manado Malay the form *akang* has only an applicative function, not a causative one. Also, unlike in Riau and Jakarta Indonesian, the the associative marker *-nya* is rare or absent.

permutations or the periphrastic passive construction with *kena*. It is these alternative means that are used also in varieties that do not make use of *di-*, such as, for example, periphrastic passives with *kena* in Kuala Lumpur Malay, *dapa* in Papuan Malay and either *kena* or *dapa* in Kupang Malay.

The distribution of the various end-point-oriented generalised voice markers exhibits a somewhat more complex distribution, though also with a rough east-to-west cline encompassing the koiné varieties of both Indonesia and Malaysia. A further element of complexity is that there are several distinct forms in competition. Riau Indonesian offers two alternatives, basilectal *-kan* and somewhat more mesolectal *-in*, the latter most probably a borrowing from Jakarta Indonesian, ultimately from Balinese (Kahler 1966:i; Ikranagara 1980: 137). These two forms compete with each other in complex and intricate ways throughout the western Indonesian archipelago: some places, such as Malang, have only *-kan*, others, such as Jakarta, only *-in,* while yet others, such as Medan have both. In addition, a third form, *-i*, occurs sporadically in the localised koinés in a few places such as Rantau Prapat on the east coast of north-central Sumatra, Bengkulu on the west coast of southern Sumatra, and Makassar in southern Sulawesi. Finally, some varieties have no end-point-oriented generalised voice marker whatsoever, for example, Tapan Indonesian, spoken in the eponymous town in West Sumatra, as well as most eastern Indonesian koinés, such as Kupang Malay and Papuan Malay. Again, a number of alternative strategies expressing similar functions are available either alongside or in place of an end-point marker. For the causative function, the most common strategy is the periphrastic construction with 'give', such as *bagi* in Kuala Lumpur Malay, or *kasi* in Kupang Malay, Papuan Malay and many others. For the applicative function, for benefactives at least, the most widespread strategy is with a preposition *untuk* 'for', available in Riau Indonesian, Kuala Lumpur Malay, Papuan Malay and numerous others. An alternative strategy, present in Kupang Malay, involves the use of *kasi* 'give' but in a different construction to the causative *kasi*: whereas causative *kasi* precedes its host, as in *kasi masok* (give enter) 'put in', applicative *kasi* follows it, as in benefactive *beli kasi* (buy give) 'buy for'. For certain specific words, an alternative lexical strategy may also be available. For example, for the most common way of saying 'turn off' (in the context of an electrical appliance), where Riau Indonesian makes use of an end-point-oriented generalised voice marker in *matikan* or *matiin* (die:EP), Papuan Malay makes use of a simple word *padam* 'extinguish', Southeast Maluku Malay uses *bunuh* 'kill', while Kuala Lumpur Malay uses the English loan word *off*. A somewhat more complicated example of variation involving both periphrastic and lexical alternatives involves the ways of saying 'do what'. As noted in Section 2.2 above, in Riau Indonesian one common way of expressing the notion is with *ngapain*, consisting of the question word *apa* followed by the end-point marker *-in* (and also preceded by the agent-oriented marker *(me-)(N-)*). In contrast, the

corresponding expression is *buat apa* (do what) in Sabah Malay and *bikin apa* (do what) in Papuan Malay. Finally, like in Riau Indonesian, in most or all other koiné varieties, the zero-marking option is also available; see Conners, Bowden and Gil (2015) for Jakarta Indonesian.

The preceding discussion of the generalised voice markers *(me-)(N-)*, *di-* and *-kan* portray Riau Indonesian as, if anything, a little bit less isolating than at least some of the other Malay/Indonesian koinés. A few other examples could be adduced pointing towards a similar conclusion. Thus, the Riau Indonesian personal marker *si-* (see Gil this volume, Chapter 1, Section 3.3.6), while present also in Jakarta Indonesian and Sabah Malay, is absent from eastern Indonesian koinés such as Kupang Malay and Papuan Malay. Similarly, the Riau Indonesian reduced form *se-* 'one', used, among others, in construction with numeral bases, as in *seratus* (one:hundred), *seribu* (one:thousand), occurs in this construction in most other Malay/Indonesian koiné varieties but not in Sabah Malay, where instead the full form of the numeral 'one' is used, as in *satu ratus* (one hundred), *satu ribu* (one thousand). However, by the same token, there are also a number of morphological features that point in the other direction, towards Riau Indonesian as being somewhat more isolating than at least some of the other Malay/Indonesian koinés; such features involve morphological constructions in other varieties that are not present in Riau Indonesian.

One such example is that of the suffix *-an* in its use as a comparative marker: whereas in Riau Indonesian, and in fact most other Malay/Indonesian koinés, comparative constructions are formed periphrastically, typically with a word such as *lebih* 'more', in Jakarta Indonesian, the suffix *-an* is used productively for this purpose. A second example is that of the depatientive generalised voice marker *ber-* (and its cognates), already discussed in Section 2.2 above. In several eastern Indonesian koinés, *ber-* seems to be much more commonly used than in Riau Indonesian, both in terms of its distribution across word types, and, for specific word types, across word tokens. Some examples of words that rarely or never take *ber-* in Riau Indonesian but do so commonly or even obligatorily in Papuan Malay include *babingung* 'confused', *bataria* 'scream', *bakalai* 'fight' and *baduri* 'thorny'; their Riau Indonesian equivalents would be *bingung, teriak, kelahi* and *duri* respectively. A third example is that of a separate series of enclitic pronouns that occur after their hosts in a number of different constructions of which the most common is to mark possession. Such enclitic pronouns are absent from Riau Indonesian, as well as from Kuala Lumpur Malay and most eastern Indonesian koinés such as Kupang Malay and Papuan Malay; however, they are present in Sumatra Barat Indonesian, which is otherwise very close to Riau Indonesian, as well as in Jakarta Indonesian, Makassarese Malay, Sabah Malay and elsewhere – for example Sumatra Barat Indonesian *bukuku* (book:POSS.1SG) 'my book'.

Another potential case of morphological complexity, also involving pronouns, specific to Papuan Malay, is that of a nascent system of argument indexation (in the sense of Haspelmath 2013), as represented in the following table:

Table 4. Argument indexation in Papuan Malay

ko pukul (2sG see) 'you hit' agent/experiencer/theme	*pukul ko* (see 2sG) 'hit you' patient/stimulus
ko anjing (2sG dog) 'your dog' possessor	*anjing ko* (dog 2sG) 'you dog' essant

Table 4 above illustrates a generalised notion of argument indexation in accordance with two independent dimensions. Rows classify argument indexation in accordance with the category of the host, either an activity-denoting word such as *pukul* 'hit' or a thing-denoting word such as *anjing* 'dog'. Columns distinguish between the position of the pronoun, here exemplified with the 2nd person singular *ko*, either before or after its host word. The four cells thus represent four different subcases, or subconstructions, of a generalised argument-indexation construction, distinguished by their thematic role assignments; within each cell, the second line specifies the thematic role typically assigned to the pronoun in that particular configuration. Whereas for activity-denoting-word hosts, these thematic role assignments reflect the basic agent-activity-patient word order, for thing-denoting-word hosts, the left-hand cell represents a preposed possessor construction, while the right-hand cell represents a construction shared with several other languages in various parts of New Guinea, in which words denoting things are marked for person, number and in some languages also gender.[26]

In Table 4, the pronominal forms are written separately, in accordance with the convention, proposed in Gil (this volume, Chapter 1, end of Section 2.3), to the effect that in the absence of positive evidence to the contrary, the default decision should be to write each morpheme separately. However, at least in the case of pronouns preceding their hosts, as in the left-hand column of the table, there is a modest amount of paralinguistic evidence suggesting that they may in

---

26. The bottom row presupposes an enrichment of the traditional notion of thematic role. While the role of possessor may be considered as representing a default maximally-underspecified thematic role, the role of essant, introduced in Gil (2013), is that cross-linguistically prototypically instantiated by the subject of a predicate nominal construction; thus, for *anjing ko* above, *anjing* assigns the essant role to *ko*, thereby asserting that the referent of the 2nd person singular pronoun *is* a dog.

some instances be part of a larger word including also their following hosts.[27] One source of evidence comes from naturalistic spelling: while pronouns are most often written as separate words, in a minority of cases they may be written joined on to their following hosts, as in *alapar* (1SG:hungry) 'I'm hungry', from a corpus of SMS messages. A second source of evidence is the Bahasose ludling (named after how it applies to the word *bahasa* 'language'). The Bahasose ludling applies to each word as a whole, replacing the final disyllabic foot $-C_1V_1C_2V_2C_3$ with the sequence $-C_1V_1C_2oC_2eC_3$ (where C stands for zero, one or two consonants and V for zero or one vowels). Examination of a small corpus of naturalistic speech in the Bahasose ludling suggests that in most cases, pronouns are considered as separate words; however, in a small number of examples, they are treated as joined on to their following hosts, for example *desudode*, derived from *desuda* (3SG:PFCT) 'He has already'. Thus, it would seem to be the case that the Papuan Malay pronouns, or at least the short forms of the pronouns when occurring in front of their hosts, may be in the process of becoming clitics.[28]

In, summary, then, a comparison of word structure in Riau Indonesian and other Malay/Indonesian koinés reveals some cases of greater complexity in Riau Indonesian and other cases of greater simplicity. However, all of these differences are of relatively minor import; the overall picture that emerges is one in which, with respect to word structure, Riau Indonesian and the other Malay/Indonesian koinés exhibit broadly similar levels of complexity, sharing a similarly isolating typological profile.

4.2   Malay/Indonesian koinés and other varieties of Malay/Indonesian

Broadening the vista from the koinés to other more traditional varieties of Malay, a similar picture emerges. In Gil (2001), a contrastive analysis of Riau Indonesian with Siak Malay, a traditional dialect also spoken in Riau province, reveals a tendency

---

27. In Papuan Malay, as in many other eastern Indonesian koinés, pronouns occur in two variants "long" (disyllabic) and "short" (monosyllabic). As far as I have been able to ascertain, there are no syntactic or semantic differences in the distribution of the two variants; in particular, both long and short pronouns occur in all four of the constructions in Table 4. However, only the short pronouns may undergo the process of cliticisation described here.

28. In particular, when occurring in front of words denoting activities, such pronouns may sometimes occur in conjunction with a coreferential word or phrase, in what Haspelmath (2013) refers to as a conominal construction. In such cases, the construction in question takes on a resemblance to a subject-verb agreement construction of the kind that violates McWhorter's Creole Prototype, specifically, item (2-i) above stipulating "little or no inflectional affixation", though in the cases at hand this is clearly a recent and in fact ongoing innovation.

for Siak Malay to be just a shade more complex than Riau Indonesian, albeit still extremely simple overall (and indeed, simpler than McWhorter's stock example of a creole language, Saramaccan). However, only one of the differences observed pertains to word structure, and even it is of a quantitative and rather impressionistic nature: the use of *(me-)(N-)* would appear to be somewhat more frequent in Siak Malay than in Riau Indonesian – though both varieties allow for both options, with *(me-)(N-)* prefixation occurring as an alternative to zero-marked forms. Indeed, in general, word structure in Riau Indonesian and Siak Malay is remarkably similar, exhibiting a near-identical degree of isolating structure.

Moreover, just as Riau Indonesian is not exceptional among Malay/Indonesian koinés, Siak Malay is typical in its isolating profile of traditional varieties of Malay, as may readily be observed in descriptions of dialects such as Jambi Malay (Yanti 2010), Ulu Muar Malay (Hendon 1966), Brunei Malay (Clynes 2001) and others, and in corpora of dialects such as Langkat Malay, Bangka Malay, Pontianak Malay and Balai Berkuak Malay available in Gil et al. (2015). To be sure, there are differences of detail, and the available material on these dialects does not always provide adequate explicit argumentation in support for assignments of wordhood. Nevertheless, the overall picture that emerges is quite clear: there is no systematic difference in word complexity between the Malay/Indonesian koinés and the traditional Malay dialects.

For example, the agent- and patient-oriented generalised voice markers *(me-)(N-)* and *di-* are present in Jambi Malay, Ulu Muar Malay and Balai Berkuak Malay but absent in Patani Malay. End-point-oriented generalised voice markers are also present in some dialects, such as Jambi Malay, Pontianak Malay and Balai-Berkuak Malay; indeed, in some dialects there are two different markers with different functions, such as *-kə* and *-i* in Palembang Malay, *-kan* and *-i* in Ulu Muar Malay and Brunei Malay – in the latter, the two forms may sometimes cooccur in the same word. On the other hand, there are no such markers in Kedah Malay and Patani Malay. Similarly, the suffix *-an* is present productively in some dialects, such as Ulu Muar Malay and Brunei Malay, but not in others, such as Jambi Malay and Balai Berkuak Malay – though for this suffix in particular it is often difficult to establish an objective threshold for what counts as productive. Thus, examples such as these suggest that for the most part, traditional Malay varieties share the isolating grammatical profile of their koiné counterparts.

Looming large, however, within the Malay/Indonesian linguistic landscape is an outlier language variety with a somewhat different grammatical profile: Standard Malay/Indonesian. The most cursory inspection of a written text in Standard Malay/ Indonesian will show that words are typically long; much longer than in colloquial varieties, traditional or koiné. While some of these words are monomorphemic, often borrowed from Indic languages, Arabic, Dutch and English, many others are

polymorphemic. Lauder (2008: 124) reports that "Indonesian word structure consists of a large number of derivational morphemes, at least 380 by one unpublished corpus study"; however, this figure, based on the author's own work in computational linguistics, includes a large number of forms of an esoteric and marginal nature, typically occurring in texts of a rather rarefied style. Often, conscious language engineering tries to make Standard Malay/Indonesian bear a greater resemblance to its proponents' notion of what a respectable national language should look like, typically modelled after English and other languages with lots of morphology. For example, in Malaysia, Malay subtitles for many English-language soap operas typically render every occurrence of the English plural marker *-s* with a reduplicated form in Malay, even though such usage runs counter not only to the colloquial varieties of Malay but also to other, more widespread versions of the standard language. Still, whatever the style, the complexity of word structure in Standard Malay/Indonesian remains consistently greater than in most or all of its colloquial varieties.

For example, with a few lexical exceptions, words denoting activities never occur in bare form, but only with a generalised voice marker such as the agent-oriented *(me-)(N-)* or the patient-oriented *di-*. Moreover, the use of these markers is grammatically constrained in a way that it is not in many or most colloquial varieties of Malay/Indonesian – see Section 5.3 below for more discussion. Two end-point markers, *-kan* and *-i*, are present, as in a few colloquial varieties, but not most others, which have either one such marker or none. Some of the numerous other affixes occurring in Standard Malay/Indonesian but not in most colloquial varieties include *-wan*, *-wati*, *-anda*, *pra-*, *tuna-*, and *antar-* (Sneddon 1996: 44–53).

Given the above, although Dryer's (2005) *World Atlas of Language Structures* map of morphological structures, mentioned in the introduction to this volume, characterises 47% of the Austronesian languages in his sample as having "little or no affixation"; Standard Indonesian is assigned to the remaining 53% of the Austronesian languages, those with significant amounts of morphology. Nevertheless, from a world-wide perspective, it is important not to overstate the degree of complexity of word structure in Standard Malay/Indonesian. In Dryer's world-wide sample, a full 86% of the world's languages fail to have "little or no affixation", and among these, Standard Malay/Indonesian is undoubtedly among the languages with the least amount of morphological complexity – nowhere near the ballpark of, say, McWhorter's (2001a and elsewhere) favourite morphologically complex language Tsez. Even in comparison with English, also in Dryer's 86%, and a language not renowned for its complex word structure, the simplicity of Standard Malay/Indonesian is striking: among the many complex features of English morphology completely lacking in Standard Malay/Indonesian are suppletion, ablaut, tense-aspect marking, agreement marking, and essentially any kind of inflectional morphology. Indeed, with regard to the latter, Standard Malay/Indonesian would

seem to be in full compliance with McWhorter's Creole Prototype property (2/ii) "little or no inflectional affixation".

In summary, then, Standard Malay/Indonesian emerges as somewhat more morphologically complex than other colloquial varieties of the language, albeit not dramatically so. Accordingly, adding the standard language to the rather mixed bag of non-koiné varieties of Malay/Indonesian contributes to the characterisation, as per (4b), of the Malay/Indonesian koinés as being at most only moderately less complex than other Malay/Indonesian varieties.

4.3 Malay/Indonesian and other Malayic varieties

Expanding the circle one step further to include other Malayic varieties, the overall picture still does not change much, with widespread isolating structure punctuated sporadically by occasional innovations introducing a modest degree of morphological complexification. This is not really all that surprising, given that the Malay/Indonesian language does not constitute a coherent genealogical unit definable in terms of shared innovations: whether or not a Malayic variety is considered to be a variety of Malay or Indonesian is dependent on a complex interplay of extralinguistic factors, as a result of which the boundary between Malay/Indonesian and other Malayic varieties is often fuzzy and ill-defined.[29] Still, many Malayic varieties are significantly different from most varieties of Malay/Indonesian, reflecting the substantially greater time depth of their diversification, and thereby warranting their separate consideration here.

A substantial number of Malayic varieties exhibit an isolating profile that is generally similar to that of Malay/Indonesian. This can be readily observed in the grammatical descriptions of varieties such as Minangkabau (Moussay 1981, see also Crouch this volume on colloquial Padang Minangkabau), Urak Lawoi' (Hogan 1988), Sakai (Kalipke & Kalipke 2001), Besemah (McDonnell 2016), Mualang (Tjia 2007), Iban (Asmah 1981) and others. Further support for this conclusion may be

---

29. For example, Minangkabau, once considered by some scholars to be a dialect of Malay, is now commonly viewed as a separate language, given its strong cultural identity associated with properties such as matrilineal inheritance and a famous cuisine. In southern Sumatra, in many Malayic-speaking areas, there is no coherent notion of language as a shared collection of speech patterns across a substantial area: when asked what they speak, people simply respond with an expression of the form *bahasa T*, 'the language of T', where T is a toponym usually referring to the village or town in which they live. For the most part, there is little or no identification with Malay ethnicity and culture of the kind found further north, in Riau, which would justify referring to such varieties as dialects of Malay. Finally, in both Sumatra and Borneo, the identification of a Malayic variety as a dialect of Malay is often dependent on religion, and the widespread assumption that in order to be Malay and speak Malay, one must believe in Islam.

obtained by perusal of the naturalistic corpora in Gil et al. (2015), for varieties such as Minangkabau, Besemah, Rantau Panjang and Sungai Jernih Sarolangun (the latter two spoken in western Jambi province).

Once again, this conclusion can be further supported through the examination of selected affixes. Crouch (this volume) provides a detailed description of the optionality of the agent-oriented generalised voice affix in colloquial Minangkabau and the corresponding prevalence of bare unmarked stems; other varieties in which a stem marked with *(me-)(N-)* may alternate with a zero-marked form include Urak Lawoi', Tapan, Besemah, Mualang and Iban – though the conditions governing such alternations may differ from one variety to another. Cognates of the patient-oriented generalised voice affix are also present in most Malayic languages, though once again, their functions may differ somewhat from one language to another, as evident, for example, from Adelaar's (2005) discussion of *di-* in Salako; however, in other varieties, such as Urak Lawoi', they are absent. End-point-oriented generalised voice markers vary somewhat more in both form and function; some languages have two, for example Besemah *-ka* and *-i*, Sakai *-kat* and *-ng,* others just one, such as Padang Minangkabau *-an*, Tapus Minangkabau *-ge*, while yet others, including Urak Lawoi', Tapan and Mualang, do not have any such markers. Variation is also present in associative markers: some varieties have two, for example Padang Minangkabau *-nyo* and *-e*, others just one, as in Besemah *-nye* (with morphophonemic variation), while yet others have none, as is the case for Mualang. Examples such as these and several others attest to a largely similar word structure shared by Malay/Indonesian varieties and several other Malayic languages.

Against this shared background, however, several cases stand out of morphological complexification in Malayic languages that have no obvious counterparts within the more limited domain of Malay/Indonesian. Some of these are relatively minor and localised. Thus, in Bangkinang, spoken in the eponymous city and surrounding regions in western Riau province, the end-point-oriented generalised voice marker exhibits morphophonemic adjustments suggesting that it is more closely bound to its host than in other varieties; for example, adding it to *mati* 'die' yields the form *motiin*, triggering change of the first vowel from *a* to *o* – under one plausible analysis, the tight bond between the stem and the suffix results in a trisyllabic word in which the antepenultimate syllable is reduced to the neutral vowel *o*. In Tapan, the associative marker *-ă* exhibits increased phonological bonding of a rather exceptional nature. As argued in McKinnon & Gil (2014), when *-ă* is added to typical disyllabic words such as *asap* 'smoke', the resulting form, *asapă*, is actually disyllabic, with *-apă* constituting a single syllable in which the short *ă* vowel effects a secondary sonority peak within the coda. Indeed, when the final coda is a weak consonant, the short vowel spreads leftwards across the consonant, as in *pilih ~ piliăhă* 'choose'. In terms of a structure of the kind proposed in Gil (this

volume, Chapter 1), the associative marker -*ă* would thus seem to be contained within the core foot, thereby contributing, albeit rather modestly, to an increase in word structure complexity in Tapan.

A productive source of incipient innovative morphological structure in Malayic languages is provided by a variety of rules of phrase-final phonology, whose effect is to mark the final syllable of a phonological phrase as different from all of the preceding ones (Gil & McKinnon 2015). The most common example of this involves stress, which, in most Malayic languages of Western Nusantara falls on the final syllable of the phonological phrase (see Gil 2006 for Riau Indonesian); however, in addition to stress, individual dialects and languages make use of a wide array of phonological features in order to effect a similar pattern. Amongst the Malay/Indonesian koinés, Jakarta Indonesian exhibits an phrasally-conditioned alternation between zero and glottal stop in final coda position, for example phrase-medial [mata] ~ phrase-final [mataʔ]. In Kupang Malay, words of the form *CaCV[high]* undergo reduction in phrase-medial position, with the final high vowel undergoing deletion or metathesis, for example phrase-medial [kas] / [kays] ~ phrase-final [kasi] 'give'. In these and other similar cases, the alternations are not categorical but rather tendencies, albeit significant ones, and they do not bear any grammatical consequences. Similar alternations, occur in other Malayic languages. Thus in Minangkabau, in certain phonological environments an epenthetic vowel is inserted before a word-final consonant in phrase-final position, for example phrase-medial [aɲjiŋ] ~ phrase-final [aɲjiăŋ] 'dog'. An alternation common to many varieties in Sumatra and Borneo involves the pre-oralisation of final nasal stops, for example, in the Orang Asli dialect of Padang and Bengkalis islands in Riau province, phrase-medial [makan] ~ phrase-final [makatn] 'eat'. Yet another common alternation in many varieties of Sumatra is that which introduces an excrescent nasal following a final high vowel, for example, in the Pasia Sabalah subdialect of Padang Minangkabau (briefly discussed in Gil 2016: 444–445), phrase-medial [jawi] ~ phrase-final [jawiŋ] 'cow'. The preceding alternations are also, for the most part, phonological and of a quantitative nature. However, in a contrastive analysis of three Malayic varieties of West Sumatra, Gil and McKinnon (2014) present evidence for a path of grammaticalisation of the excrescent nasal alternation: whereas in the Pasia Sabalah subdialect of Padang Minangkabau it is purely phonological, in the city dialect of Tapan it is in the process of becoming systematised, and in the Binjai village dialect of Tapan it is largely grammaticalised.

Perhaps the most remarkable case of grammaticalisation of phrase-final phonology is that provided by a cluster of Malayic varieties in western Jambi province centering on the mountain valley of Kerinci. In Kerinci, almost every word in the language may occur in either of two forms, commonly referred to as absolute and oblique – see Prentice and Usman (1978), Steinhauer and Usman (1978),

McKinnon (2011) and Ernanda (2017). Phonologically, the alternations between absolute and oblique forms are of considerable complexity, focusing on the final rhyme of the word, and typically involving complex patterns of vowel gradation. For each rhyme, the absolute-oblique alternation needs to be specified individually; there are no obvious phonological generalisations governing the forms of the alternations. Thus, learning the language requires mastery of a page-sized table specifying, for each rhyme, the phonological forms of the corresponding absolute and oblique forms. The functions of the alternation are also variegated. The absolute form is the one commonly used in citation, but it also occurs in other syntactic environments. The oblique form occurs in a mixed bag of syntactic constructions, typically but not always involving a tighter syntactic nexus to a following word or phrase. Further complicating the picture is the existence of a truly mind-boggling amount of dialectal variation, whereby in each village, occasionally even in different parts of the same village, the alternation assumes different phonological forms, and sometimes also different functions.

A diachronic account for the absolute-oblique alternation is provided in a series of articles by McKinnon, Cole and Hermon (2011), McKinnon et al. (2015), and McKinnon et al. (2018). The leading idea behind their account is that the alternation derives from the convergence and coalescence of two different paths of grammaticalisation and lexicalisation starting from two distinct sources. The first source is the associative marker, which gets phonologically "sucked in" to its host word, as was illustrated above for the Tapan forms *asapă* and *piliăhă*. And the second source involves phrase-final phonological alternations of the kind considered above. Historically, then, the oblique form is argued to originate in some combination of the associative marker and the phrase-medial form of the word in question.

An illustration of how these two sources contribute to the grammar of the absolute-oblique alternation is given in the contrast between Riau Indonesian (5) and the two corresponding constructions in the Tanjung Pauh Mudik dialect of Kerinci (from Gil & McKinnon 2014) in (6):

(5) Riau Indonesian
    Pintu hitam
    door   black
    i.   'The door is black'   predicative
    ii. 'black door'   attributive

(6) Tanjung Pauh Mudik Kerinci
    a. *Pint$^y\widehat{ao}$ it^{y}a*
        door:ABS black:ABS
        'Doors are black'   generic, predicative

b. *Pintʸɨw   itʸa*
   door:OBL black:ABS
   i. 'black door'   attributive
   ii. 'The door is black'   specific, predicative

In Riau Indonesian, with no such alternations, the construction in (5) may be understood either attributively or predicatively; as argued in Gil (2005, 2017a), this is an instance of vagueness, rather than ambiguity, consistent with the language's IMA grammatical profile. In contrast, Tanjung Pauh Mudik Kerinci provides the option of marking the first word as either absolute as in (6a) or oblique as in (6b), a choice that affects the available interpretations of the expression. Diachronically, these interpretations provide a clear reflection of the two historical sources of the absolute-oblique alternation. The contrast between the predicative interpretation of (6a) and the attributive interpretation of (6b-i) mirrors the phrasal-phonological source of the alternation, with the erstwhile phrase-final absolute form *pintʸa͡o* occurring before the major syntactic break between subject and predicate, while the originally phrase-medial oblique form *pintʸɨw* occurs within the tighter syntactic environment of a head-plus-attribute construction. As for the interpretation in (6b-ii), this is a particular case of the generalisation to the effect that words in Kerinci may occur in the oblique form even in positions typically associated with absolute forms, in which case the oblique form results in a definite interpretation; this generalisation in turn reflects the origin of the alternation in the associative marker, one of whose functions, across the Malayic family, is to mark definiteness.

Example (6) gives just a little taste of the complexity of the grammatical functions associated with the absolute-oblique alternation in Kerinci. McWhorter (2007: 213,233) is right in commenting that there is nothing comparable to this level of complexity within Malay/Indonesian. Indeed, the Kerinci absolute-oblique alternation, both phonologically and grammatically, is up there amongst the most baroque linguistic patterns observable anywhere. Nevertheless, it is all too easy to be blinded by this exuberance and lose sight of the broader grammatical profile of the language. The absolute-oblique alternation aside, a big aside but still, the remainder of the language is actually very Malayic-like in its overall ground plans, including even a word structure that is otherwise on a par with other Malayic languages in its largely isolating character. Thus, as summarised in McKinnon (2011: 77), Tanjung Pauh Mudik Kerinci has agent- and patient- but no end-point-oriented generalised voice affixes; and it has just a small handful of other affixes, while lacking counterparts to various others, such as the suffix *-an* and the causative prefix *per-*, which are both absent.

In summary, then, while many Malayic languages are every bit as isolating as their Malay/Indonesian counterparts, some exhibit a certain amount of secondary complexification in their word structure. Thus, in accordance with (4c), Malay/Indonesian varieties are at most only moderately less complex than other Malayic varieties.

4.4   Malayic varieties and other languages of Western Nusantara

Moving beyond Malayic, the other languages of Western Nusantara, although all Austronesian, exhibit, as one might expect, an increasing degree of grammatical diversity. In terms of their morphological profiles, these languages can be divided into three broad groups: those with isolating structure comparable to that of most Malayic varieties, those with a significant amount of innovated morphological complexity, and those with a significant amount of inherited morphological complexity shared with Austronesian languages outside of the region.

It is the latter two groups of languages, presumably, that lead McWhorter to the claim that Malayic languages are systematically simpler than their non-Malayic counterparts to the region. This claim forms the basis for his characterisation of Malay as an NCSL discussed in Section 2.3 above, see especially Table 2. The exceptionality of Malayic and Malay amongst the languages of the region is a recurrent refrain in his writings – see, for example, McWhorter (2001c: 407, 2007: 207, 2008: 171–2, 2019). Why aren't there any other similarly reduced languages in the region outside of Malayic?, he asks, again and again, Where is the "Riau Javanese"? To which my response is: But there are such languages, lots of them, and they're all over the place, hiding in plain sight – languages with isolating structure on a par with that typical of Malay/Indonesian and other Malayic varieties.

Consider, first, that other "elephant in the Austronesian room", namely Javanese. As argued in detail in Conners (this volume), linguists' impression of what Javanese is like is unduly influenced by the prominence of the Central Javanese dialect which forms the basis for what is considered to be Standard Javanese. However, if, following Conners' suggestion, we shift our attention to the other, more geographically peripheral dialects of Javanese, then suddenly a much more isolating structure is in evidence – if not every bit as isolating as Riau Indonesian, then pretty close. The strongly isolating nature of peripheral Javanese dialects is evident in the extensive corpora compiled by Conners, accessible in Gil et al. (2015) – see Example (17) below.

And it's not just Javanese. Varying degrees of isolating structure are evident, *mutatis mutandis*, in many other non-Malayic languages of the region, including Acehnese (Durie 1985), Rejang (McGinn 1982), Nasal (Anderbeck and Aprilani

2013), Madurese (Davies 2010) and Maanyan (Gudai 1985). A similar isolating pro-
file is also evident in the Gil et al. (2015) corpora for various dialects of Onya Darat
and Kenyah. Following are some typical naturalistic utterances in the Samandang
dialect of Onya Darat (compiled by Uri Tadmor) and the Óma Lóngh dialect of
Kenyah (compiled by Antonia Soriente):[30]

(7) Samandang Onya Darat   [150826060908385293720941]
     eʔ, naʔ biis te,   naʔ mo   jane   te,   dah
     yes want sleep later want 2SG.SAME/YNG admonition later after
     mo   mondiʔ te.
     2SG.SAME/YNG come   later
     'Yes, use it later when you want to go to sleep'

(8) Óma Lóngh Kenyah   [134759162344160806]
     tè te có alem la'a ateq ii lèny deq òèj   nè   ngkiny
     go PRT one night again very who really REL NEG.IMP come carry
     ataq di   sèket alem lèny dae Bòngèny té
     fish DEM.PROX every night really voice Bòngèny there
     'The next night, who really wants to bring fish every night, said Bòngèny'

As evident from the interlinear glosses in (7) and (8) above, there is a one-to-one
correspondence between morphemes and orthographic words, which, in the ab-
sence of evidence to the contrary, can be taken as indicative of a strongly isolating
structure. While some other utterances in the Onya Darat and Kenyah corpora may
contain a small number of bound morphemes, utterances such as the above are not
exceptional in any way, but, rather, representative of the corpora as a whole, and
the isolating structure of the two languages.

Thus, the prevalence of isolating structure across Sumatra, Java and Borneo
shows that, in this respect, Malay/Indonesian and Malayic are typical languages of
their region – anything but exceptional. Some further discussion of the grammat-
ical simplicity of Western Nusantara languages in the domain of clause structure
is provided in Section 5.3 below. To be sure, there are also many languages in the
region with significantly more complex word structure, in some cases innovative,
in other cases apparent retentions from earlier Malayo-Polynesian morphology.
Broadly speaking, such languages appear to be concentrated in peripheral regions
of Western Nusantara, such as northern Sumatra and the Barrier islands of the
west coast of Sumatra, plus also in northern Borneo. But the general overall picture

---

30. In the naturalistic data from the Gil et al. (2015) corpora cited in this chapter, the utterance
ID number is indicated in square brackets. The interlinear glosses follow the Leipzig Glossing
Rules, with occasional additions. In Example (7) "SAME/YNG" means 'same or younger generation'.

remains one of widespread isolating structure, not significantly different from that of most Malayic languages.

In summary, then, taking all of the languages of the region into consideration, Malayic varieties emerge as at most only moderately less complex than other languages of Western Nusantara, in accordance with (4d).

4.5   Languages of Western Nusantara and other Mekong-Mamberamo languages

Of all the language networks referred to in Figure 1 and subsequently in (4), the languages of Western Nusantara probably constitute the least well motivated group. As suggested in Figure 1, they form part of two larger partly overlapping networks, of which one is areally defined, namely Mekong-Mamberamo.

Named after the two major rivers at its western and eastern extremities, the Mekong-Mamberamo linguistic area, proposed in Gil (2015), comprises the entirety of mainland Southeast Asia, most of Nusantara, albeit shading off somewhat in northern parts of Borneo and Sulawesi, plus parts of western New Guinea, including the Bird's Head and, to a lesser extent, some adjacent regions. The Mekong-Mamberamo linguistic area contains languages belonging to some twenty or more different linguistic families: in the west are Sino-Tibetan, Austroasiatic, Tai-Kadai and Hmong-Mien, in a large central swathe Austronesian, and to the east a larger number of smaller language families and isolates such as Timor-Alor-Pantar, North Halmahera, West Bird's Head, Maybrat, Abun, Mpur, Hatam, East Bird's Head, Yawa and others. The Mekong-Mamberamo linguistic area is characterised by the following 17 properties:

(9)   Mekong-Mamberamo Properties
1. passing gesture
2. repeated dental clicks expressing amazement
3. conventionalised greeting with 'where'
4. 'eye day' > 'sun' lexicalisation
5. d/t place-of-articulation asymmetry
6. numeral classifiers
7. verby adjectives
8. basic SVO word order
9. iamitive perfects
10. 'give' causatives
11. low differentiation of adnominal attributive constructions
12. weakly developed grammatical voice
13. isolating word structure

14. short words
15. low grammatical-morpheme density
16. optional thematic-role flagging
17. optional TAM marking

In general, the above 17 properties occur significantly more frequently within the Mekong-Mamberamo area than outside of it, thereby demarcating it and setting it apart from neighbouring linguistic areas such as South Asia, Northeast Asia, Taiwan and the Philippines, the remainder of New Guinea, and Australia. In particular, Riau Indonesian and many of its relatives exhibit almost all of the 17 Mekong-Mamberamo properties.[31]

Of the above 17 properties, the last 7, that is to say numbers 11–17, characterise Mekong-Mamberamo languages as being of simpler grammatical structure than their non-Mekong-Mamberamo counterparts. The remainder are of equal complexity to their non-Mekong-Mamberamo counterparts, except for number 6, numeral classifiers, that is actually more complex. Among the last 7 properties, at number 13, is isolating word structure. What this says, then, is that from an areal point of view, the isolating grammatical profile of many of the languages of Western Nusantara is in fact characteristic of a much broader area containing both Austronesian and non-Austronesian languages.

In order to compare languages on such a large scale, it is not practical to go into the same level of detailed analysis adopted in the preceding sections; depth must be sacrificed for breadth, as instead we avail ourselves of various cross-linguistic studies at a world-wide level. In Dryer's (2005) map, mentioned in Section 4.2 above, 37 out of 62 Mekong-Mamberamo languages are characterised as having "little or no affixation", a rate of 60%, in contrast to the 14% rate world-wide.[32] Similarly, in Bickel and Nichols' (2005) map of morphological type as manifest in the expression

---

31. For example, Riau Indonesian exhibits 16 of the properties, lacking only number 10, 'give' causatives; similar facts hold also for Jakarta Indonesian, Siak Malay, Padang Minangkabau and many other Malayic varieties. A somewhat different pattern is evident in Papuan Malay, which has 'give' causatives, but lacks property number 11, low differentiation of adnominal attributive constructions: genitive modifiers are distinguished from adjectival and relative-clause modifiers in that they occur before rather than after their heads – it thus also totals 16 properties. An example of a Malayic variety satisfying all 17 properties is Sabah Malay. Another possible example might be Tapus Minangkabau, contingent on the possible analysis of the causative suffix *-ge* as being a reduced form of *agiah* 'give'.

32. For the purposes of this calculation, the Mekong-Mamberamo area was taken to consist of China south of 30N, all the countries of mainland Southeast Asia, plus Nusantara west of 135E. It should be noted that Dryer's map significantly under-samples the languages of Western Nusantara, perhaps because he considers them less interesting; a fairer sample of such languages would probably increase the isolating quotient of the Mekong-Mamberamo area.

of case and tense-aspect-mood categories, 5 out of 11 Mekong-Mamberamo languages are characterised as "exclusively isolating", a rate of 45%, as opposed to a 10% rate world-wide.[33]

Using similar criteria but a much larger sample of some 1500 languages world-wide, Donohue and Denham (this volume) present a map showing the largest concentration of isolating languages in Mainland Southeast Asia, a secondary cluster in eastern Nusantara and western New Guinea, and sporadic cases elsewhere in the Mekong-Mamberamo area. Donohue and Denham's map thus clearly displays a Mekong-Mamberamo distribution for isolating languages. In subsequent maps, Donohue and Denham decompose the holistic notion of isolating language into more specific instances of isolating structure pertaining to particular morphosyntactic features, providing maps showing the absence of verbal agreement, tense marking, bound causatives, core case marking, and subordinating morphology. In their Map 8 they then take the intersection of these maps, and the resulting languages, those in which all of the above are absent, are almost exclusively located in the Mekong-Mamberamo area.[34]

Two cases of strongly isolating Austronesian languages outside the Western Nusantara region but within the Mekong-Mamberamo linguistic area are Cham, spoken in Vietnam, discussed by Brunelle (this volume), and Lio, spoken on the island of Flores in eastern Indonesia, discussed by Elias (this volume); both chapters attest to the pervasiveness of the isolating linguistic profile in Austronesian languages across the Mekong-Mamberamo area. Further to the east, Gasser, Arnold and Kamholz (to appear) comment on the presence of varying degrees of isolating structure in the languages of the South Halmahera West New Guinea subgroup, such as, for example, Ambel, as described by Arnold (2017). In his paper titled "Why Does a Language Undress: Strange Cases in Indonesia", dealing with Riau Indonesian, central Flores, and Timor, McWhorter (2008: 170) expresses his puzzlement that "for reasons unknown to me at present" (as he puts it), the three most serious challenges to his hypothesis regarding the relationship between language

---

33. The slightly different figure of six isolating Mekong-Mamberamo languages in Bickel and Nichols (2005) cited in Gil (2015: 311) is due to the marginally different boundaries adopted for the Mekong-Mamberamo area; whereas the earlier calculation included some languages of northeast India, here they were excluded.

34. Closer inspection of the database on which Donohue's map is based suggests the presence, within the Mekong-Mamberamo area, of a west-to-east cline, whereby isolating structure is most pronounced in Mainland Southeast Asia, gradually decreasing through western Nusantara, eastern Nusantara and into western New Guinea, where it is at its relative weakest. Significantly, though, even in western New Guinea, isolating structure is more prevalent than it is worldwide, and, in particular, more prevalent than in neighbouring non-Mekong-Mamberamo regions such as Taiwan and the Philippines to the north, and the remainder of New Guinea to the east.

contact and simplification all come from Indonesia. But the reason for this is obvious: they are all part of a single Mekong-Mamberamo linguistic area, which, like all linguistic areas, is the currently-observable product of multiple past events of language contact.

Outside of the Austronesian family, isolating languages are equally plentiful at both ends of the Mekong-Mamberamo linguistic area. To the west, the isolating profile of Mainland Southeast Asia is well known; see for example Bybee (1997) and Enfield (2005, 2011). Less familiar, perhaps, is the widespread occurrence of isolating structure among the non-Austronesian languages of eastern Indonesia and western New Guinea. Abun, an isolate spoken on the north coast of the New Guinea Bird's Head, is described by Berry and Berry (1999) as highly isolating. Many other Bird's Head languages have a set of person index prefixes, but relatively little morphology otherwise; for example Maybrat, Mpur and Hatam. Elsewhere in eastern Indonesia, a moderately isolating profile is characteristic also of North Halmahera languages such as Tidore and Tobelo, as well as some of the Timor-Alor-Pantar languages, in particular those of Timor – see discussion in Schapper (this volume).[35]

Thus, isolating structure is a characteristic feature of the entire Mekong-Mamberamo area, from Mainland Southeast Asia though Nusantara and into western New Guinea. Of course, in a linguistic area of this size, there are bound to be exceptions, pockets or regions of greater morphological complexity. As foreshadowed in Section 4.4 above in the context of Western Nusantara, such exceptions may fall into two main types: innovations and intrusions. Innovations leading to greater morphological complexity are far too many to list here; just a single example will be mentioned. Recall the grammaticalisation of phrase-final phonology into a morphosyntactic absolute-oblique alternation in Kerinci discussed in Section 4.3. An analogous development can be observed in several languages of Timor, in which a process of metathesis, originally marking the end of a phonological phrase, develops into an inflectional paradigm, which, as in Kerinci, is of substantial phonological and morphosyntactic complexity, and also exhibiting a substantial amount of variation between dialects and closely related languages – see Steinhauer (1996), Engelenhoven (2004), Edwards (2016), and Schapper (this volume, Section 4.4) for details. A second source of morphological complexity involves the intrusion, into the Mekong-Mamberamo area, of languages and language families from outside

---

35. Commenting on the apparent simplicity of the languages of Bird's Head and Halmahera, Reesink (2002: 26) mentions

predominant CV(C) syllable structure, a five-vowel system, a dearth of morphological complexity, except for subject and possessor prefixation of verbs and inalienable nouns, lack of tense-mood-aspect making, SVO word order and asyndetic conjunctions, suspiciously similar to what are known as serial verb constructions, and so on.

the area, bringing their more complex word structures with them. One example is the spread of Tibeto-Burman languages from the north and west into Mainland Southeast Asia, while another is the spread of Trans-New-Guinea and possibly other language families of New Guinea westwards into the southern Bird's Head, the Bomberai peninsula, and perhaps also the islands of Timor, Alor and Pantar.[36] But of course, the most striking example is that which is of central concern to this chapter, namely the spread of Austronesian languages into Nusantara.

In summary, then, in accordance with (4e), the languages of Western Nusantara are at most only moderately less complex than other languages of the Mekong-Mamberamo region with respect to word structure. Indeed, notwithstanding the inevitable bumps and wiggles, the Mekong-Mamberamo area, as argued in Gil (2015), constitutes the largest concentration in the world of isolating structure. Thus, as suggested in Section 4 up to this point, Riau Indonesian is isolating because Malay/Indonesian koinés are, Malay/Indonesian koinés are isolating because Malay/Indonesian is, Malay/Indonesian is isolating because Malayic is, Malayic is isolating because Western Nusantara is, and Western Nusantara is isolating because the Mekong-Mamberamo area is. This is the path represented in Figure 1, back at the beginning of this chapter, with a double line. However, to complete the picture, we must now turn to consider an network of languages that does not share the isolating profile of the Mekong-Mamberamo area.

4.6   Languages of Western Nusantara and other Austronesian languages

It is generally accepted that the Austronesian languages originated in Taiwan, spread south into the Philippines, and from there further south into Nusantara. However, as is patently obvious to even the most casual observer, the Austronesian languages of Taiwan and the Philippines are anything but isolating: any reference grammar of such a language will contain lengthy chapters describing a variety of complex morphological patterns. The resulting boundary between isolating languages in Indonesia and non-isolating ones in the Philippines is a crucial component of the demarcation of the Mekong-Mamberamo linguistic area proposed in Gil (2015).

In recognition of the challenges associated with the determination of word-boundaries discussed at length in Gil (this volume, Chapter 1), four surrogate properties are proposed in Gil (2015) – linguistic features that, while logically

---

36. The intrusive nature of the latter case would of course be contingent on the establishment of a genealogical connection between the Trans-New-Guinea and Timor-Alor-Pantar language families, as first proposed by Wurm, Voorhoeve and McElhanon (1975) and Stokhof (1975), which in turn would entail that the latter family represents the product of a migration westwards from New Guinea.

independent of isolating word-structure, tend to correlate with it empirically. Items 14–17 in (9) above, these four properties are short words, low grammatical-morpheme density, optional thematic-role flagging, and optional TAM marking. Since the values of these features are more easily ascertainable without recourse to the kind of in-depth analysis provided for Riau Indonesian in Gil (this volume, Chapter 1), they lend themselves more readily to broad, large-scale cross-linguistic surveys of the kind presented in Gil (2015). These four properties are illustrated below in a contrastive analysis of Riau Indonesian and colloquial Tagalog, showing the simplest and most natural way of saying 'The chicken is eating':[37]

(10) Riau Indonesian
Ayam makan
eat   chicken
'The chicken is eating'

(11) colloquial Tagalog
Kumakain   *yung manok*
<AT.REAL>IPFV~eat TOP   chicken
'The chicken is eating'

The property of word-length as a distinguishing feature of Mekong-Mamberamo as opposed to other Austronesian languages is evidenced in the presence of the four-syllable word *kumakain* in (11); whereas Malay/Indonesian and many other Mekong-Mamberamo languages have a maximal-word constraint strongly dispreferring words of more than two syllables, such a constraint is absent from the Austronesian languages of Taiwan and the Philippines, where longer words abound. Perhaps the most salient and far-reaching typological property of Mekong-Mamberamo languages, though, is low grammatical-morpheme density. In Mekong-Mamberamo languages, grammatical morphemes are mostly optional and relatively infrequent, resulting in long stretches of text consisting entirely of content words. In contrast, in the Austronesian languages of Taiwan and the Philippines, grammatical morphemes are often obligatory, occurring relatively frequently in texts. This difference is clearly visible in the comparison of (10), with no grammatical morphemes, and (11), with three: partial reduplication of *kain* > *kakain* expressing imperfectivity, infixation of

---

37. The choice of colloquial rather than standard Tagalog as the comparandum for Riau Indonesian is intended to facilitate a sociolinguistically like-for-like comparison involving two colloquial varieties of a large national language. It should be noted that the conventional orthography for Tagalog adopted here somewhat downplays the complexity of word structure of Tagalog. As argued in Gil (1996), markers such as *yung* are not separate words but rather clitics that attach to their following hosts; if this had been represented as such in (11) above, the contrast between (11) and its Riau Indonesian counterpart in (10) would have been even more striking.

-*um*- combining realis aspect/mood and assignment of the actor thematic role to the clause's topic, and *yung* flagging its host *manok* as topic and further marking it as definite. The property of low grammatical-morpheme density underlies two more specific properties associated with Mekong-Mamberamo languages. The first is optional thematic-role flagging. In (10), *ayam* is not flagged, and indeed, it can bear any thematic role whatsoever: actor (as in the given translation) but also patient, or any other role that might make sense in the context of the utterance. In contrast, in the Austronesian languages of Taiwan and the Philippines, thematic-role flagging is often obligatory; thus in (11), *manok* is flagged with *yung*, which, in conjunction with the actor-topic infix -*um*- on *kain*, marks *manok* as the actor, while ruling out alternative assignments of thematic roles. The second more specific property is optional TAM marking. In (10) it is absent, and the sentence can accordingly be interpreted not only as present progressive (as per the translation) but with any other possible conceivable combination of tense, aspect and mood. In contrast, in many Austronesian languages of Taiwan and the Philippines, including Tagalog, TAM marking is obligatory; thus in (10), *kain* bears two such markers, initial CV-reduplication and the infix -*um*-.

The contrast between Riau Indonesian (10) and colloquial Tagalog (11) is not fortuitous. Rather, it is emblematic of a systematic difference in grammatical profile between Mekong-Mamberamo languages and those spoken further north, in Taiwan and the Philippines, as argued for in detail in Gil (2015). In particular, whereas Mekong-Mamberamo languages typically exhibit an isolating grammatical profile, the Austronesian languages of Taiwan and the Philippines are generally characterised by richer and more complex morphological structures, in accordance with (4f) above. In terms of their grammatical structure and isolating profile, then, Riau Indonesian and its relatives bear a closer resemblance to Mekong-Mamberamo languages such as Vietnamese and Abun than they do to Austronesian languages such as Tagalog. It is this observation that lies at the heart of the story of Riau Indonesian and its relatives, and for which, in the next and final section, an account is proposed.

5.   The story of Riau Indonesian and its relatives

Sections 3 and 4 above surveyed the relevant sociohistorical and linguistic landscapes of Riau Indonesian and its relatives, close and distant. We now bring these two surveys together in order to propose an account for why Riau Indonesian and its relatives exhibit their isolating grammatical profile.

The sociohistorical facts described in Section 3 provide no positive evidence whatsoever for any radical disruption in transmission brought about by imperfect

adult second-language acquisition in the recent history of Riau Indonesian, and plenty of reason to believe that there was unlikely to have been any such disruption in the course of the last couple of millennia. The linguistic facts described in Section 4 show that the isolating profile of Riau Indonesian and its relatives is shared by a swathe of languages across the large territorial expanse that constitutes the Mekong-Mamberamo linguistic area, and therefore that it is of great antiquity. In conjunction, the sociohistorical and the linguistic facts lead inexorably towards the dual-heritage story of Riau Indonesian and its relatives represented in Figure 1.

From the perspective of its Austronesian heritage, the isolating profile of Riau Indonesian and its relatives is argued to be a product of contact-induced simplification, much or most of which took place at an early stage, at or soon after the original intrusion of Austronesian languages into Nusantara. Alternatively, from the perspective of its Mekong-Mamberamo heritage, the isolating character of Riau Indonesian and its relatives may be viewed as a direct descendant of a Mekong-Mamberamo isolating profile dating back to time immemorial.

The story of Riau Indonesian and its relatives is thus played out in three acts: the first involving two separate Austronesian and Mekong-Mamberamo heritages, the second associated with their coming together some 3500–4000 years ago, and the third consisting of their subsequent life together culminating in contemporary Riau Indonesian and its relatives. We present each of these three acts in turn.

5.1   Act 1: Two separate heritages

The story of the Austronesian origins in Taiwan is well-known and need not be repeated here; instead we focus here on the second, Mekong-Mamberamo heritage.

Imagine we had access to a time machine which could take us on a linguistic field-trip back into the past. First we would pick a location, somewhere in the Malay/Indonesian heartland, in central Sumatra or western Borneo. Then we would go back, say, ten thousand years. And from there we would proceed on a journey back towards the present, stopping off every few hundred years for observations. What would we find?

In terms of populations, we would observe peoples coming and going; this much we know from genetics. In particular, several studies provide evidence for one or more migrations out of Mainland Southeast Asia into our region prior to say four thousand years ago (Karafet et al. 2010; Jinam et al. 2012 and Vallee; Luciani and Cox 2016). Similarly, in terms of cultures, we would see various crops and material artefacts appearing and sometimes also disappearing. For example, also at different times prior to say four thousand years ago, we would see bananas arriving from the east (Kennedy 2008; Denham and Donohue 2009), pigs brought in from the northwest (Larson et al. 2007; Dobney, Cucchi and Larson 2008), and a tradition

of cord or basketry-wrapped paddle-impressed pottery introduced from Mainland Southeast Asia (Spriggs 2007). But what about the languages?

Even as far back as ten thousand years ago, it is likely that we would find languages conforming to the overall Mekong-Mamberamo profile as described in (9), including, of course, isolating structure. A similar conclusion is reached also by Donohue and Denham (this volume), who suggest that "many of the languages of the region that were spoken in place prior to the dispersal of Austronesian languages also showed an isolating profile". Of course we cannot be certain; it has been suggested, for example by Goldammer (2011), that geographical regions may change their typological profiles over time. However, given the geographical extent of the current Mekong-Mamberamo region, it is reasonable to assume that much or most of it was in place for at least several thousand years. And of course, as we proceed forward in time, the prevalence of Mekong-Mamberamo properties would generally increase.

The question arises: How did Mekong-Mamberamo languages develop their characteristic grammatical profile, and more specifically, how did they acquire their characteristic simplicity, as reflected in properties 11–17 in (9)? Gil (2015: 336–340) offers a number of highly speculative answers. One suggests that it might be a relic from an earlier stage in the evolution of language itself. A second appeals to the huge number of islands that occupy a large proportion of the relevant area. And a third attributes the simple profile to contact with other hominin species.[38] Unfortunately, at present, we simply do not know enough to make an informed choice between these and possibly other accounts.

For McWhorter, however, such agnosticism is intrinsically ill-conceived. Central to his word-view is the insistence that a simple grammatical profile of the kind characteristic of Mekong-Mamberamo languages can *only* be due to language contact.[39] True, several scholars, including LaPolla (2001), Enfield

---

38. The field of hominin evolution is in a state of continual flux. The appeal to contact with hominins in Gil (2015) was motivated in part by the original dating of *Homo floresiensis* remains, by Morwood et al. (2004), to a time as recent as 12,000 years ago. Subsequent work however, by Sutikna et al. (2016), has pushed that date back to some 50,000 years ago, thereby substantially decreasing the likelihood of *Homo floresiensis* being relevant to the currently observable linguistic landscape. On the other hand, a more recent report by Jacobs et al. (2019) provides evidence that Denisovans may have been alive and mating with modern humans as recently as 15,000 years ago in New Guinea, suggesting that the possibility of some linguistic signal of hominin admixture being reflected in the Mekong-Mamberamo linguistic area cannot be entirely written off.

39. Thus, McWhorter (2008: 169) writes:

In the uninterrupted transmission of a human language, radical loss of complexity throughout the grammar is neither normal, occasional, nor rare, but **impossible**. The natural state of human language is one saddled with accreted complexity unnecessary to communication. Wherever this complexity is radically abbreviated overall rather than in scattered, local fashion, this is not just sometimes, but

(2011), and Post (2015) have pointed to the "creole-like" appearance of Mainland Southeast Asian languages, while for the other end of the Mekong-Mamberamo area, Reesink (2002: 26) writes that "At the level of coffee-table talks about the languages of the Bird's head and Halmahera, characterisations like 'these languages are rather creole-like' can be heard." But still, this is not to say that they *are* creoles. Indeed, asserting that the entirety of the ancient Mekong-Mamberamo area is the product of ancient creolisation or other such disruptions in transmission would be no less speculative than any of the speculations offered in Gil (2015), and if anything even more lacking in empirical support. In particular, one cannot but wonder why such rampant creolisation might have taken place in the ancient Mekong-Mamberamo region and not in any number of other regions of high linguistic diversity around the world, such as, for example, the Caucasus, East Africa or Central America. Better at present to conclude that we simply do not know why the ancient Mekong-Mamberamo area developed in the way that it did.

The geographical extent and presumed antiquity of the Mekong-Mamberamo area fly in the face of the widely accepted universal tendency for languages to accumulate morphological complexity over the course of time, for which McWhorter is a leading spokesperson. Thus, McWhorter (2007: 5) asserts that "[t]he natural state of human language is one saddled with accreted complexity unnecessary to communication." To make this point, he has a stock collection of languages illustrating what he considers to be the default state of complexity, including Tsez (2001a, 2007: 30–31), Chechen (2007: 11–12), Navajo (2007: 17, 228), and a typical "Algonquian language" (2007: 7–8, 273), which he often draws upon in order to emphasise the ways in which they contrast with creole languages. Indeed, it may be the case that, in terms of morphological complexity, more languages are like Tsez. Chechen, and Navajo than are like Riau Indonesian and other Mekong-Mamberamo languages, but the issue is not one to be decided by the world's languages all lining up to vote. The facts are that the Mekong-Mamberamo is exceptional as the only area of its size in the world with such a predominance of isolating and otherwise grammatically simple languages; this calls for an explanation; but as argued above, we can at present only speculate what form such an explanation might take.

In the meantime, in the absence of such an explanation, one is led to the alternative conclusion that the isolating profile of Mekong-Mamberamo languages represents a potential pole of stability around which a large number of languages may coalesce for an extended period of time. Such diachronic stability dovetails nicely with arguments, presented in Gil (2000, 2013, 2017a), to the effect that the

---

always caused by a sociohistorical situation in which non-native acquisition of the language was widespread enough that grammar was transmitted to new generations in a significantly simplified form. This is true not only in the extreme case of pidgins and creoles, but also to a lesser but robust extent in many languages of the world.

IMA (Isolating-Monocategorial-Associational) language profile of Riau Indonesian and other similar languages represents a default setting from cognitive, ontogenetic and phylogenetic points of view. In doing so, it turns the table on McWhorter's puzzlement as to why some languages are isolating, suggesting instead that we should be wondering why it is not the case that all languages look like Riau Indonesian. Functionally, there is no reason for them not to; as argued in Gil (2009a), IMA language structure is sufficient to sail a boat, and indeed to run an advanced modern nation state. Rather, it is Tsez, Chechen, Navajo and for that matter also Standard Average European that represent a curious dissonance, with their layers upon layers of accreted complexity, resulting from diachronic processes, leading them to a place far removed from the pristine simplicity of a more IMA-like language such as Riau Indonesian and other Mekong-Mamberamo languages.

Meanwhile, back in our time machine at ten thousand years ago, while the grammatical profiles we would be observing would be the largely familiar Mekong-Mamberamo ones, we would probably not recognise most of the words. Indeed, this would probably remain the case at eight, six, or even four thousand years ago. In part this is because of the rate at which the lexical signal decays, but this is not the whole story: in other parts of the world, such as, for example, the Middle East, or the Pontic steppe of Eastern Europe, we would in fact be able to recognise many of the words six or four thousand years ago. However, at some point in time, around 3500–4000 years ago, the words in our central Sumatra or western Borneo site would rather suddenly become familiar to us, clearly related to those in contemporary Malay/Indonesian. Which brings us to Act 2 of the story.

5.2   Act 2: The coming together

Around 3500 to 4000 years ago, the Austronesian and Mekong-Mamberamo heritages of Riau Indonesian and its relatives converged. The conventional way of describing this event is that Austronesian languages spread into Nusantara, undergoing simplification and adopting the isolating profile. But looking at things a bit differently, why not say instead that there were Mekong-Mamberamo languages that, while staying put, underwent an infusion of lexical material originating from outside the region, from locations further to the north? Or, to co-opt the title of Donohue and Denham (this volume), "became Austronesian".

While this coming together of Austronesian and Mekong-Mamberamo heritages constitutes language contact par excellence, a more precise determination of the mechanisms underlying such contact is harder to come by, given the time depth and lack of direct evidence. Three potential mechanisms for such contact are available. The first two involve assimilation, either through metatypy, along the lines proposed by Ross (1996) for the Austronesian language Takia of

Papua New Guinea, or through relexification, following Lefebvre (1986, 1997) and others. Broadly speaking, these two mechanisms are consonant with the two alternative perspectives associated, respectively, with the Austronesian and Mekong-Mamberamo heritages. Adopting an Austronesian perspective, the incoming Austronesian languages maintained their lexicons, while picking up grammatical features of the languages that were there before them, including their isolating profile, by means of metatypy. Conversely, looking at things the Mekong-Mamberamo way, the in situ Mekong-Mamberamo languages kept the lions' share of their grammar, including their isolating structure, while adopting a new lexicon, in a process of relexification.

The third potential mechanism is creolisation, a break in transmission due to rampant imperfect adult second-language acquisition. Whereas metatypy and relexification both acknowledge the dual Austronesian / Mekong-Mamberamo heritage of the languages in question, creolisation, at least in its more radical guises, essentially denies the Mekong-Mamberamo heritage, viewing the story of Riau Indonesian and its relatives entirely as "something that happened to" Austronesian. In particular, it views the adoption of an isolating profile by the incoming Austronesian languages as being due in its entirety to processes such as imperfect adult second-language acquisition, while having nothing at all to do with the isolating nature of the languages that were there already, and with which the incoming Austronesian languages came into contact.

However, as argued in Gil (2015: 335), contact-induced simplification simply cannot be the whole story behind the convergence of Austronesian and Mekong-Mamberamo heritages. Whereas assimilation, be it via metatypy or relexification, may potentially account for all 17 Mekong-Mamberamo properties in (9) above, simplification is relevant only for those Mekong-Mamberamo properties that are simpler than their non-Mekong-Mamberamo counterparts, namely, numbers 11–17, among which, of course, is isolating structure. In contrast, the first 10 Mekong-Mamberamo properties can only be accounted for in terms of general processes of assimilation.

One could conceivably construct an Occam's Razor argument to the effect that the remaining seven Mekong-Mamberamo properties, including isolating structure, should also be accounted for in terms of general processes of assimilation, without any need to invoke contact-induced simplification. However, such an Occam Razor's argument only works under the assumption that we are dealing with a single event whose nature we are trying to ascertain. In actual fact, however, there is every reason to believe that what we now observe as the Austronesian mid-section of the Mekong-Mamberamo linguistic area is the product of multiple events taking place in different places, at different times, and involving different subsets of the 17 Mekong-Mamberamo properties.

Some specific linguistic arguments in support of the heterogeneity of the Austronesian spread into Nusantara are provided in Donohue and Denham (this volume), who suggest that "it is best to think of the process of 'becoming Austronesian' as not being a single uniform process, but rather different processes involving different starting points, different trajectories, and different time spans in different societies". In particular, in terms of their three "dimensions of Austronesian-ness", Donohue and Denham characterise Riau Indonesian and Papuan Malay as having Austronesian-like lexicon and phonology but lacking Austronesian-like morphosyntax, and argue that these two Malay/Indonesian koinés are creoles. Crucially though, unlike the creole accounts proposed by Adelaar and Prentice and by McWhorter, the creolisation that Donohue and Denham are referring to here is ancient, possibly dating as far back as the original expansion of Austronesian languages into Nusantara. As such, the arguments put forward in this chapter against a more recent creolisation account do not apply to the Donohue and Denham proposal.

In summary, then, at the present stage of our knowledge, we are not in a position to determine whether the isolating profile emerging from the coming together of the Austronesian and Mekong-Mamberamo heritages is a product of general assimilation, involving either metatypy or relexification, or of creolisation. Indeed, it is more than likely to be a complex combination of all of these mechanisms applying at different times and places. In particular, although, pace McWhorter, Riau Indonesian and its relatives are not recently creolised languages, they may possibly owe their isolating profiles to processes of creolisation undergone by their Austronesian ancestors in the distant past.

One apparent argument in favour of metatypy and against alternative relexification or creolisation accounts might involve the presence of supposedly conservative morphology, and the suggestion that, given the general resistance of morphology to borrowing, such morphology can only be accounted for by in terms of direct inheritance. Thus, to the extent that the languages of Nusantara exhibit Austronesian or Malayo-Polynesian morphology, one might consider this to be a clear indicator of direct descent from proto-Austronesian or Malayo-Polynesian. However, the force of this argument is weakened by two observations. The first is that in the case of isolating languages there is precious little morphology to appeal to, while the second is that what little morphology there is would in fact appear to be readily borrowable.

Table 5 summarises the major affixes of Rian Indonesian and some of its Malayic relatives, and provides their reconstructions. The first two columns present the affixes in Riau Indonesian and their classification as weakly bound (W), strongly bound (S), or both (W/S), in accordance with Gil (this volume, Chapter 1, Table 2). The third column presents Proto-Malayic reconstructions, when present,

Table 5. Reconstructions of major affixes in Riau Indonesian and its relatives

	Riau Indonesian	Morpheme type	Proto-Malayic	Earlier reconstruction			Function
1	-kan	W	*akAn	Ø			applicative/causative
2	ber-	W/S	*(mb)Ar-	Ø			medial
3	di-	W	*di	*di	PAN		locative
4	-nya	W	*-ɲa	*ni-a	PAN		3SG agent/possessor
5	se-	W/S	*saʔ-	*sa$_3$	PAN		one
6	ke-	W/S	*kə-	*ka$_3$	PAN	1	oblique
7	i-	W/S	*i-	*i$_2$	PAN	1	location
8	s-	W/S	*s- [1]	*sa$_2$	PAN	1	locative
9	g-	W/S	*bagay [1]	*bagay	PWMP		same kind/type
10	si-	S	*si- [1]	*si$_1$	PAN		nominative for names
11	-in	W	Ø	*-i$_2$/*-in-/*-en	PAN	1	location/perfective/passive
12	ter-	W	*tAr-	*taR-	PAN		spontaneous/involuntary
13	Ø		*pAr-	*paR-$_1$	PWMP		deverbal nouns
14	(me-)(N-)	W/S S	*mAN-	*maŋ-	PMP	2	active
15	Ø		*pAN-	*paŋ-	PWMP	1	instruments/products
16	Ø		*-aʔ	*-a$_3$	PAN	1	subjunctive
17	Ø		*-i	*-i$_2$	PMP		transitive
18	-an	W/S	*-an/*-An	*-an	PAN		locative/location
19	Ø		*kA- -an	*ka- -an$_{123}$	PAN		adversative/location/abstract
20	Ø		*pAr- -an	*paR- -an	PWMP		deverbal nouns of location
21	Ø		*pAN- -an	*paŋ- -an	PWMP	3	abstract

of the corresponding Riau Indonesian affixes as well as of other affixes not present
in Riau Indonesian, following Adelaar (1984, 1992).[40] The fourth column presents
earlier reconstructions, when present, of the same affixes for the earliest avail-
able stage, Proto-Western-Malayo-Polynesian (PWMP), Proto-Malayo-Polynesian

---

40. In three cases, marked with a "1", Adelaar does not provide reconstructions; instead the
reconstructions indicated are proposed by the present author.

(PMP), or Proto-Austronesian (PAN), following Blust and Trussel (2016).[41] And the fifth column provides a brief and simplified indication of the function of each affix, largely following Blust and Trussel. The rows of the table represent the union of two sets: the affixes of Riau Indonesian and the affixes of Proto-Malayic as reconstructed by Adelaar. Omitted, therefore, from Table 2 are other affixes present in other Malayic varieties but absent from both Riau Indonesian and Proto-Malayic, for example the associative marker -ă in Tapan, discussed in Section 4.3 above.

The affixes represented in Table 5 are arranged in rough and ready groups in accordance with the nature of the challenge that they pose to relexification and creole accounts of Riau Indonesian and its relatives. To being with, items 1 and 2 do not have reconstructions earlier than Proto-Malayic; accordingly, they may be considered to be innovations that arose subsequent to the process of simplification associated with the original spread of Austronesian languages into the region. In contrast, the remaining items in the table all reconstruct to PWMP, PMP or PAN and therefore constitute apparent instances of conservative morphology that need to be accounted for in one way or another. Of these 19 items, the first eight, namely items 3–10, consist of items that are reconstructed as free forms. Although there is no reason to believe that reconstructions of wordhood properties are any more well-founded than the often unreliable synchronic descriptions of such properties, the presence of such reconstructions suggests that in many cases, what appears to be a affix may instead be a form that was originally free, and only became an affix subsequent to the intrusion of Austronesian languages into the region and ensuing simplification. Thus, although conservative, such forms also do not stand in conflict with the relexification and creole accounts.[42]

However, the remaining 11 affixes, items 11–21, are reconstructed as affixes for PWMP, PMP or PAN, raising the question how they managed to survive the radical disruption posited to have taken place at the original expansion of

---

41. For the most part, Blust and Trussel are explicit in assigning the appropriate Malay forms to their cognate sets; however, in a few cases they do not provide a Malay form, in which case, the assignment is made by the present author; these cases are marked with a "1". For two cases, Blust and Trussel do not provide a reconstruction. In one of these, marked with a "2", the reconstruction is from Blust (2009: 359), while in the other, marked with a "3", the reconstruction is by the present author, drawing on the reconstructions of the two constituent parts of the circumfix given elsewhere in the table.

42. Cf. An analogous point can be made with regard to the bound person-number indexes in the languages of the Central Malayo-Polynesian group. In recognition of their obvious cognacy with Austronesian independent pronouns, Blust (1993: 258–259) reconstructs a paradigm of bound indexes for Proto-Central-Malayo-Polynesian; however, in response, Donohue and Grimes (2008: 131–132) argue that such bound forms are the product of independent innovations.

Austronesian languages into Nusantara. The answer proposed here is that they didn't survive the disruption, but instead were lost, and then subsequently reintroduced into the daughter languages by means of borrowing. Such an account faces two obvious problems. First, it violates Occam's razor, positing two separate events, loss and reintroduction, rather than straightforward continuity. And secondly, it runs counter to the commonly-held assumption that morphology is rarely borrowed, though as suggested in Seifart (2013, 2015), morphological borrowings are actually more widespread than is usually supposed; see also van Klinken and Hajek (this volume) for a detailed case study from Timor. Nevertheless, there would seem to be good reasons to countenance such an account for at least some of the cases in question.

First, there is one case where an analogous account stands out as the most likely story for what happened, namely, the Riau Indonesian suffix -in (item 11 in Table 5). As noted above, Riau Indonesian -in was borrowed from Jakarta Indonesian, which in turn borrowed it from Balinese; it is not reconstructible to Proto-Malayic. However, Balinese -in is most likely inherited from a Proto-Austronesian suffix; in fact, there are three possible candidates, *i, *-in- and *-en (though the literature does not come down definitely in support of any one of them). Thus, Riau Indonesian as well as Jakarta Indonesian -in would appear to be a descendant from a Proto-Austronesian suffix that was lost but then reintroduced through borrowing. The claim, then, is that the other affixes may have undergone a similar trajectory of loss and subsequent reintroduction, albeit at a much earlier stage, as evidenced by the fact that they are already present once again by Proto-Malayic times.

As we shall now see, there are numerous clear examples of affix borrowing in Austronesian languages, in both the near and the distant past, involving the affixes listed in Table 5. Evidence for affix borrowing in contemporary Malayic languages is readily observable in naturalistic speech examples such as the following, cited in Fadlul et al. (2013):

(12) Padang Minangkabau   [977818100824190107]
      A   sabana nyo   kan ambo lah mencubo
      DEIC one:true ASSOC Q   1SG   PRF AG:try
      'You see, actually I've already tried'
(13) Padang Minangkabau   [302152093638160207]
      Tu   bara   beradiak   siko
      DEM.DIST NON.PAT:what NON.PAT:younger.sibling LOC:DEM:DEM.PROX
      tu   Butet Dalima
      DEM.DIST Butet Dalima
      'So how many brothers and sisters does Butet Dalima have here?'

Although the above utterances are in the Padang dialect of Minangkabau, in each example there is a prefix that is associated with the local variety of Indonesian, Sumatra Barat Indonesian: in (12) the agent-oriented generalised voice prefix *men-* in *mencubo*, and in (13) the depatientive prefix *ber-* – the Minangkabau equivalents of these two prefixes would be *man-* and *bar-*. Fadlul et al consider three possible analyses for examples such as these, borrowing, code-switching and register-switching, and opt tentatively for borrowing.[43] But even if these forms have not yet acquired the status of borrowings, examples such as these suggest that they may be on the way to doing so, thereby highlighting the propensity for these two prefixes to undergo borrowing from one language variety into another.

Numerous other examples can be adduced of recent cross-dialectal borrowings of affixes within Malay/Indonesian and/or the Malayic family. Two further cases of borrowings from Indonesian into Minangkabau are mentioned in Crouch (this volume), involving the generalised voice markers *(me-)(N-)* and *-kan*. For the agent-oriented generalised voice marker, the original Minangkabau forms are either *maN-* or *ma-*; however, in colloquial Minangkabau the form *N-*, borrowed from Sumatra Barat Indonesian, is gaining currency. For the end-point-oriented generalised voice marker, the original Minangkabau form is *-an*; however, *-kan* also occurs, both colloquially and formally, suggesting that it might have a dual source in both Sumatra Barat Indonesian and in Standard Indonesian, both of which have *-kan*. Another example, mentioned in Section 2.2 above, is the borrowing of the suffix *-an* into Riau Indonesian from either Jakarta Indonesian or Standard Indonesian. Yet another example is the borrowing of the Minangkabau instrument-forming prefix *paN-* into Sumatra Barat Indonesian, in forms such as *pancas* 'charger'. At the other end of the archipelago, Kluge (2016) presents a detailed description of the alternation in Papuan Malay, between two realisations of the depatientive prefix, *ba-* and *ber-*, arguing that the latter is most appropriately analyzed as a borrowing from Standard Indonesian. Examples such as these and many others provide a vivid instantiation of the ease in which morphology can be borrowed from one Malayic variety into another.

Casting a wider net, there is ample evidence that similar morphological borrowings took place also in earlier periods, and among a wider range of languages. One example is provided by the prefix *di-* in its function as patient-oriented generalised voice marker (item 3 in Table 5). In Riau Indonesian, *di-* has two functions, patient-oriented generalised voice marker and locative, though in Gil (2002) it is argued that these are two instantiations of a single macrofunction. In contrast, it is only the locative function that is reconstructed for Proto-Malayic and PAN;

---

43. Two arguments presented by Fadlul et al. in support of a borrowing analysis are (a) such forms occur also in texts where there is no independent evidence for code or register switching, and (b) such forms are not necessarily associated with specific sociolinguistic contexts of the kind that are conducive to code or register switching,

the voice marker function is considered by Adelaar (1992) to be an innovation within Malayic – one shared by Riau Indonesian and many other Malayic varieties. A number of different accounts have been proposed for the development of the patient-oriented generalised voice marker *di-*; see van den Berg (2004) for a useful survey. However, its origins are of no concern to us here. Rather, the crucial point is that the patient-oriented generalised voice marker *di-* occurs also outside of Malayic, in languages such as Toba Batak, Lampung, Sundanese, Javanese, Mandar, and Toraja, to name just a few. Since the languages with patient-oriented generalised voice marker *di-* do not form a genealogical subgroup, the remaining explanation for the distribution of voice marker *di-* is language contact. Thus, without knowing when and where and in what language the patient-oriented generalised voice marking function of *di-* originated, we can conclude that in some of the languages where it is present today, it must have been borrowed from elsewhere.

More dramatically, several cases may be observed of Malay or otherwise Austronesian morphology being borrowed into non-Austronesian languages, ranging from the Malay Peninsula to the New Guinea Bird's Head. For example, in the Aslian language Semelai, described by Kruspe (2004), there are two systems of affixation, non-concatenative and concatenative, where the concatenative system consists entirely of affixes borrowed from Malay, including *br-, tr-, par- ~ pr-, m(N)-, -iʔ*, and *-an* (corresponding respectively to items 2,12,13,15,18,19 in Table 5). Were it not for the conservative Aslian non-concatenative morphology, one could almost imagine characterising Semelai as an Austronesian language with conservative morphology that underwent Aslian relexification. But more to the point here, given the manifest borrowability of these suffixes, what is to stop one from turning the tables and characterising Malay as a non-Austronesian language that underwent Austronesian relexification, and, in addition, just like Semelai, borrowed a battery of affixes from some other Austronesian language?

The Semelai case is anything but exceptional. The borrowing of affixes from Malay or other Austronesian languages is in fact widespread throughout the Aslian language family. Several examples are cited by Matisoff (2003), relying largely on personal communications from Geoffrey Benjamin; these include Temiar *bar-, tɛr-, pə*, and *ma-* (items 2,12,13,14 in Table 5), Semai *br-* and *pr-* (items 2,13), Jah Hut *pr- and mʔ-* (items 13,14), and Kentaqbong *pi-* and *maʔ-* (items 13,14). Similar examples are also cited by Burenhult (2005) for Jahai, including *br-, tr-* and *pr-* (items 2,12,13), though for the most part he does not does not dwell on their Malay or Austronesian provenance.[44] The ubiquity of these borrowings raises several inter-

---

44. It is possible that one or more of the borrowed forms based on *p* do not originate from the cognate set represented in item 13 Table 5, but rather from an alternative cognate set associated with the PAN causative prefix **pa-₂*, which is not indicated in Table 5 because it is not present in either Riau Indonesian or Adelaar's Proto-Malayic reconstructions. See also footnote 46 below.

related questions which must be left for future research: whether the donor language was Malay or perhaps some other Austronesian language, at what period the borrowings took place, and whether the borrowings were into each language individually or alternatively into various higher-level language groupings within Aslian. Still, what these examples make patently clear is that many if not all of the affixes in Table 5 are potentially borrowable.

What is more, the borrowing of Malay or Austronesian affixes into Aslian languages is echoed right at the other end of the archipelago, with similar morphological borrowings, sometimes of the very same forms, into diverse non-Austronesian languages of Wallacea and western New Guinea. One striking example, from Reesink (2002: 20–21, 2009) is the borrowing of prenasalisation (item 14 in Table 5) into the Bird's Head isolate language Hatam, as manifest in alternations such as *kes ~ ngges* 'drop', 'let go'.[45] Another borrowing of the same morpheme, described by Voorhoeve (1982: 11) and Reesink (1998: 633, 2002: 20), is into the North-Halmaheran language West Makian, where it surfaces as the prefix *ma-*. Also found in West Makian are demonstrative and locative paradigms, described by Voorhoeve (1982: 18) and attributed by Reesink (2002: 20) to borrowing from Austronesian; among others, these paradigms would appear to contain reflexes of *i-* and *s-* (items 7, 8 in Table 5).[46] Again, as with the Aslian languages above, the historical details of these borrowings remain to be worked out, but whatever the diachronic scenarios involved, they provide yet further striking evidence for the borrowability of Austronesian morphology into non-Austronesian languages.[47]

---

45. Somewhat similarly, in the North Halmaheran language Ternate, Hayami-Allen (2001: 113–115) reports a root-initial process of lenition, which "appears to be a prefixation of a nasal phoneme"; although she does not offer any suggestions with regard to its origins, an Austronesian provenance, again in the item 14 cognate set, would seem likely.

46. Yet additional morphological borrowings from Austronesian into non-Austronesian would seem to involve reflexes of the PAN causative prefix **pa-$_2$* mentioned in footnote 44 above. One example from the Timor-Alor-Pantar language Bunaq, cited by Schapper (2010: 345), is the causative prefix *ha-*, borrowed from the Austronesian language Tetun. Another possible example, from the North-Halmaheran language West Makian, cited by Voorhoeve (1982: 11), is the verbal prefix *fV-*, which also has causative among its described functions. Additional examples, from the Bird's Head, are Mpur *fa-* (Odé 2002: 56–57, Reesink 2002: 20, 27) and Hatam *ha-* (Reesink 2002: 20,27).

47. Another borrowing proposal is that of van Hasselt (1905: 11), who suggests that the Malay prefix *ber-* is the source of a wide range of forms with a verbalising function throughout the Bird's Head and Cenderawasih Bay region, in both Austronesian languages, e.g. Biak *ve*, Ansus *ve*, Wamesa *ve*, Waropen *we*, Moor *ve*, and also non-Austronesian languages, e.g. Yawa *ve*, Meyah *ebe*, Hatam *bi*, Mpur *bi,* Abun *bi*. However, Gil (2017b) argues against van Hasselt's proposal, suggesting that the source for all these forms is Proto-Central-Eastern-Malayo-Polynesian **bai* 'do', and that the resemblance to Malay *ber-* is mere coincidence.

Thus, as documented in the preceding pages, most if not all of the affixes represented in Table 5 are eminently borrowable, be it inter-dialectal borrowing from one Malay/Indonesian variety to another, borrowing between a Malayic variety and some other Austronesian language, or, most notably, from Malay or another related Austronesian language into some non-Austronesian language, ranging from Aslian in the west to Bird's Head languages in the east. The widespread borrowability of Malay and Austronesian morphology thus significantly weakens the force of any argument to the effect that Malayic varieties *must* be Austronesian simply because of the presence of a dozen or so affixes that are reconstructible to PAN. In doing so, it also refutes the argument that Riau Indonesian and its relatives must be the product of metatypy, not relexification or creolisation, because of the presence of supposedly inherited morphology. Of course, this is not to deny the Austronesian heritage of Malay, a heritage that it shares with, say, Javanese and Tagalog but not Semelai or Hatam. Rather, the conclusion to be drawn from the borrowability of Malay and Austronesian morphology is that the presence of such supposedly conservative affixes cannot be invoked in order to deny the significance, alongside Austronesian, of a second and parallel heritage for Riau Indonesian and its relatives: the Mekong-Mamberamo heritage.

Whatever the mechanisms involved in the coming together of the Austronesian and Mekong-Mamberamo heritages, metatypy, relexification or creolisation, these processes took place in large part at their first coming together some 3500–4000 years ago – after which the isolating character of the Mekong-Mamberamo Austronesian languages was preserved, more or less, all the way to the present. As argued in Gil (2015: 329–330), we can make use of various subsequent expansions of Austronesian languages out of the Mekong-Mamberamo linguistic area in order to gauge the antiquity of specific Mekong-Mamberamo properties, suggesting that such properties might have been present at their Mekong-Mamberamo points of origin at the time when the languages embarked on their dispersals. Thus, for example, the isolating nature of many Oceanic languages suggests that this isolating profile might have been largely present in Proto-Oceanic, spoken in the Bismarck archipelago around 3,500–4,000 years ago.[48]

This view of isolating structure as being of considerable antiquity in the Austronesian languages of Nusantara is one that is shared, to variable degrees, by

---

48. A somewhat different picture, however, emerges in the case of Malagasy, which is argued to represent a migration out of Southeast Borneo some 1500 years ago. Unlike its closest stay-behind relatives of the South Barito subgroup, which are largely isolating, Malagasy maintains a significant amount of older Austronesian morphology. This would seem to suggest that in one part of the Mekong-Mamberamo area at least, the adoption of an isolating profile was a more protracted process.

several scholars in the present volume. For Cham, Brunelle (this volume) suggests that most of its isolating profile was already in place prior to its arrival in Vietnam at the middle of the first millennium AD. For Javanese, Connors (this volume) argues that the complexification characteristic of Central Javanese is innovative, and that the more isolating profile of the peripheral dialects is conservative. For Lio and other languages of central Flores, Elias (this volume) dates their isolating profile back to the original arrival of Austronesian languages in the region. And for at least some languages of Timor, both Austronesian and non-Austronesian, Schapper (this volume) also argues for an ancient presence of isolating grammatical structure.

As mentioned earlier, one of the most striking properties of the isolating profile in the Mekong-Mamberamo area is its typological stability and resistance to the usual processes of accretion of morphological complexity. Some of the ways in which this plays out are dealt with below, in Act 3.

5.3   Act 3: A new hybrid identity

Since the coming together of the Austronesian and Mekong-Mamberamo heritages some 3500 to 4000 years ago, a new hybrid identity has emerged, that of Austronesian Mekong-Mamberamo languages. And like other hybrids in nature, it would seem to be endowed with a measure of hybrid vigour.

One important factor underpinning the typological stability of Austronesian Mekong-Mamberamo languages would seem to be the presence of a critical mass that acts as a counterweight to the potential development of morphological complexity. A case in point is provided by the complex morphological alternations in Kerinci discussed in Section 4.3. On the one hand, as suggested there, these innovations originate in phrasal phonological processes that are productive throughout the region, as a result of which one might expect similar morphological complexifications to emerge in other Malayic varieties in the future. On the other hand, observation of the sociolinguistic landscape of Kerinci suggests that the complex system of absolute-oblique alternations is under severe threat of erosion.

As noted earlier, these alternations display substantial dialectal variation. For example, in three locations in close proximity, the alternating absolute and oblique forms for the word for 'girl' are *gado^ht ~ gadi^hk* in Koto Pudung, *gadoyh ~ gadih* in Koto Tuo, and *gadʌ͡ɛh ~ gadɨyh* in Tanjung Pauh Mudik (McKinnon 2011: 7). So how do people communicate with each other, for example when coming to market in the main urban centre of Sungai Penuh? In fact, they have three choices available: the local variety of Indonesian, the local variety of Minangkabau, but also a common Kerinci koiné that would appear to be in the process of developing throughout the Kerinci valley. And not surprisingly, in this emerging Kerinci koiné, the absolute-oblique alternations would seem to be in the course of breaking down.

For example, for 'girl', the single unalternating form that appears to be wining out is *gadih*. What is more, some younger people, in Sungai Penuh but also in other villages, appear to be acquiring some version of this koiné as their first language: these, then, are native speakers of Kerinci whose mastery of the complex absolute-oblique alternations ranges from partial to non-existent.

As always, one can ask whether such morphological simplification is the product of imperfect adult second-language acquisition or of assimilation to some other available typological profile. And again, the answer can be obtained by comparing the situation in Kerinci to analogous cases of inter-dialectal communication in other parts of the world. For example, when speakers of Czech and Slovak communicate with each other, for a form such as 'street' in the nominative singular, they might use either *ulice* with the Czech suffix *-e,* or *ulica* with the Slovak suffix *-a*. Crucially, however, they will never simplify by using a bare form **ulic* (Jan Chromý pers. comm.). Similarly, when speakers of the northern (Galilee) and central ("triangle") dialects of Rural Palestinian Arabic are talking to each other, for a form meaning 'with you (plural)' they might choose either *ʕendkom*, with the northern-dialect suffix *-kom* (also marked for masculine), or *ʕendkū* with the central-dialect suffix *-kū* (unmarked for gender), but never the bare form **ʕend* (Jad Kadan pers. comm.). Examples such as these can be multiplied at will. What is common to all of them is that there is no move towards simplification and the lowest common denominator: solving the inter-dialectal conflict by simply omitting the problematic affix and using a bare form of its host instead is not an option. And the reason for this is obvious: these languages are spoken in parts of the world where rich inflectional systems are the norm, as a result of which there is no available model for the use of bare forms.

Contrast this now with Kerinci. In Kerinci, most or all people are fluent also in the local Indonesian and the local Minangkabau, both of which lack the absolute-oblique alternation. Accordingly, these two languages provide a readily available model for a common Kerinci koiné that is simpler than any of the original Kerinci varieties – an option that is not available in the Czech/Slovak and Rural Palestinian Arabic cases. Thus, the loss of the absolute-oblique distinction in the Kerinci koiné may be attributed to assimilation to other languages with which it is in contact, thereby showing how the Mekong-Mamberamo isolating profile may rein in languages that are threatening to escape it, and in doing so preserve its typological stability. It is precisely this critical mass of the Mekong-Mamberamo area that prevents a language such as Kerinci from keeping on complexifying until it comes to resemble a language spoken in areas such as Europe, the Middle East, or Daghestan.

Of course, the story of Kerinci is not necessarily applicable to other locations in the Austronesian Mekong-Mamberamo speaking area: different things happen in different places. In some cases, simplification may indeed be due also to imperfect adult second-language acquisition; indeed, this may also be a contributing factor,

albeit of secondary importance, in the case of Riau Indonesian and other Malay/Indonesian koinés. On the other hand, assimilation may also lead in some cases to complexification. One example of this might be the Papuan Malay system of argument indexation, as represented in Table 4, Section 4.1, which is probably due in large part to contact with Biak and other languages of the South Halmahera West New Guinea subgroup of Austronesian. Another perhaps more striking example would be Makassarese Malay (Hanan 2015), sometimes characterised as a mixed language, containing several features of the relatively more complex morphology of Makassarese. Ultimately, however, all of these instances of complexification remain with the confines of the overall simplicity characteristic of the Mekong-Mamberamo area.

Alongside the above, another important kind of secondary complexification observable in the Mekong-Mamberamo region is that associated with the development of formal or standardised linguistic registers, of the type described by Crouch (this volume) for Minangkabau, Conners (this volume) for Javanese, and of course Malay/Indonesian here. On the face of it, such complexification would seem to run counter to the inverse correlation between grammatical and social complexity proposed by the likes of Dahl (2004, 2009), Nichols (2009), Sinnemäki (2009), and Trudgill (2009, 2011). On the other hand, it is consistent with the direct correlation between grammatical and social complexity proposed by Gil and Shen (2019) and Raviv, Meyer and Lev-Ari (2019), as discussed in Section 2.4.

At issue in cases such as these is the directionality of the diachronic process: Do standard varieties indeed develop by complexification from colloquial varieties, as suggested above, or do perhaps colloquial varieties result from the simplification of their standard counterparts? While Crouch takes no stand on this issue for Minangkabau, Conners argues in support of complexification for Javanese, and complexification is surely the direction of the process in Malay/Indonesian, as in many other similar cases.

Nevertheless, the opposite view would appear to be the one more commonly held. In its basest form, it reflects the prejudices that many people, including even quite a few linguists, hold against colloquial language varieties, deeming them to be "broken" forms of their standard-language counterparts – recall the discussion at the end of Section 2.1 above. But this view is also held by scholars who are free of such prejudices, such as McWhorter. In the passage cited in Section 2.3 above, McWhorter talks of "two passes" of simplification for Malay, one resulting in NCSLs such as Standard Malay/Indonesian, the second, applying to the output of the first, producing putative creoles such as the Malay/Indonesian koinés. Thus, McWhorter considers the grammatical profile of Standard Malay/Indonesian to be a way station on the diachronic path from Taiwanese and Philippine complexity to colloquial Malay/Indonesian simplicity.

The issue of directionality comes to the fore in the domain of clause structure and the contrast between the voice systems of Standard Malay/Indonesian and the colloquial varieties. The facts, as described in Gil (2002) and revisited in Gil (2015: 302–309), are that the voice systems of Malay/Indonesian varieties lie on a range between two abstract systems, which may be represented schematically as follows:

(14) *The Indonesian-Type Voice System*
    a. Clause Type 1: active
        AG-V
        preverbal argument is agent
    b. Clause Type 2: passive
        PAT-V
        preverbal argument is patient

(15) *The Sundic-Type Voice System*
    a. Clause Type 1: bare neutral
        V
        preference for preverbal argument to be agent
    b. Clause Type 2: generalised active
        AG-V
        strong preference for preverbal argument to be agent
    c. Clause Type 3: generalised passive
        PAT-V
        strong preference for preverbal argument to be patient
    d. Clause Type 4 (uncommon): doubly-marked neutral
        PAT-AG-V

Both Indonesian- and Sundic-Type voice systems make use of two generalised voice prefixes, one agent-oriented, the other patient-oriented; however, they do so in different ways.

In the more familiar Indonesian-Type voice system, represented in (14), the two prefixes are in complementary distribution, resulting in two voices: an active associated with the agent-oriented prefix, and a passive associated with the patient-oriented prefix. As in prototypical voice systems, active and passive voice differ grammatically; in particular, they unequivocally determine the thematic role of a preceding argument: agent in the case of active voice, patient in the case of passive voice.[49]

---

49. Some descriptions of the Indonesian-Type voice system make reference to a third clause type, often referred to as a "second passive", which is of the form AG=V, where a usually pronominal agent is cliticised to the verb, and the preceding argument, if present, is interpreted as patient.

In contrast, in the Sundic-Type voice system, represented in (15), each of the two prefixes may be either absent or present independently of the other, resulting in four generalised voices: a bare neutral voice associated with the absence of both prefixes, a generalised active associated with the agent-oriented prefix, a generalised passive associated with the patient-oriented prefix, and a doubly-marked neutral voice, associated with the cooccurrence of both prefixes. Unlike in prototypical voice systems, the thematic role of a preceding argument is not determined unequivocally, but only as a matter of preference; this is the reason for the use of the term "generalised". Unlike in the Indonesian-type voice system, where the markers have well-defined syntactic functions, in the Sundic-Type voice system, the markers' functions are semantic – asserting the presence of an argument bearing the relevant thematic role, agent or patient.[50]

While Standard Malay/Indonesian exhibits an Indonesian-Type voice system, many colloquial varieties of Malay/Indonesian exhibit generalised voice systems that are closer to, or identical to, the Sundic Type. For Malay/Indonesian koinés, Gil (2002) provides examples of all four clause types in (15) for Riau Indonesian, while Conners, Bowden and Gil (2015) provide similar examples for Jakarta Indonesian. For traditional Malay varieties, the four clause types in (15) are illustrated with the following examples from Siak Malay:[51]

(16)   Siak Malay
    a.   Clause Type 1: neutral
        Sepeda tarok *mano Din*   *a?*
        bicycle put   which HYP\Kudin DEIC
        'Where did you put the bicycle, Kudin?'

---

50. McWhorter (2007: 226) takes issue with the characterisation of these markers' functions as semantic, arguing instead that they are of a pragmatic nature; the distinction is important to him for his arguments concerning the directionality of the development. To be sure, the choice of generalised voice marker in a particular utterance is dependent on its discourse context, but having pragmatic consequences is hardly a feature that is specific to Sundic-Type generalised voice systems; the same is true also for the choice of active or passive markers in prototypical voice systems – as indeed is the case for grammatical markers in general.

51. In (16) and then (17) below, the Sundic-Type clause types are illustrated by constructions that exemplify the dispreferred word order that would not be available for the corresponding Indonesian-Type clause types and is hence diagnostic for the Sundic-Type system: for Clause Type 1 V preceded by patient, for Clause Type 2 AG-v preceded by patient, and for Clause Type 3 PAT-v preceded by agent. In (16) and (17), the V and (where present) its relevant argument are indicated in boldface. For example, in (16a), the V *tarok* is preceded by the patient *sepeda*, both in boldface.

   b. Clause Type 2: generalised active
       Poto tadi   nengok Vid
       photo PST.PROX AG:see HYP\David
       'Can I see the pictures you just took, David'
   c. Clause Type 3: generalised passive
       Mister   dipotret   teros
       white.person PAT:picture continue
       'You keep on taking pictures of me'
   d. Clause Type 4: doubly-marked neutral
       Dinunggu   situ
       PAT:AG:wait LOC:DEM:DEM.DIST
       'Somebody's waiting for you over there'

Several other Malay/Indonesian varieties exhibit subsets of the four Sundic-Type clause types. For example, Sabah Malay has no productive agent-oriented generalised voice marker and hence has only clause types (15a, c), while Kuala Lumpur Malay and Papuan Malay, have no productive generalised voice markers at all, and hence have only clause type (15a).

The question arises which of the two voice systems came first, the Indonesian-Type system of Standard Malay/Indonesian, or the Sundic-Type system of the colloquial varieties. The default position, as represented for example in Ross (2002: 470), would seem to be that the Standard Indonesian system came first, and then broke down in the colloquial varieties. In contrast, Benjamin (1993) suggests that the colloquial voice system was the original one, and then underwent grammaticalisation to result in the standard-language voice system.[52] Gil (2002) weighs the pros and cons of both possible directionalities but leaves the question open. However, subsequent work provides good reason to believe that Benjamin was right, and that it is the Sundic-Type voice system that is prior, and the Indonesian-Type voice system that developed out of it.

One argument in support of this, put forward already in Gil (2002), pertains to general principles of grammaticalisation: it is much more common for the functions of a given form to change from semantic to grammatical than the other way around. A second argument for the anteriority of the Sundic-Type voice system derives from its geographic distribution. Unbeknownst to me in 2002, the distribution of the Sundic-Type voice system is actually much wider than just Malay/Indonesian; it is present, in one guise or another, throughout much of Western Nusantara – hence the appellation "Sundic".

---

52. Thus Benjamin (1993: 363) writes:

Their present-day status as syntactic-function markers is the result of their having been standardised into a single paradigmatic set, where previously they were independent elements used optionally for whatever nuance of meaning they could bring to an utterance.

For colloquial Minangkabau, Crouch (this volume) provides examples of clause types (16a, b), and (16c) is also attested, the only one seeming to be lacking being (16d). Further afield, colloquial Javanese varieties typically instantiate all four Sundic-Type clause types. The following examples are from the Pemalang dialect of Javanese, a subvariety of the Pesisir Lor dialect discussed in Conners (this volume), taken from the Gil et al. (2015) corpus (compiled by Thomas Conners):

(17)   Pemalang Javanese
    a.   Clause Type 1: neutral   [393474152327220107]
        *Ya, kuwé **cockék**   **tampeg** rak nganggó rakèt*
        yes that shuttlecock:ASSOC slap   Q   AG:use racket
        'Well you should hit the shuttlecock with a racket'
    b.   Clause Type 2: generalised active   [497872204336170806]
        *Kuwé **Baurekso kuwé** asliné   **mendhemé**   nang kònòk*
        that   Baurekso that   original:ASSOC AG:bury:ASSOC LOC there
        'Baurekso was actually buried there'
    c.   Clause Type 3: generalised passive   [45713105800150103]
        Ci   ***Lingé***   *wis* ***didol?***
        HYP\older.sister Ling:ASSOC PRF PAT:sell
        'Has Ling already sold it?'
    d.   Clause Type 4: doubly-marked neutral   [500225143944260902]
        *Lampuné **dingèi**,   **dingèi**   siji-siji,   palu   céngkrong*
        lamp:ASSOC PAT:AG:give PAT:AG:give DISTR~one hammer sickle
        'They put a lamp with a hammer and sickle in each room'

Further examples of clause types (15b, c) in the Tengger dialect of Javanese are cited in Conners (2008: 137, 151). Moreover, it's not just Malay/Indonesian and Javanese, which, in spite of their early genealogical separation, have been in close contact with each other for many centuries. As shown in Gil (2015: 307–309) clause types (15b, c) are present also in remote Mentawai, spoken on some of the Barrier Islands to the west of Sumatra.

The widespread distribution of Sundic-Type voice systems across Western Nusantara suggests that they are of ancient provenance, pre-dating the formation of the Malayic language family, let alone the rise of Standard Malay/Indonesian. Moreover, the presence of at least some aspects of a Sundic-Type voice system in Mentawai, whose overall morphological profile is of relatively high complexity, shows that such systems are not restricted to isolating languages, and therefore not necessarily the products of the kind of simplification that leads to isolating structure. All this leads to the conclusion that the Sundic-Type voice system is diachronically prior, and that its presence in colloquial Malay/Indonesian languages was the source which gave rise, through natural processes of grammaticalisation, to the voice system of Standard Malay/Indonesian.

The present account thus departs from the commonly held view that the voice systems of Standard Malay/Indonesian and other similar "Indonesian-Type" languages represent the outcome of a more-or-less monotonic process of dissolution of the earlier Philippine-Type voice systems; see Starosta, Pawley and Reid (1982), Cole, Hermon and Yanti (2008), Blust (2009: 450–451), and others. More specifically, it differs from McWhorter (2007: 226–7, 2019), who considers the voice systems of colloquial Malay/Indonesian varieties to be derived from that of Standard Malay/Indonesian through contact-induced simplification and creolisation.

A potential objection to the diachronic scenario proposed here, raised also by McWhorter (2007: 226), is that the marker *(me-)(N-)* is reconstructed back to Proto-Malayo-Polynesian **maŋ-* (see item 14, Table 4 above), with functions similar to those observable in Standard Malay/Indonesian. However, the story being proposed here suggests that the PMP form **maŋ-* with its associated grammatical function was simply lost at the original coming together of Austronesian and Mekong-Mamberamo, following which, at some later stage, it was borrowed back into some ancestor of Proto-Malayic from some other Austronesian language, but with its semantic function, in accordance with the Sundic-Type voice system. As argued in Section 5.2 above, such a scenario involving loss and subsequent reborrowing, is not at all implausible; on the contrary, it represents a viable developmental trajectory for much or most of the apparently conservative Austronesian morphology of Riau Indonesian and its relatives.

The process of language standardisation, associated primarily with languages of empire such as Malay and Javanese, but to a lesser extent also with other languages such as Minangkabau, has a substantial impact on the linguistic profile of the region. Throughout this chapter, a number of examples were provided of morphological borrowings from Standard Malay/Indonesian into colloquial varieties thereof. In balance, however, it would be fair to characterise the semi-artificial standardised languages as fighting a losing battle against the more natural colloquial varieties. Suffice it to take note of prescriptive grammars of Malay or Indonesian railing against "incorrect" usages to realise that it is the colloquial varieties, with their more organic Mekong-Mamberamo profiles, that are on the ascendant.

Ironically, one domain where the standardised languages are if anything too successful is in effecting a bias amongst linguists with regard to what a typical language of the region looks like. Both Crouch (this volume) for Minangkabau and Conners (this volume) for Javanese show how the standardised versions of these languages provide a distorted impression of the grammatical profiles of the real colloquial varieties that underlie the standardisations, and as argued here, the same is true *a fortiori* also for Malay/Indonesian. Thus, as suggested in Gil (2015: 298–310), the notion of an "Indonesian-Type" clause structure reflects a bias towards the grammatical patterns of Standard Malay/Indonesian, which is then imposed,

often with little empirical justification, on analyses of other languages – see, for example, the discussion in Gil (2015: 308–309) of how this bias impacted on the description of Mentawai by Jufrizal and Arka (2006). Indeed, if it were not for the overbearing presence of Standard Malay/Indonesian and Javanese, the existence of Sundic-Type voice systems would probably have been recognised at a much earlier date, as would have the presence of an isolating crescent extending from the Mekong to the Mamberamo rivers.

As argued in this section, the simple grammatical structure and isolating profile that emerged from the coming together of the Austronesian and Mekong-Mamberamo heritages some 3500–4000 years ago resulted in simple grammatical structures and an isolating profile that persisted broadly, subject to the inevitable ups and downs and twists and turns, all the way to the present. Thus, the convergence of Austronesian and Mekong-Mamberamo heritages was the constituting event in the story of Riau Indonesian and its relatives; by comparison, everything else since then was smooth sailing.

The story of Riau Indonesian and its relatives can be usefully summarised in terms of McWhorter's (2007: 254) typology of language contact presented in Table 1 earlier, as follows:

Table 6. Riau Indonesian and its relatives in McWhorter's typology of language contact

	Lexicon only	Lexicon and syntax	Lexicon, syntax, morphology, phonology
No Simplification	Standard Malay Papuan Malay Siak Malay		Nonthaburi Malay Sri Lankan Malay
Moderate Simplification	Riau Indonesian		
Extreme Simplification			Pre-Proto-Malayic?

If there is a creole in the history of Riau Indonesian and its relatives, it is one that might perhaps have formed a few thousand years ago, at the original coming together of the Austronesian and Mekong-Mamberamo heritages. This is Pre-Proto-Malayic in the bottom right cell of Table 6. Its location in the bottom "creole" row reflects the radical contact-induced simplification that it would represent, relative to its Austronesian heritage. Its position in the rightmost column reflects the fact that, relative to its Austronesian heritage, the contact effects are evident in all domains of grammar – lexicon, syntax, morphology and phonology. In accordance with McWhorter's terminology, Pre-Proto-Malayic would thus be a deep or radical creole. There is just one problem with this: as argued in Section 5.2 above, radical simplification due to imperfect adult second-language acquisition is only one possible model for what happened at the convergence of Austronesian and

Mekong-Mamberamo, the two other available models being metatypy and relexification. In particular, from the alternative Mekong-Mamberamo perspective, there was no simplification at all; the simple grammatical profile was there all along. Hence the question mark beside Pre-Proto-Malayic in Table 6.

The remaining varieties of Malay/Indonesian in Table 6 are all contemporary varieties whose classifications reflect their recent histories, and descent from their immediate ancestors. In rather different ways, all of the varieties in the top row represent states of affairs in which no recent simplification has occurred – on the contrary, in some cases, complexification may have taken place. As argued above, Standard Malay represents a complexification of the colloquial varieties on which it is based. Similarly, as suggested in Section 4.1, the innovative argument-indexing system of Papuan Malay constitutes a complexification with respect to a presumed ancestral eastern Malay variety spoken a few hundred years previously in Maluku. As for Siak Malay, it, like many other traditional Malay dialects in Sumatra and Borneo, is argued in this chapter to represent the continuation of a grammatical profile that extends far back in time, and therefore exhibits no traces of recent simplification.

In contrast to these, Nonthaburi Malay, as argued by Tadmor (1995), has adopted several phonological and grammatical features from the surrounding Thai; examples include the introduction of aspirated stops, and the development of an innovative series of "pseudo affixes", as in *hɔʔ makʌŋ* (thing eat) 'food', replacing the earlier original affixes, as in *makɛnɛ* (eat:AUG) 'food' in its immediate ancestor Patani Malay. However, Nonthaburi Malay provides no evidence of simplification with respect to Patani Malay. As for Sri Lankan Malay, as described in Slomanson (2006, 2008, 2016), it exhibits influences from neighbouring languages, primarily Tamil, across all grammatical domains, ranging from a distinction between dental and retroflex consonants to an SOV word order typology with a substantial amount of verbal morphology including a distinction between finite and non-finite forms. Compared to its presumed Southeast Asian Malay ancestor, Sri Lankan Malay is not simplified; quite the contrary, in many respects it has undergone a significant amount of complexification.

Which brings us back to Riau Indonesian. As argued in detail in this chapter, Riau Indonesian does not have any recent history of radical contact-induced simplification due to imperfect adult second-language acquisition; it is therefore not a creole language. To the extent that it is somewhat simpler than neighbouring traditional dialects such as Siak Malay, as suggested in Gil (2001) and discussed further in Section 4.1 above, it may be tentatively assigned to the moderate simplification row in Table 6. However, with respect to its overall grammatical patterns, it is actually very similar to other traditional dialects; again, there is no evidence for any kind of recent restructuring. Thus, as suggested in Table 6, Riau Indonesian ends up as being an NCSL – the category that McWhorter originally introduced for

larger languages such as English and Standard Malay. Though even this classification is making the most out of the very minor differences between Riau Indonesian and traditional varieties such as Siak Malay; a more fine-grained representation might place Riau Indonesian somewhere in-between the no-simplification and moderate-simplification rows in Table 6.

McWhorter goes to some effort to emphasise the refutability of his characterisation of Riau Indonesian and its relatives as creoles, as in the following passage (2007: 236–7):

> We must assume, then, that both standard and PMD Malay are the results of grammatical simplification caused by incomplete acquisition. There are two discoveries that would refute this hypothesis. […] One would be [Indonesian-Type] languages which had undergone no significant non-native acquisition over their histories, which were nevertheless similarly reduced compared to Ross's early [Indonesian-Type] reconstruction and in other features such as those in living [Indonesian-Type] languages such as the ones of Sulawesi. Surely if the transition from Ross's early [Indonesian-Type] reconstruction to standard Malay is simply business as usual in the evolution of [Indonesian-Type] languages, then we would expect that not just one, but at least several groups of isolated [Indonesian-Type] speakers would speak languages similarly abbreviated compared to a Tukang Besi or a Batak variety. The same applies to varieties like Riau Indonesian. If this homology can result from uninterrupted transmission of a grammar, then we would expect at least a few other [Indonesian-Type] languages of similar grammatical profile. To my knowledge, at present there has been discovered not even a single one such variety.

But as demonstrated at length in this chapter, Riau Indonesian is that language. Moreover, as argued here and in other chapters in this volume, there are actually lots and lots of such languages with similar grammatical profiles: Minangkabau, Javanese, Lio, and numerous others – all with "no significant non-native acquisition over their histories" that can be demonstrated to have taken place any time in the course of the past three millennia. McWhorter continues:

> A second refutation would be several nonstandard Malay varieties spoken in regions with a long tradition of interethnic mixture that have ample overspecifications, structural elaborations, and/or irregularities absent in the standard. I refer to a hypothetical Malay in which, perhaps, there are obligatory subject-marking prefixes, a three-way distinction in demonstratives, a good dozen numeral classifiers in regular use, and imperative affixes. Varieties like this would vastly weaken the link I propose between non-native acquisition and the contrast between standard and colloquial Malay, and suggest that the difference between Adelaar and Prentice's "PMDs" and these radically conservative dialects was a matter of a mere roll of the dice.

Again, such varieties do exist. Kerinci, discussed in Section 4.3, albeit not included here within Malay proper, would definitely qualify as having "ample overspecifications". Crucially, although spoken in a mountain valley, Kerinci is actually host to at least as much language contact and inter-ethnic mixture as various lowland Sumatran varieties such as Siak Malay, Akit and Sakai, which do not exhibit similar overspecifications.

McWhorter is to be applauded for spelling out in detail how his characterisation of Riau Indonesian as a creole can be put to empirical test. But this chapter has put it to the test and found it wanting: Riau Indonesian is not a creole language.

6.   Conclusion

As argued in this chapter, Riau Indonesian and its relatives are the product of a dual heritage, Austronesian and Mekong-Mamberamo. While the lexicon is largely Austronesian, most of the grammatical patterns are Mekong-Mamberamo. The major constitutive event in the history of Riau Indonesian and its relatives was the coming together of these two heritages some 3500 to 4000 years ago.

The notion of a language having dual heritage is hardly new. Nichols and Warnow (2008: 762) talk of "dual parentage" in the case of creolisation and language mixture, as do Meakins, Green and Turpin (2018: 238) in the context of contact languages. The present proposal differs, however, from the above ones in that the dual ancestry is posited not for recent times but rather for the distant past: hence the choice of the term "heritage" rather than "parentage".

One might wonder: Why stop at two? Adopting a radical epidemiological approach as in Enfield (2003), specific linguistic features are what count and have histories, and obviously there are large numbers of them; languages, conceived of as clusters of such features, are merely epiphenomenal. A perhaps less extreme version of this position is hinted at by Donohue and Denham (this volume), who suggest that "a classification which truly reflects a language's social history should include information about the different sources of the different modules of that language".

There would seem, however, to be a general tendency for the totality of features to cluster into two large groups, lexical and grammatical. In part, this dichotomy may be due to the higher salience, in the consciousness of native speakers, of the lexicon as opposed to the grammar, resulting in a linguistic folk ontology in which people identify a language as the totality of its words, to the exclusion of all the other stuff. But whatever the reason, the typical situation is one in which the lexicon comes from one source and the grammar from another – as indeed is argued here to be the case for languages with Austronesian lexicon and Mekong-Mamberamo grammar.

Of course, things are rarely that neat. Thus, Michaelis (2017), drawing upon data from the *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013), shows that word-order patterns in creoles tend to resemble those of the lexifier language rather than the contact language that provides for most of the other grammatical features. Hopefully, an enhanced appreciation of the phenomenon of dual heritage, in the recent and distant past, will set the stage for the formulation of more detailed hypotheses concerning the possible ways in which languages may come together, and the nature of the potential outcomes.

Acknowledgements

References

Aboh, Enoch O. 2016. Creole distinctiveness: A dead end. *Journal of Pidgin and Creole Languages* 31(2): 400–418. https://doi.org/10.1075/jpcl.31.2.07abo

Adelaar, K. Alexander. 1984. Some Proto-Malayic affixes. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140(4): 402–421.

Adelaar, K. Alexander. 1992. *Proto Malayic: The Reconstruction of Its Phonology and Parts of Its Lexicon and Morphology*. Canberra: Pacific Linguistics.

Adelaar, K. Alexander. 1994. The classification of the Tamanic languages. In *Language Contact and Change in the Austronesian-Speaking World*, Tom E. Dutton & Darrell T. Tryon (eds), 1–42. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110883091.1

Adelaar, K. Alexander. 2005. *Salako or Badamea: Sketch Grammar, Text and Lexicon of a Kanayatn Dialect in West Borneo*. Wiesbaden: Harrassowitz.

Adelaar, K. Alexander & Prentice, David J. 1996. Malay: Its history, role and spread. In *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas* [Trends in Linguistics. Documentation 13], Stephen A. Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds), 673–693. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110819724.2.673

Anderbeck, Karl & Herdian, Aprilani. 2013. *The Improbable Language: Survey Report on the Nasal Language of Bengkulu, Sumatra* [SIL Electronic Survey Report 2013–012]. Dallas TX: SIL International.

Ansaldo, Umberto. 2004. Contact, typology and the speaker: The essentials of language. *Language Sciences* 26: 485–494. https://doi.org/10.1016/j.langsci.2003.11.004

Ansaldo, Umberto & Matthews, Stephen J. 1999. The Minnan substrate and creolisation in Baba Malay. *Journal of Chinese Linguistics* 27(1): 38–68.

Arnold, Laura. 2017. A Grammar of Ambel, An Austronesian Language of West New Guinea. PhD dissertation. University of Edinburgh.

Asmah, Haji Omar. 1981. *The Iban Language of Sarawak: A Grammatical Description*. Kuala Lumpur: Kementerian Pelajaran Malaysia.

Bakker, Peter, Daval-Markussen Aymeric, Parkvall, Mikael & Plag Ingo. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26: 5–42. https://doi.org/10.1075/jpcl.26.1.02bak

Benjamin, Geoffrey. 1993. Grammar and polity: The cultural and political background to Standard Malay. In *The Role of Theory in Language Description* [Trends in Linguistics, Studies and Monographs 69], William A. Foley (ed.), 341–392. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110872835.341

Berry, Keith & Berry, Christine. 1999. *A Description of Abun: A West Papuan Language of Irian Jaya*. Canberra: Pacific Linguistics.

Bickel, Balthasar & Nichols, Johanna. 2005. Fusion of selected inflectional formatives. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds), 86–89. Oxford: OUP.

Blasi, Damián E., Michaelis, Susanne M. & Haspelmath, Martin. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1: 723–729. https://doi.org/10.1038/s41562-017-0192-4

Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32(2): 241–293. https://doi.org/10.2307/3623195

Blust, Robert. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Blust, Robert & Trussel, Stephen. 2016. Austronesian comparative dictionary, web edn. <http://www.trussel2.com/acd> (8 December 2019).

Burenhult, Niclas. 2005. *A Grammar of Jahai*, Canberra: Pacific Linguistics.

Burgess, Anthony. 1975. *Language Made Plain*, revised edn. Glasgow: Fontana/Collins.

Bybee, Joan. 1997. Semantic aspects of morphological typology. In *Essays on Language Function and Language Type, Dedicated to T. Givón*, Joan L. Bybee, John Haiman & Sandra A. Thompson (eds), 25–37. Amsterdam: John Benjamins. https://doi.org/10.1075/z.82.05byb

Clynes, Adrian. 2001. Brunei Malay: An overview. *Occasional Papers in Language Studies* 7: 11–43. Department of English Language and Applied Linguistics, Universiti Brunei Darussalam.

Cole, Peter, Hermon, Gabriella & Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118: 1500–1553. https://doi.org/10.1016/j.lingua.2007.08.008

Collins, James T. 1980. *Ambonese Malay and Creolisation Theory* [Penerbitan Ilmiah 5]. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Conners, Thomas. 2008. Tengger Javanese. PhD dissertation, Yale University.

Conners, Thomas, Bowden, John & Gil, David. 2015. Valency classes in Jakarta Indonesian. In *Valency Classes in the World's Languages*, Andrej Malchukov & Bernard Comrie (eds), 941–986. Berlin: De Gruyter Mouton.

Dahl, Östen. 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, Östen. 2009. Testing the assumption of complexity invariance: The case of Elfdalian and Swedish. In *Language Complexity as an Evolving Variable*, Geoffrey Sampson, David Gil & Peter Trudgill (eds), 50–63. Oxford: OUP.

Dalby, Andrew. 1998. *Dictionary of Languages*. New York NY: Columbia University Press.

Daniel, Michael, Dobrushina, Nina & Knyazev, Sergey. 2010. Highlanders' Russian: Case study in bilingualism and language interference in central Daghestan. In *Sociolinguistic Approaches to Non-Standard Russian* [Slavica Helsingiensia 40], Arto Mustajoki, Ekaterin Protassova & Nikolai Vakhtin (eds), 65–93. Helsinki: Helsinki University.

Davies, William D. 2010. *A Grammar of Madurese*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110224443

de Vries, Lourens. 2004. *A Short Grammar of Inanwatan. An Endangered Language of the Bird's Head of Papua*. Canberra: Pacific Linguistics.

DeGraff, Michel. 2001. On the origin of Creoles: A Cartesian critique of 'neo'-Darwinian linguistics. *Linguistic Typology* 203: 213–310.

DeGraff, Michel. 2003. Against creole exceptionalism. *Language* 79(2): 391–410. https://doi.org/10.1353/lan.2003.0114

DeGraff, Michel. 2005. Linguists' most dangerous myth: The fallacy of Creole exceptionalism. *Language in Society* 34(4): 533–591. https://doi.org/10.1017/S0047404505050207

Denham, Tim & Donohue, Mark. 2009. Pre-Austronesian dispersal of banana cultivars West from New Guinea: Linguistic relics from Eastern Indonesia. *Archaeology in Oceania* 44: 18–28. https://doi.org/10.1002/j.1834-4453.2009.tb00041.x

Dobney, Keith, Cucchi, Thomas & Larson, Greger. 2008. The pigs of Island Southeast Asia and the Pacific: New evidence for taxonomic status and human-mediated dispersal. *Asian Perspectives* 47: 59–74. https://doi.org/10.1353/asi.2008.0009

Donohue, Mark & Grimes, Charles E. 2008. Yet more on the position of the languages of Eastern Indonesia and East Timor. *Oceanic Linguistics* 47(1): 114–158. https://doi.org/10.1353/ol.0.0008

Donohue, Mark & Sawaki, Yusuf. 2007. Papuan Malay pronominals: Forms and functions. *Oceanic Linguistics* 47: 253–276. https://doi.org/10.1353/ol.2007.0017

Durie, Mark. 1985. *A Grammar of Acehnese, on the Basis of a Dialect of North Aceh*. Dordrecht: Foris.

Dryer, Matthew S. 2005. Prefixing vs. suffixing in inflectional morphology. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie (eds), 110–113. Oxford: OUP.

Edwards, Owen. 2016. Metathesis and Unmetathesis: Parallelism and Complementarity in Amarasi, Timor. PhD dissertation, Australian National University.

Enfield, Nicholas J. 2003. *Linguistic Epidemiology: Semantics and Grammar of Language Contact in Mainland Southeast Asia*. London: Routledge Curzon.

Enfield, Nicholas J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34: 181–206. https://doi.org/10.1146/annurev.anthro.34.081804.120406

Enfield, Nicholas J. 2011. Linguistic diversity in Mainland Southeast Asia. In *Dynamics of Human Diversity: The case of Mainland Southeast Asia*, Nicholas J. Enfield (ed.), 63–79. Canberra: Pacific Linguistics.

van Engelenhoven, Aone. 2004. *Leti: A Language of Southwest Maluku*. Leiden: KITLV.

Ernanda, Ernanda. 2017. *Phrasal Alternation in Kerinci*. Amsterdam: LOT.

Fadlul Rahman, Fitri, Kurniati, Santi, Prima Putri, Yessy & Gil, David. 2013. Word-internal language mixing: Borrowing, code-switching or register-switching. Paper presented at the *Seventeenth International Symposium on Malay/Indonesian Linguistics*, Padang, Indonesia, 9 June 2019.

Gasser, Emily, Arnold, Laura & Kamholz, David. To appear. The languages of Halmahera and West New Guinea. In *Handbook of the Austronesian Languages of Southeast Asia* [tentative title], S. A. Adelaar & Antoinette Schapper (eds). Oxford: OUP.

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17: 179–200. https://doi.org/10.1017/S0332586500003000

Gil, David. 1996. How to speak backwards in Tagalog. In *Pan-Asiatic Linguistics, Proceedings of the Fourth International Symposium on Language and Linguistics, January 8–10, 1996*, Vol. 1, 297–306. Salaya: Mahidol University, Institute of Language and Culture for Rural Development.

Gil, David. 2000. Syntactic categories, cross-linguistic variation and Universal Grammar. In *Approaches to the Typology of Word Classes, Empirical Approaches to Language Typology*, Petra M. Vogel & Bernard Comrie (eds), 173–216. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110806120.173

Gil, David. 2001. Creoles, complexity and Riau Indonesian. *Linguistic Typology* 5: 325–371.

Gil, David. 2002. The prefixes *di-* and *N-* in Malay / Indonesian dialects. In *The History and Typology of Western Austronesian Voice*, Fay Wouk & Malcolm Ross (eds), 241–283. Canberra: Pacific Linguistics.

Gil, David. 2005. Word order without syntactic categories: How Riau Indonesian does it. In *Verb First: On the Syntax of Verb-Initial Languages Systems* [Linguistik Aktuell/Linguistics Today 73], Andrew Carnie, Heidi Harley & Sheila A. Dooley (eds), 243–263. Amsterdam: John Benjamins. https://doi.org/10.1075/la.73.17gil

Gil, David. 2006. Intonation and thematic roles in Riau Indonesian. In *Topic and Focus, Cross-Linguistic Perspectives on Meaning and Intonation*, Chungmin Lee, Matthew Gordon & Daniel Büring (eds), 41–68. Dordrecht: Springer.

Gil, David. 2009a. How much grammar does it take to sail a boat? In *Language Complexity as an Evolving Variable*, Geoffrey Sampson, David Gil & Peter Trudgill (eds), 19–33. Oxford: OUP.

Gil, David. 2009b. Riau Indonesian: What kind of a language is it? *Kongres Internasional Masyarakat Linguistik Indonesia (KIMLI) 2009* (Indonesia Linguistics Society International Conference), 29–61. Malang: Masyarakat Linguistik Indonesia and Universitas Negeri Malang.

Gil, David. 2013. Riau Indonesian: A language without nouns and verbs. In *Flexible Word Classes, Typological Studies of Underspecified Parts of Speech*, Jan Rijkhoff & Eva van Lier (eds), 89–130. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199668441.003.0004

Gil, David. 2015. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia, The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: De Gruyter Mouton.

Gil, David. 2016. Describing languoids: When incommensurability meets the language-dialect continuum. *Linguistic Typology* 20: 439–462. https://doi.org/10.1515/lingty-2016-0017

Gil, David. 2017a. Isolating-monocategorial-associational language. In *Handbook of Categorization in Cognitive Science*, 2nd edn, Henri Cohen & Claire Lefebvre (eds), 471–510. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-08-101107-2.00020-8

Gil, David. 2017b. Roon ve, DO/GIVE coexpression, and language contact in Northwest New Guinea. In *Contact and Substrate in the Languages of Wallacea*, Part 1, Antionette Schapper (ed.). *NUSA* 62: 41–100.

Gil, David & McKinnon, Timothy. 2014. Excrescent nasals in Western Sumatra. Paper presented at the *24th Annual Meeting of the Southeast Asian Linguistics Society*, Yangon University, Yangon, Myanmar, 30 May 2014.

Gil, David & McKinnon, Timothy. 2015. Phrasal phonological alternations in Malayic languages. Paper presented at the *11th Meeting of the Association for Linguistic Typology*, Albuquerque, NM, 1 August 2015.

Gil, David & Shen, Yeshayahu. 2019. How grammar introduces asymmetry into cognitive structures: Compositional semantics, metaphors and schematological hybrids. *Frontiers in Psychology – Language Sciences*. https://doi.org/10.3389/fpsyg.2019.02275

Gil, David, Tadmor, Uri, Bowden, John & Taylor, Bradley. 2015. *Data from the Jakarta Field Station, Department of Linguistics, Max Planck Institute for Evolutionary Anthropology*, 1999–2015. <https://jakarta.shh.mpg.de/data.php> (30 May 2020).

Goldammer, Thomas. 2011. Areal Patterns in Ancient Western Eurasia. PhD dissertation, University of Leipzig.

Gudai, Darmansyah H. 1985. A Grammar of Maanyan, A Language of Central Kalimantan, PhD dissertation, Australian National University.

Hanan, Sandra Safitri. 2015. *Berkenalan dengan dialek Melayu Makassar* (Getting to Know the Makassar Malay Dialect). Yogyakarta: Ombak.

Haspelmath, Martin. 2013. Argument indexing: A conceptual framework for the syntactic status of bound person forms. In *Languages Across Boundaries: Studies in Memory of Anna Siewierska*, Dik Bakker & Martin Haspelmath (eds), 197–226. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110331127.197

Haspelmath, Martin, Dryer, Matthew, Gil, David & Comrie, Bernard (eds). 2005. *The World Atlas of Language Structures*. Oxford: OUP.

Hayami-Allen, Rika. 2001. A Descriptive Study of the Language of Ternate, The Northern Moluccas, Indonesia. PhD dissertation, University of Pittsburgh.

Hendon, Rufus S. 1966. *The Phonology and Morphology of Ulu Muar Malay (Kuala Pilah District, Negri Sembilan, Malaya)*. New Haven CT: Yale University, Dept. of Anthroplogy.

Hogan, David W. 1988. *Urak Lawoiʾ: Basic Structures and a Dictionary*. Canberra: Australian National University.

Ikranagara, Kay. 1980. *Melayu Betawi Grammar* [NUSA Linguistic Studies in Indonesian and Languages in Indonesia 9]. Jakarta: Universitas Atma Jaya.

Jacobs, Guy S., Hudjashov, Georgi, Lauri, Saag, Kusuma, Pradiptajati, Darusallam, Chelzie C., Lawson, Daniel J., Mondal, Mayukh, Pagani, Luca, Ricaut, François-Xavier, Stoneking, Mark, Metspalu, Mait, Sudoyo, Herawati, Lansing, J. Stephen & Cox, Murray P. 2019. Multiple deeply divergent Denisovan ancestries in Papuans. *Cell* 177(4): 1010–1021. https://doi.org/10.1016/j.cell.2019.02.035

Jinam, Timothy A., Hong, Lih-Chun, Phipps, Maude E., Stoneking, Mark, Ameen, Mahmood, Edo, Juli, HUGO Pan-Asian SNP Consortium; Naruya Saitou. 2012. Evolutionary history of continental Southeast Asians: Early train hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Molecular Biology and Evolution* 29: 3513–3527. https://doi.org/10.1093/molbev/mss169

Jufrizal & Arka, I. Wayan. 2006. Voice in Mentawai: A note on its typological position in the Austronesian languages of Indonesia. Paper presented at the *16th Annual Meeting of the Southeast Asian Linguistics Society*, Jakarta, Indonesia, 21 September 2006.

Kahler, Hans. 1966. *Wörterverzeichnis des Omong Djakarta*. Berlin: Dietrich Reimer.

Kalipke, Hans & Agar Kalipke, Mohamad. 2001. *Wörterbuch Sakai Indonesisch-Deutsch*. Hamburg: Helmut Buske.

Karafet, Tatiana M., Hallmark, Brian, Cox, Murray P., Sudoyo, Herawati, Downey, Sean, Lansing, J. Stephen & Hammer, Michael F. 2010. Major east-west division underlies Y chromosome stratification across Indonesia. *Molecular Biology and Evolution* 27: 1833–1844. https://doi.org/10.1093/molbev/msq063

Kennedy, Jean. 2008. Pacific bananas: Complex origins, multiple dispersals? *Asian Perspectives* 47: 75–94. https://doi.org/10.1353/asi.2008.0004

Kluge, Angela. 2014. *A Grammar of Papuan Malay*. Utrecht: LOT.

Kluge, Angela. 2016. Productivity of morphological patterns and social domain analysis in Papuan Malay. *Journal of the Southeast Asian Linguistics Society* 9: 200–218.

Kruspe, Nicole. 2004. *A Grammar of Semelai*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511550713

LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds), 225–254. Oxford: OUP.

Larson, Greger, Cucchi, Thomas, Fujita, Masakatsu, Matisoo-Smith, Elizabeth, Robins, Judith, Anderson, Atholl, Rolett, Barry, Spriggs, Matthew, Dolman, Gayner, Kim, Tae-Hun, Nguyen Thi Dieu Thuy, Randi, Ettore, Doherty, Moira, Due, Rokua Awe, Bollt, Robert, Djubiantono, Tony, Griffin, Bion, Intoh, Michiko, Keane, Emile, Kirch, Patrick, Li, Kuang-Ti, Morwood, Michael, Pedriña, Lolita M., Piper, Philip J., Rabett, Ryan J., Shooter, Peter, Van den Bergh, Gert, West, Eric, Wickler, Stephen, Yuan, Jing, Cooper, Alan & Dobney, Keith. 2007. Phylogeny and ancient DNA of Sus provides insights into Neolithic expansion in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Sciences (Washington DC)* 104: 4834–4839. https://doi.org/10.1073/pnas.0607753104

Lauder, Alan. 2008. Review of 'Jean Aitchison, Words in the Mind (3rd edition)'. *Linguistik Indonesia* 26(1): 115–127.

Lefebvre, Claire. 1986. Relexification in creole genesis revisited: The case of Haitian Creole. In *Substrata Versus Universals in Creole Genesis* [Creole Language Library 1], Pieter Muysken & Norval S. H. Smith (eds), 279–300. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.1.13lef

Lefebvre, Claire. 1997. On the cognitive process of relexification. In *Relexification in Creole and Non-Creole Languages. (With Special Reference to Modern Hebrew, Haitian Creole, Romani and Rumanian)*, Julia Horvath & Paul Wexler (eds), 72–99. Wiesbaden: Otto Harrassowitz.

Lim, Sonny. 1988. Baba Malay: The language of the 'Straits-born' Chinese. In *Papers in Western Austronesian Linguistics* No.3., Hein Steinhauer (ed.), 1–61. Canberra: Pacific Linguistics.

McDonnell, Bradley. 2016. Symmetrical Voice Constructions in Besemah: A Usage-based Approach. PhD dissertation, University of California, Santa Barbara.

McGinn, Richard. 1982. *Outline of Rejang Syntax* [NUSA 14]. Jakarta: Universitas Atma Jaya.

McKinnon, Timothy. 2011. The Morphophonology and Morphosyntax of Kerinci Word-Shape Alternations. PhD dissertation, University of Delaware, Newark.

McKinnon, Timothy, Cole, Peter & Hermon, Gabriella. 2011. Object agreement and 'pro-drop' in Kerinci Malay. *Language* 87(4): 715–750. https://doi.org/10.1353/lan.2011.0092

McKinnon, Timothy & Gil, David. 2014. Codas with vowels in Western Sumatra. Paper presented at the *24th Annual Meeting of the Southeast Asian Linguistics Society*, Yangon University, Yangon, Myanmar, 30 May 2014.

McKinnon, Timothy, Hermon, Gabriella, Yanti & Cole Peter. 2018. From phonology to syntax: Insights from Malay. In *Boundaries Crossed, at the Interfaces of Morphosyntax, Phonology, Pragmatics and Semantics*, Huba Bartos, Marcel den Dikken, Zoltán Bánreti & Tamás Varadi (eds), 349–371. Berlin: Springer. https://doi.org/10.1007/978-3-319-90710-9_22

McKinnon, Timothy, Yanti, Cole Peter & Hermon, Gabriella. 2015. Infixation and apophony in Malay: Description and developmental stages. *Linguistik Indonesia* 33(1): 1–19.

McWhorter, John H. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74: 788–818. https://doi.org/10.2307/417003

McWhorter, John H. 2000. Defining creole as a synchronic term. In *Degrees of Restructuring in Creole Languages* [Creole Language Library 22], Ingrid Neumann-Holzschuh & Edgar Schneider (eds), 85–123. Amsterdam: John Benjamins.

McWhorter, John H. 2001a. The rest of the story: Restoring pidginisation to creole genesis theory. *Journal of Pidgin and Creole Languages* 17: 1–48. https://doi.org/10.1075/jpcl.17.1.02mcw

McWhorter, John H. 2001b. The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–166. https://doi.org/10.1515/lity.2001.001

McWhorter, John H. 2001c. What people ask David Gil and why? *Linguistic Typology* 5: 388–412. https://doi.org/10.1515/lity.2001.003

McWhorter, John H. 2005. *Defining Creole*. Oxford: OUP.

McWhorter, John H. 2007. *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780195309805.001.0001

McWhorter, John H. 2008. Why does a language undress: Strange cases in Indonesia. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94] Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 167–190. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.12mcw

McWhorter, John H. 2011a. *Linguistic Simplicity and Complexity. Why Do Languages Undress?* Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781934078402

McWhorter, John H. 2011b. Tying up loose ends: The creole prototype after all. *Diachronica* 28(1): 82–117. https://doi.org/10.1075/dia.28.1.04mcw

McWhorter, John H. 2018. *The Creole Debate*. Cambridge: CUP. https://doi.org/10.1017/9781108553308

McWhorter, John H. 2019. *On counterexamples to the creole prototype: Why Riau Indonesian is a creole and Chinese isn't*. Ms.

Matisoff, James A. 2003. Aslian: Mon-Khmer of the Malay Peninsula. *Mon-Khmer Studies* 33: 1–58.

Mead, David. 2003. Evidence for a Celebic subgroup. In *Issues in Austronesian Historical Phonology*, John Lynch (ed.), 115–141. Canberra: Pacific Linguistics.

Meakins, Felicity, Green, Jennifer & Turpin, Myfany. 2018. *Understanding Linguistic Fieldwork*. New York NY: Routledge. https://doi.org/10.4324/9780203701294

Michaelis, Susanne. 2017. Explaining the features of creole languages: Language shift and the division of labour between substrates and lexifiers. Paper presented at the *Workshop on Language Shift and Substratum Influence in (Pre)history*, Max Planck Institute for the Science of Human History, Jena, 12 July 2017.

Michaelis, Susanna, Maurer, Philippe, Haspelmath, Martin & Huber, Magnus (eds). 2013. *The Atlas of Pidgin and Creole Language Structures (APiCS)*. Oxford: OUP.

Miestamo, Matti. 2008. *Standard Negation, The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin: Mouton de Gruyter.

Morwood, Michael J., Soejono, Radien P., Roberts, Richard G., Sutikna, Thomas, Turney, Chris S. M., Westaway, Kira E., Rink, W. Jack, Zhao, Jian-xin, van den Bergh, Gert, Rokus Awe Due, Hobbs, D. R., Moore, Mark W., Bird, Michael I. & Fifield, L. K. 2004. Archaeology and age of a new homonin from Flores in Eastern Indonesia. *Nature* 431(7012): 1087–1091. https://doi.org/10.1038/nature02956

Moussay, Gerard. 1981. *La langue Minangkabau*. Paris: Archipel.

Mufwene, Salikoko S. 1986. Les langues créoles peuvent-elles ĕtre définies sans allusion à leur histoire? *Études Créoles* 9: 135–50.

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

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

Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In *Language Complexity as an Evolving Variable*, Geoffrey Sampson, David Gil & Peter Trudgill (eds), 110–125. Oxford: OUP.

Nichols, Johanna & Warnow, Tandy. 2008. Tutorial in computational linguistic phylogeny. *Language and Linguistics Compass* 2(5): 760–820. https://doi.org/10.1111/j.1749-818X.2008.00082.x

Nordhoff, Sebastian (ed.). 2012. *The Genesis of Sri Lanka Malay, A Case of Extreme Language Contact*. Leiden: Brill.

Nothofer, Bernd. 1986. The Barrier Island languages in the Austronesian language family. In *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, Paul Geraghty, Lois Carrington & Stephen A. Wurm (eds), 87–109. Canberra: Pacific Linguistics.

Nothofer, Bernd. 1994. The relationships between the languages of the Barrier Islands and the Sulawesi-Philippine languages. In *Language Contact and Change in the Austronesian World*, Tom E. Dutton & Darrell T. Tryon (eds), 389–409. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110883091.389

Odé, Cecilia. 2002. A sketch of Mpur. In *Languages of the Eastern Bird's Head*, Ger P. Reesink (ed.), 45–107. Canberra: Pacific Linguistics.

Paauw, Scott. 2009. The Malay Contact Varieties of Eastern Indonesia: A Typological Comparison. PhD dissertation. State University of New York at Buffalo.

Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94] Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 265–285. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.17par

Platt, John T. 1975. The Singapore English speech continuum and its basilect 'Singlish' as a 'Creoloid'. *Anthropological Linguistics* 17: 363–374.

Post, Mark W. 2015. Morphosyntactic reconstruction in an areal-historical context: A pre-historical relationship between North East India and Mainland Southwest Asia? In *Languages of Mainland Southeast Asia, The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 209–265. Berlin: De Gruyter Mouton.

Prentice, David J. 1987. Malay (Indonesian and Malaysian). In *The World's Major Languages*, Bernard Comrie (eds), 913–935. Routledge: London.

Prentice, David J. & Usman, A. Hakim. 1978. Kerinci sound-changes and phonotactics. In *Second International Conference on Austronesian Linguistics, Fascicle I, Western Austronesian*, Stephen A. Wurm & Lois Carrington (eds), 121–163. Canberra: Pacific Linguistics.

Raviv, Limor, Meyer, Antje & Lev-Ari, Shiri. 2019. Larger communities create more systematic languages. *Proceedings of the Royal Society B* 286: 20191262.
https://doi.org/10.1098/rspb.2019.1262

Reesink, Ger P. 1998. The Bird's Head as sprachbund. In *Perspectives on the Bird's Head of Irian Jaya*, Jelle Miedama, Cecilia Odé & Rien A. C. Dam (eds), 603–642. Amsterdam: Rodopi.

Reesink, Ger P. 2002. The Eastern Bird's Head languages compared. In *Languages of the Eastern Bird's Head*, Ger P. Reesink (ed.), 1–44. Canberra: Pacific Linguistics.

Reesink, Ger P. 2009. A connection between Bird's Head and (Proto) Oceanic. In *Discovering History through Language, Papers in Honour of Malcolm Ross*, Bethwyn Evans (ed.), 181–192. Canberra: Pacific Linguistics.

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

Ross, Malcolm D. 2002. Final words: Research themes in the history and typology of Western Austronesian languages. In *The History and Typology of Western Austronesian Voice Systems*, Fay Wouk & Malcolm Ross (eds), 451–474. Canberra: Pacific Linguistics.

Schapper, Antoinette. 2010. Bunaq: A Papuan Language of Central Timor. PhD dissertation. Australian National University.

Seifart, Frank. 2013. *AfBo: A World-Wide Survey of Affix Borrowing*. Max Planck Institute for Evolutionary Anthropology, Leipzig. <http://afbo.info> (15 November 2019).

Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91(3): 511–532.
https://doi.org/10.1353/lan.2015.0044

Sinnemäki, Kaius. 2009. Complexity in core argument marking and population size. In *Language Complexity as an Evolving Variable*, Geoffrey Sampson, David Gil & Peter Trudgill (eds), 126–141. Oxford: OUP.

Slomanson, Peter. 2006. Sri Lankan Malay morphosyntax, Lankan or Malay? In *Structure and Variation in Language Contact* [Creole Language Library 29], Ana Deumert & Stephany Durrleman-Tame (eds), 135–158. Amsterdam: John Benjamins. https://doi.org/10.1075/cll.29.08slo

Slomanson, Peter. 2008. The perfect construction and complexity drift in Sri Lankan Malay. *Lingua* 118: 1640–1655. https://doi.org/10.1016/j.lingua.2007.08.009

Slomanson, Peter. 2016. Pragmatic accommodation as a catalyst for the development of (non-)finiteness. *The Linguistic Review* 33(3): 365–396. https://doi.org/10.1515/tlr-2016-0002

Sneddon, James Neil. 1996. *Indonesian, A Comprehensive Grammar*. London: Routledge.

Sneddon, James Neil. 2006. *Colloquial Jakartan Indonesian*. Canberra: Pacific Linguistics.

Spriggs, Matthew. 2007. The Neolithic and Austronesian expansion within Island Southeast Asia and into the Pacific. In *From Southeast Asia to the Pacific: Archaeological Perspectives on the Austronesian Expansion and the Lapita Cultural Complex*, Scarlett Chiu & Christophe Sand (eds), 104–125. Taipei: Centres for Archaeological Studies Research and for Humanities and Social Sciences, Academica Sinica.

Starosta, Stanley, Pawley, Andrew K. & Reid, Laurie. 1982. The evolution of focus in Austronesian. In *Papers from TICAL*, Vol. 2, Amran Halim, Lois Carrington & Stephen A. Wurm (eds), 201–216. Canberra: Pacific Linguistics.

Steinhauer, Hein. 1991. Malay in east Indonesia: The case of Larantuka (Flores). In *Papers in Austronesian Linguistics*, Vol. 1, Hein Steinhauer (ed.), 177–195. Canberra: Pacific Linguistics.

Steinhauer, Hein. 1996. Morphemic metathesis in Dawanese (Timor). In Hein Steinhauer (ed.), *Papers in Austronesian Linguistics*, 217–232. Canberra: Pacific Linguistics.

Steinhauer, Hein & Usman, A. Hakim. 1978. Notes on the morphemics of Kerinci (Sumatra). In *Second International Conference on Austronesian Linguistics, Fascicle I, Western Austronesian*, Stephen A. Wurm & Lois Carrington (eds), 483–502. Canberra: Pacific Linguistics.

Stokhof, Wim A. L. 1975. *Preliminary Notes on the Alor and Pantar Languages (East Indonesia)*. Canberra: Pacific Linguistics.

Sutikna, Thomas, Tocheri, Matthew W., Morwood, Michael J., Saptomo, E. Wahyu, Jatmiko, Rokus Due Awe, Wasisto, Sri, Westaway, Kira E., Aubert, Maxime, Li, Bo, Zhao, Jian-xin, Storey, Michael, Alloway, Brent V., Morley, Mike W., Meijer, Hanneke J. M., van den Bergh, Gerrit D., Grün, Rainer, Dosseto, Anthony, Brumm, Adam, Jungers, William L. & Roberts, Richard G. 2016. Revised stratigraphy and chronology for *Homo floresiensis* at Liang Bua in Indonesia. *Nature* 532: 366–369. https://doi.org/10.1038/nature17179

Tadmor, Uri. 1995. Language Contact and Systemic Restructuring: The Malay Dialect of Nonthaburi, Central Thailand. PhD dissertation, University of Hawai'i at Manoa.

Tjia, Johnny. 2007. *A Grammar of Mualang, An Ibanic Language of Western Kalimantan*, Indonesia. Utrecht: LOT.

Trudgill, Peter. 2009. Sociolinguistic typology and complexification. In *Language Complexity as an Evolving Variable*, Geoffrey Sampson, David Gil & Peter Trudgill (eds), 98–109. Oxford: OUP.

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

Vallee, François, Luciani, Aurélien & Cox, Murray P. 2016. Reconstructing demography and social behavior during the Neolithic expansion from genomic diversity across Island Southeast Asia. *Genetics* 204: 1495–1506. https://doi.org/10.1534/genetics.116.191379

van den Berg, René. 2004. Some notes on the origin of Malay *di-*. *Bijdragen tot de Taal-, Land- en Volkenkunde (Journal of the Humanities and Social Sciences of Southeast Asia)* 60(4): 532–554.

van Hasselt, Frans J. F. 1905. *Spraakkunst der Numfoorsche Taal* (The art of speech in the Numfoor language). The Hague: Martinus Nijhoff. https://doi.org/10.1007/978-94-015-0872-8

Voorhoeve, Clemens L. 1982. The West Makian language, North Moluccas, Indonesia: A fieldwork report. In *The Makian Languages and Their Neighbors*, Clemens L. Voorhoeve (ed.). Canberra: Pacific Linguistics.

Wexler, Paul. 2002. *Two-Tiered Relexification in Yiddish: Jews, Sorbs, Khazars and the Kiev-Polessian Dialect*. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110898736

Wolff, John U. 1988. The contribution of Banjar Masin Malay to the reconstruction of Proto-Malay. In *Rekonstruksi dan Cabang-Cabang Bahasa Melayu Induk* (The Proto-Malay Language: Reconstruction and Branches), Mohd Thani Ahmad & Zaini Mohamed Zain (eds), 85–98. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Wurm, Stephen A., Voorhoeve, Clemens L. & McElhanon, Kenneth A. 1975. The Trans-New Guinea phylum in general. In *New Guinea Area Languages and Language Study*, Stephen A. Wurm (ed.), 299–322. Canberra: Pacific Linguistics.

Yanti. 2010. A Reference Grammar of Jambi Malay. PhD dissertation, University of Delaware, Newark.

CHAPTER 4

Voice and bare verbs in Colloquial Minangkabau

Sophie Crouch

The University of Western Australia

Minangkabau is an Austronesian language spoken primarily in West Sumatra. Previous studies of voice and morphosyntax, which have largely relied on elicitation-based methodology, suggest that Minangkabau can be characterised as an Indonesian-type language since its active/passive voice system resembles that of Malay/Indonesian. This study, which makes use of a corpus of naturalistic Minangkabau data, finds that the use of bare verbs (i.e. verbs that are unmarked for voice) is pervasive in informal and conversational contexts. Morphological underspecification for voice in the naturalistic data suggests that Colloquial Minangkabau is a distinct variety. The apparent optional nature of voice marking in Colloquial Minangkabau indicates that its function is primarily semantic and conceptual, and that Colloquial Minangkabau is better characterised as having a Sundic-type voice system.

Keywords: Minangkabau, voice, bare verbs, Sundic-type voice system, Indonesian-type voice system

1. Introduction

Minangkabau is an Austronesian language with approximately 7 million speakers (Gordon 2005). It is primarily spoken around Padang and in the highlands of West Sumatra and also throughout the Indonesian Archipelago and the Malay Peninsula, due to the Minangkabau *marantau* tradition of migration (Drakard 1999). The Minangkabau homeland borders on areas where, among others, the Batak languages, Kerinci and Riau Indonesian are spoken (Moussay 1998).

Previous studies have shown that, typologically speaking, Minangkabau can be characterised as an Indonesian-type language since its voice system resembles that of Malay/Indonesian. However, this study aims to show that this is true only of Standard Minangkabau. In fact, Colloquial Minangkabau shows a unique character and is therefore better considered a 'Sundic-type' language (Gil 2015a). Bare verbs (i.e. verbs that are not marked for voice) are a prevalent feature of Colloquial Minangkabau and I argue that the presence of these

forms demonstrates that morphological underspecification for voice is grammatically acceptable in Colloquial Minangkabau. Colloquial Minangkabau thus fits Gil's (2001, 2007) criteria for being an 'Associational' Language and its voice system has more in common with isolating Austronesian languages than typical Indonesian-type languages.

This chapter explores the idea that Colloquial Minangkabau leans towards the isolating end of the spectrum of Austronesian languages. The chapter first clarifies the differences in form and function between Standard and Colloquial Minangkabau. Section 3 describes the pragmatic basis for voice in Standard Minangkabau and Section 4 then describes the use of bare verbs and the ambiguity inherent in the voice system of Colloquial Minangkabau.

2.   Standard Minangkabau and Colloquial Minangkabau

2.1   A focus on the standard

Despite its large number of speakers and the spread of Minangkabau people throughout the Indonesian Archipelago, Minangkabau remains relatively under-described compared to other Indonesian languages. Some descriptive work on the language was carried out by Dutch scholars in the 19th and early 20th centuries. This early work mainly focuses on Minangkabau's Arabic-based orthography and the creation and implementation of a Romanised orthography (for further discussion see Voorhoeve 1955). Also a number of Minangkabau word lists, dictionaries, collections of folk tales and a grammatical sketch of the language were published during this period (cf. van der Toorn 1899). However, the scholarship of these early Dutch works has long been surpassed by studies grounded in modern descriptive linguistics.

More recently published works on the language range in scope from semantics (cf. Adnani 1971; Anwar 1992; Puspawati 1997; Ramadian 1992), to phonology (cf. Adelaar 1992a; Williams 1961), morphology (cf. Ansjar 1971; Be 1978/1979, 1984), syntax (cf. Arifin et al, 1979; Brodkin & Fortin 2017; Fortin 2001, 2004), and pragmatics (cf. Genta 1999; Tanner 1972). The historical-genetic features of the language are discussed by Adelaar (1995) and a good dialect survey of the language is presented by Medan (n.d.). More recent dialect surveys with some diachronic analysis can be found in Nadra (2006) and Nadra et al. (2008). In addition, Moussay has published a detailed descriptive grammar of Minangkabau (cf. Moussay 1998) as well as a three-way Minangkabau-Indonesian-French dictionary (Moussay 1995).

A number of these studies (cf. Brodkin & Fortin 2017; Fortin 2001, 2004; Moussay 1998; Williams 1961) also examine Minangkabau voice and morphosyntax. However, a shortcoming of these works is that they rely primarily on elicited

data and/or examples of Standard Minangkabau for their analysis. As this study demonstrates, using naturalistic, informal and conversational Minangkabau data reveals a different and more complete picture of the nature of voice in Minangkabau than is presented in previous studies. Specifically, whereas previous studies have leaned towards classifying Minangkabau as an Indonesian-type language, new data and analysis shows that this is true only of Standard Minangkabau and not of Colloquial Minangkabau.

2.2   A new focus: Naturalistic data

The data for this study originate primarily from an electronic database of naturalistic Minangkabau, the MPI EVA Minangkabau corpus (Gil 2015b), which was developed by researchers at Universitas Bung Hatta in Indonesia. The corpus is one of a number of corpora of Indonesian languages maintained as part of the MPI EVA Jakarta Field Station Project. The project aims to collect naturalistic data from a range of Indonesian minority languages across the archipelago to enable language description and investigate language acquisition and language contact in the region.

At the time of research, the MPI EVA Minangkabau corpus contained over sixty-thousand utterances, including data from a wide variety of discourse types including personal narratives and personal histories, traditional folk tales, gossip, conversations and interviews. In order to give an accurate cross section of Minangkabau society and a profile of a language with some degree of internal diversity, the data was collected in the urban region of Padang as well as across the West Sumatra region from speakers who vary in age, gender, regional background and social status.[1]

For this study, elicitation was also used in combination with the naturalistic data. However, an electronic corpus like the MPI EVA Minangkabau corpus counteracts the methodological shortcomings of purely elicitation-based analyses of language which often rely on intuition, introspection and potentially biased native speaker judgements. Arguments for the existence of certain constructions, or arguments about the motivation for certain constructions, can instead be made on an empirical basis (Biber 2000; Bybee & Hopper 2001; Keller 1999; Schütze 1996).

---

1.   Minangkabau data has been transcribed according to the Standard Minangkabau orthography as established at the *Seminar Bahasa Minang* in 1976. The orthography and its phonetic correspondences are listed below. Please note that in accordance with orthographic conventions, word-final glottal stops are transcribed as 'k' and word-internal glottal stops (i.e. between two vowels) are not transcribed. Vowels: a – [a], e – [ɛ], i – [i], o – [ɔ], u – [u]; Diphthongs: ai – [ai], au – [au], ia – [ia], ua – [ua], ui – [ui]; Consonants: b – [b], c – [ʧ], d – [d], g – [g], h – [h], j – [ʤ], k – [k]/[ʔ], l – [l], m – [m], n – [n], ng – [ŋ], ny – [ɲ], p – [p], r – [r], s – [s], t – [t], w – [w], y – [j].

Another benefit of a corpus of this kind is the type of data it contains. The variety of discourse types in the corpus, as well as the large volume of naturalistic data, enabled comparisons to be made against the less spontaneous and more formal Minangkabau data that I had seen in previous studies of the language, and that I had elicited myself. One of the most striking differences between these two data sets, i.e. the elicited data as opposed to the naturalistic data, is the high frequency of bare verbs found in the naturalistic data (see Section 4). Naturalistic data, especially conversational data, is particularly important to consider for linguistic analysis since it reveals a range of linguistic structures not found in elicited data (cf. Biber, Conrad & Reppen 1998; Biber 2000; Francis 1993; Mosel 2006; Nerbonne 1998; Thompson & Hopper 2001: 41; Wouk 1999). In fact, the discrepancies found between the naturalistic and elicited Minangkabau data form the basis for categorising Standard and Colloquial Minangkabau as two distinct varieties.

2.3 Standard and Colloquial Minangkabau:
Formal and functional differences

There are around a dozen dialects of Minangkabau (Gordon 2005; Medan n.d.) and there is evidence that a standard form of the language based on the Padang dialect, i.e. Standard Minangkabau, also exists (Adnani 1971: 4; Moussay 1998). Standard Minangkabau is the prestige form and is used in formal contexts and in the written medium. Standard Minangkabau is similar to Standard Indonesian in terms of its syntactic structures and voice system and may also be influenced by the prescriptivist rules that inform the use of Standard Indonesian. Colloquial Minangkabau, on the other hand, is used in informal contexts and can be characterised by its freer word order and its varying degrees of morphological complexity, including the use of bare verbs. It is important to note that Colloquial Minangkabau is not a single dialect but exists as a continuum of varieties used over the geographically large and linguistically diverse area of West Sumatra. Colloquial Minangkabau also is not unique to one group of people, as all speakers of Minangkabau command a colloquial variety as well as other varieties. It exhibits a great degree of internal diversity, including morphological variation (Noviatri et al. 2017). Standard and Colloquial Minangkabau are also not mutually exclusive and speakers may switch between the two varieties within the same interaction. However, for the purposes of this paper, I use 'Colloquial Minangkabau' as a catch-all term to refer to data in the MPI EVA Minangkabau corpus that represents non-standard forms of the language, as defined by Crouch (2009).

The distinction between the two varieties is recognised by speakers of Minangkabau primarily as a difference in register: Standard Minangkabau is used in formal and ritual contexts and Colloquial Minangkabau is used in more familiar

and intimate situations. My consultants also agree that Standard Minangkabau is found in formal written media, for example in newspapers and magazines (cf. Moussay 1998), whereas Colloquial Minangkabau is primarily a spoken variety and is subject to regional variation. It is not surprising then that the naturalistic data of the MPI EVA Minangkabau corpus reveals more examples of Colloquial Minangkabau, whereas much of my and other scholars' elicited data reflects Standard Minangkabau.

There are also three main formal differences between the two varieties that lend further support to making the distinction. First, borrowings of Indonesian morphology and word-internal code-mixing with Indonesian can be found in Standard Minangkabau (Fadlul et al. 2013). For instance, speakers of Standard Minangkabau often use the Standard Indonesian *kan* applicative in place of the Minangkabau applicative *an* (see Example (1)).

(1) Standard Minangkabau
a. *Ambo manggaramkan ayia lawik.*
    Ambo maN-garam-kan ayia   lawik
    1SG   TR-salt-APPL   water sea
    'I make the sea water become salt.'   (adapted from Fortin 2001: 44)[2]

Colloquial Minangkabau
b. *Aden manggaraman lauak.*
    aden   maN-garam-an lauak
    1SG   AV-salt-APPL   fish
    'I salted the fish.'   (Elicitation)

Second, morphological underspecification for voice is rife in Colloquial Minangkabau. Verbs which one would expect to be marked for voice in Standard

---

2. It is noted below each example whether it is obtained by elicitation or from some other source. Unless stated, all Minangkabau examples are taken from the MPI EVA Padang Field Station Minangkabau Corpus. Glossing follows the Leipzig Glossing Rules and includes the following additional abbreviations:

ACT – active, AV – active voice, CAUSE – cause/reason/actor marker, CNJ.OP – conjunction operator, COMP – comparative, CP – complementiser, CST – causative predicate, DIR – direction, EMPH – emphatic particle, EQV.OP – equivalency operator, EXCL – exclamative particle, FILL – filler, GOAL – goal, HORT – hortative, IMIT – imitation, IN – inchoative, INT – interrogative, INV – involuntary, k.o. – a kind of, MID – middle voice, N – homorganic nasal, NEGPOL – negative polite particle, NOM – nominaliser, ONE – one, OV – object voice, PARTRED – partial reduplication, PASS – passive, PERS – personal marker, PREP – generalised preposition, PV – passive voice, RED – reduplication, STAT – state, SUP – superlative, TRU – truncation.

Other symbols used include:

Ø – null constituent, xx – unclear speech, * – ungrammatical or a reconstructed form.

Minangkabau are very frequently affixless or 'bare' in Colloquial Minangkabau (see Section 4). Speakers of Colloquial Minangkabau also sometimes use the homorganic nasal form *N* to mark active voice verbs instead of the full form of the active voice marker *maN*. In Example (2), notice that in the Standard Minangkabau sentence the full form of the active voice marker *maN* is used on the verb, whereas in the Colloquial Minangkabau examples, the use of the reduced form of the active voice marker *N,* and even the use of the bare verb form, are both acceptable.

(2)   Standard Minangkabau
    a.   *Sia   mambuek?*
        sia   maN-buek
        who   AV-make
        'Who made it?'
    Colloquial Minangkabau
    b.   *Sia   mbuek?*
        sia   N-buek
        who   AV-make
        'Who made it?'
    Colloquial Minangkabau
    c.   *Nyo buek   agak limo puluah buah.*
        nyo Ø-buek agak limo puluah buah
        3   Ø-make quite five   ten   CLF:fruit
        'He made about fifty of them.'

Third, one can also find many examples of 'non canonical' word order constructions in Colloquial Minangkabau, whereas the syntactic rules of Standard Minangkabau are more rigidly defined (see Section 3).

Consider Examples (3) and (4), which further exemplify the differences between Standard and Colloquial Minangkabau.

(Standard Minangkabau)
(3)   a.   <u>*Rajo*</u> <u>*nan*</u> <u>*ba-kuaso*</u>   <u>*di*</u>   <u>*nagari*</u> <u>*Ruhun*</u> <u>*tu*</u>   *namo nyo*
        king   REL   POSS-power   LOC   country Ruhun   DEM:dist name 3
        Sultan I[…]
        Sultan I[…]
        'The king of *Ruhun* was called Sultan I[…]'
    b.   <u>*Nyo*</u> *punyo tigo   urang   anak.*
        3   have   three CLF:person child
        'He had three children.'
    c.   <u>*Nan*</u> <u>*patamo*</u> <u>*rajo*</u> <u>*anak*</u> <u>*rajo*</u> <u>*tu*</u>   *namo nyo Sultan Marajo Alif.*
        REL   first   king   child king   DEM:dist name 3   Sultan Marajo Alif
        'The first of the king's children was called Sultan Marajo Alif

    d. <u>*Sultan Marajo Alif* <u>*ko*</u></u>   *mengganti-kan jadi*   *rajo…*
       Sultan Marajo Alif DEM:prox AV:change-APPL become king…
       'Then Sultan Marajo Alif became the king…'
    e. *… mengganti Sultan I[…] ko*   *di*   *nagari Ruhun tu.*
       … AV:change Sultan I[…] DEM:prox LOC country Ruhun DEM:dist
       '… replacing Sultan […] in *Ruhun*.'

Colloquial Minangkabau

  (4) a. *Lari lah <u>nyo</u> ka dalam rimbo.*
       run PRF 3   to inside jungle
       'He ran into the jungle.'
    b. Ø *Lari ka dalam rimbo tu,*   *dari batang*   *kayu ka*
       Ø run to inside jungle DEM:dist from CLF:trunk wood to
       batang   *kayu.*
       CLF:trunk wood
       '(He) ran into the jungle from tree to tree.'
    c. Ø *nampak lah sarang tabuwan gadang.*
       Ø see   PRF nest bee   big
       '[…] ((And) he) saw a big beehive.'
       Nyo uruik-uruik lah dek   *kak*   *Kancia.*
       3   RED-massage PRF CAUSE TRU:older.sibling mousedeer
       '(And) Brother Mousedeer began to massage it.'

Example (3) is an excerpt from a talk about real and mythologised Minangkabau history, delivered by a respected elder. Since the context is formal, the audience is educated and the content is esoteric in nature, the speaker uses Standard Minangkabau. Some formal clues are present to help us identify this variety. Notice that the speaker is code mixing with Indonesian, using the Standard Indonesian forms *menggantikan*, 'change', and *mengganti*, 'replace'. Clause structure is also well defined in this excerpt; there are no null arguments and the pivot NP (underlined) appears in initial position in each clause.

    The sentences in (4), on the other hand, are examples of Colloquial Minangkabau. The excerpt comes from a children's folk tale about *kak Kancia*, 'Brother Mousedeer', who is a popular protagonist in Minangkabau folk stories. The content is therefore quite familiar. The context is also familiar, informal and intimate since the participants are all members of the same family. Notice that unlike Example (3), there are many null arguments and the pivot does not necessarily appear in canonical clause initial position. For example in (a), the pivot *nyo,* 'he', appears after the verb. Notice also that the verb *uruik-uruik*, 'massage', in (d) is morphologically unmarked for voice.

Morphological underspecification for voice represents the most important point of comparison between the two varieties since the presence of bare verbs in Colloquial Minangkabau gives its voice system a very different character. In fact, Standard Minangkabau resembles a typical Indonesian-type language in that it possesses features of both a pragmatically motivated voice system concerned with grammatical relations and a voice system that is used to encode conceptual properties of events. By contrast, morphological underspecification in Colloquial Minangkabau means that, in this variety, the role of voice marking is primarily conceptual. Colloquial Minangkabau is thus better characterised as a 'Sundic-type' language (see Sections 3 and 4).

3. Voice in Standard Minangkabau

3.1 An Indonesian-type language with a pivot function

Indonesian-type languages have voice systems that mark the realignment of the pivot function with semantic role (Foley and Van Valin 1984, 1985; Himmelmann 2005). However, unlike Philippine-type languages, which allow for a range of participant types to be assigned pivot status, in Indonesian-type languages voice alternation centres around the alignment of the pivot function with either the actor or the undergoer (cf. Klamer 1996). If the actor is selected as the pivot then the verb is marked for active voice, however if the undergoer is selected as the pivot then the verb is marked for passive voice. In Indonesian-type languages active voice is marked on the verb by a prefix that contains a final homorganic nasal *N*. Passive voice is marked on the verb by an oral prefix (i.e. a prefix with an initial oral, as opposed to nasal, stop) or a pronominal cliticised form of the agent (Ross 2002: 54–56).

Standard Minangkabau resembles a typical Indonesian-type language with an opposition between active voice, which is marked on the verb by the prefix *maN-*, and passive voice, which is marked on the verb by the proclitic *di-*. Like Standard Indonesian, Standard Minangkabau also has a *pasif semu* construction (or P2 construction) in which the verb is unmarked. In addition, and also like Standard Indonesian and other Malayic languages, Minangkabau has a class of lexical/derivational verbal affixes, including the involuntary marker *ta-*, the causative marker *pa-*, and the multifunctional prefix *ba-*. The active voice marker *maN-* also forms part of this class of affixes since it has a number of lexical/derivational functions as well. Minangkabau also makes use of two applicatives: *an* and *i*. Although the applicatives have a primarily syntactic function, they also have semantic and lexical/derivational functions (Brodkin & Fortin 2017; Crouch 2009).

Standard Minangkabau has a pragmatically motivated voice system. This means that one argument is assigned discourse relevance for a number of reasons, including that it: (1) has "constitutive relevance" (i.e. it refers to a first or second person speech act participant); (2) "it is most salient in the speaker's mind"; (3) "it plays an important role in the propositional act" (i.e. it refers to old information or is particularly conceptually salient); and (4) "it is the entity on which the hearer's attention is focused" (Shibatani 2006: 259). The high discourse relevance of the argument is reflected morphologically in that it triggers corresponding verbal marking, and syntactically by the fact that it has control over a number of restricted syntactic constructions. Thus in pragmatically motivated voice systems the notion of discourse relevance has been grammaticalised.

In Standard Minangkabau, there is a number of discourse-pragmatic factors that affect voice marking. Definite NPs, discourse topical NPs, and NPs that refer to speech act participants are more highly referential than other nominals. These nominals are therefore more likely to be selected as the pivot. The pivot has control over a number of syntactic constructions including raising, relativisation, extraction and zero anaphora. The pivot also has control over voice; it triggers active voice marking if it is an actor, and P2 marking or passive voice marking if it is an undergoer (Crouch 2009).

In addition, only core arguments may be selected as pivots in Standard Minangkabau. The distinction between core and non-core arguments reflects a difference in syntactic status, but it also reflects a difference in their "degree of discourse relevance" (Shibatani 2006: 261). Core arguments are unmarked and are licensed by the verb's argument structure. They are also assigned more discourse relevance than non-core arguments. Non-core arguments on the other hand, are not part of the verb's argument structure and are typically case marked or marked by adpositions (Foley & Van Valin 1984: 79; Arka 2003). In Minangkabau, non-core arguments are marked by prepositions (Crouch 2009: 79).

In Standard Minangkabau the distinction between core and non-core arguments is important to keep in mind as it affects applicative marking; non-core arguments can be assigned core status, and thus become available to be selected as the pivot, with only the use of applicative marking on the verb (Crouch 2009: 169). The distinction is also important since what makes the P2 construction distinct from a *di* passive, apart from verbal marking and well-defined clausal syntax, is that the actor remains a core argument. In Colloquial Minangkabau, on the other hand, there appears to be no distinction between core and non-core arguments as applicative marking is often optional and the P2 construction is indistinguishable from the 'bare verb' constructions (see Section 3.4 and Section 4.2).

3.2   Active voice

Active voice marking on the verb is triggered when the pivot aligns with the actor participant. This is marked on the verb by the prefix *maN-*. The prefix is a reflex of the Proto Malayic agent oriented verb marker **maN-* (Adelaar 1992a: 161; Blust 2013; Ross 2002: 54). The Minangkabau active voice prefix consists of the sequence *ma* and a homorganic nasal consonant. The phonological effects of the homorganic nasal on the verb stem are described in Adelaar (1992a, 1995) and in Crouch (2009: 121).

   Since the active voice prefix *maN-* is used to mark pivot alignment with the actor argument, it can only be used with a verb that specifies an actor as part of its argument structure. As such, the prefix can be used with active intransitive verb roots, transitive verb roots and ditransitive verb roots where its role is to show that the actor has been selected as the pivot. However, *maN-* can also affix to noun roots and stative verb roots. In this case, *maN-* not only marks the fact that the pivot is aligned with the actor role, it has the additional derivational effect of altering underlying argument structure of the root.

   The *maN* prefix also has some lexical/derivational effects effects, being able to derive active verbs from NPs and stative verbs. The prefix also has semantic effects and can add progressive aspectual properties to the NP roots and stative roots as part of the derivational process (Crouch 2009). The *maN* prefix is optional in Colloquial Minangkabau since it has less of a pragmatic role than in Standard Minangkabau, functioning primarily within this conceptual and semantic domain (see Section 4).

3.2.1   *Canonical* maN- *clauses*
In a typical *maN-* clause, the actor appears in pre-verbal pivot position. If the verb is transitive, then the undergoer appears in post-verbal position. This canonical word order is exemplified in (5).

Standard Minangkabau

(5)   *Tadi   malam den   mandanga musik.*
      before night 1SG   AV:hear   music
                    ACTOR VERB   UNDERGOER
      'Last night I listened to music.'   (Elicitation)

Non-core arguments typically occur after the undergoer and are marked by a preposition as in (6).

Standard Minangkabau

(6)   a.   *Ambo   mananam bungo   di   kabun*
           1SG   AV:plant flower   LOC   garden
           ACTOR VERB   UNDERGOER NON-CORE
           'I'm planting flowers in the garden.'   (Elicitation)

b.   *Ambo   mananam-i   kabun tu   jo   bungo.*
     1SG   AV:plant-APPL:loc garden DEM:dist with   flower
     ACTOR   VERB   CORE   NON-CORE
     'I planted the garden with flowers'   (Elicitation)

3.2.2   *Non-canonical* maN *clauses*

Although unacceptable in Standard Minangkabau, many examples of non-canonical word orders can be found in Colloquial Minangkabau. In fact, in many clauses with intransitive *maN-* verbs, the actor appears in post-verbal position. This kind of word order occurs in discourse when the actor argument is given as an afterthought, or if the actor argument is particularly pragmatically salient in terms of the development of narrative events. Examples of the 'verb + actor' word order can be found in (7), (8), and (9).

Colloquial Minangkabau

(7) *Baru awak dari pasa   manyuruak nyo.*
     new 1   from market AV:hide   3
                          VERB   ACTOR
     'I had just come out from the market when he hid.'

Colloquial Minangkabau

(8) *Eem   manggaleh nyo   Jakarta.*
     EMPH AV:sell   3   Jakarta
            VERB   ACTOR
     'Uh huh he is selling in Jakarta.'

Colloquial Minangkabau

(9) *Sadang mamanjek pareman ko   nak   tibo   pak   gaek ko.*
     PROG   AV:climb hoodlum DEM:prox EMPH arrive father old   DEM:prox
            VERB   ACTOR
     'Just as the hoodlum was climbing up, right, the old man arrived.'

An 'undergoer + verb' word order is also possible in transitive *maN* clauses in Colloquial Minangkabau. Examples of this kind of word order can be found in (10) and (11).

Colloquial Minangkabau

(10) *Beko mintak-e   tu   mangkabua.*
     later request-3   DEM:dist AV:answer
            UNDERGOER   VERB
     'Later his request would be granted.'

Colloquial Minangkabau

(11) *Lai,   lai   kami   itu   nyo   mandanga.*
    more more 1   DEM:dist 3   AV:hear
                          ACTOR   UNDERGOER VERB
    'Yeah, yeah we heard it.'

Another feature of Colloquial Minangkabau is that active transitive verbs may appear with one or more of their arguments omitted. This is only permitted in Standard Minangkabau if the verb is marked for voice (Fortin 2001: 14), whereas in Colloquial Minangkabau arguments may be omitted even if the verb is bare. Sentences (12) and (13) show two examples in which the *maN-* verb is transitive, but in which the undergoer argument is not specified.

Colloquial Minangkabau

(12) *A   ba-a   caro   mananyo Ø dek   Engki?*
    FILL POSS-what manner AV:ask   Ø CAUSE Engki
    'But how did you, Engki, ask (him)?'

Colloquial Minangkabau

(13) *Ø La tingga lo   samo jo   uda   tu   manggaleh Ø.*
    Ø PRF live   EMPH with with older.brother DEM:dist AV:sell   Ø
    '(He) lives with his older brother selling (stuff).'

In Colloquial Minangkabau bare transitive verb roots also regularly appear in clauses in which some of their arguments are omitted. For example, in (14) the verb *agiah*, 'give', appears in its bare form even though two of the verb's arguments, the undergoer and the recipient, are unexpressed. Similarly in (15), the undergoer argument of the bare verb *tokokan*, 'hit', is omitted.

Colloquial Minangkabau

(14) *Tapi ketiko ado   awak agiah Ø Ø se   nanti.*
    but when exist 1   give Ø Ø just later
    'But when it's here I'll just give (it (to you)) later.'

Colloquial Minangkabau

(15) *Ba-a   ko   nyo tokok-an Ø kayak gitu   tu   Yes?*
    POSS-what DEM:prox 3   hit-APPL Ø like DEM:dist DEM:dist TRU:Maiyes
    'How come he hit (it) like that Yes?'

The existence of these non-canonical structures in Colloquial Minangkabau suggests that 'voice' marking does not have the same effect on the syntactic organisation of the clause as it does in Standard Minangkabau. Colloquial Minangkabau is also more 'Associational', which means that syntactic clues are not necessary to determine semantic role assignment (see Section 4).

3.3 Passive voice

In Minangkabau, passive voice is encoded by the morpheme *di*, i.e. *di* shows that the pivot function is aligned with the undergoer. The Minangkabau passive voice marker is cognate with Malay/Indonesian *di*, which is described as a prefix (Dardjowidjojo 1978; Musgrave 2000; Sneddon 1996). However, *di* behaves more like a proclitic in Minangkabau and the morphophonemic and prosodic qualities of *di* differ considerably from the active voice prefix *maN* and the other verbal prefixes *ta*, *pa*, and *ba* (Crouch 2009; Williams 1961: 66–67). The clitic-like behaviour of *di* can partly be explained by its historical origins as a preposition (Adelaar 1992a, 1992b, 2005, 2008; Crouch 2009: 134–136).

3.3.1 *Canonical* di *clauses*
Passive voice clauses are highly marked structures in the standard variety and typically follow a rigid set of structural parameters.

In passive voice, undergoers occupy pre-verbal pivot position and actors appear in post-verbal position. The actor may be expressed as a post-verbal enclitic, as a full NP, or as an adjunct NP marked by the preposition *dek*.

For example, in the Standard Minangkabau sentence in (16), the pivot *awak*, which also happens to be an undergoer, appears in pre-verbal position. The verb is correspondingly marked for passive voice by the proclitic *di-* and the actor appears in post-verbal position. Notice that in (a) the actor, *nyo*, 'him', is expressed as an enclitic, whereas in (b) it is expressed as a full NP, *paman*, 'uncle', and in (c) as an adjunct NP, *dek paman*, 'by my uncle'. Note that adding *dek* in (c) clearly marks the semantic role of the actor. This is especially important if the word order is non-canonical (see (d) and (e)).

Standard Minangkabau
(16) a. *Awak di=gadang-an-nyo.*
          1   pv=big-APPL-3
          'I was raised by him.'   (Elicitation)
      b. *Awak di=gadang-an paman.*
          1   pv=big-APPL uncle
          'I was raised by my uncle.'   (Elicitation)
      c. *Awak di=gadang-an dek   paman.*
          1   pv=big-APPL CAUSE uncle
          'I was raised by my uncle.'   (Elicitation)
      d. *Dek   paman awak di=gadang-an.*
          CAUSE uncle   1   pv=big-APPL
          'I was raised by my uncle.'   (Elicitation)
      e. **Paman awak di=gadang-an.*
          uncle   1   pv=big-APPL   (Elicitation)

All passive voice verbs are necessarily transitive since they require an undergoer and an actor, although not all arguments are necessarily overtly expressed in the clause, particularly in Colloquial Minangkabau. If the verb root being passivised is stative or an active intransitive, an applicative is required to license an additional argument first.

The *an* applicative can license a benefactor argument and can also make benefactor arguments available to be selected as the pivot in a passive clause (see (17)).

Standard Minangkabau

(17) a. *Bini-nyo mambuek-an kopi   untuak Udin.*
          Wife-3   AV:make-APPL coffee for   Udin
          'Udin's wife made coffee for him.'   (Elicitation)

      b. *Udin di=buek-an   kopi dek   bini-nyo.*
          Udin PV=make-APPL coffee CAUSE wife-3
          'Udin's wife made him coffee.'   (Elicitation)

      c. **Udin di=buek kopi dek   bini-nyo.*
          Udin PV=make coffee CAUSE wife-3   (Elicitation)

One of the functions of the applicative *i*, is to license a locative argument, making it available for pivot selection in a passive clause, as (18) demonstrates.

Standard Minangkabau

(18) a. *Bungo di=tanam-nyo di   kabun.*
          flower PV=plant-3   LOC garden
          'He planted flowers in the garden.'   (Elicitation)

      b. *Kabun di=tanam-i-nyo   jo   bungo.*
          garden PV=plant-APPL:loc-3 with flower
          'The garden was planted with flowers.'   (Elicitation)

      c. **Kabun di=tanam-nyo bungo.*
          garden PV=plant-3   flower   (Elicitation)

3.3.2 *Non-canonical* di *clauses*

Many examples of non-canonical word orders in *di-* passive clauses can be found in Colloquial Minangkabau. In fact, the undergoer argument may frequently appear in post-verbal position, as shown in (19), (20) and (21) (the undergoer appears in boldface type).

Colloquial Minangkabau

(19) *Di=suruah lari **Ida** Ni   **Da**   tu.*
      PV=order run **Ida** TRU:older.sister **TRU:Ida** DEM:dist
      '*Uni*, I, Ida was told to run.'

Colloquial Minangkabau

(20) *Di=bali **pisang** tidak mau makan.*
PV=buy **banana** NEG want eat
'(He) bought a banana but didn't want to eat it.'

Colloquial Minangkabau

(21) *Di=buat-nyo **vila** kan.*
PV=make-3 **villa** EMPH
'He built a villa you know.'

Like active voice constructions, the verb's arguments in a *di* passive clause are often omitted in discourse if the context permits. Since the undergoer argument is the pivot in passive voice clauses, and therefore highly referential, it is often unexpressed in passive voice clauses, as Examples (22) and (23) demonstrate.

Colloquial Minangkabau

(22) *Uang saku punyo awak se lai tapi beko Ø di=agiah-nyo snack.*
pocket.money own 1 only more but later Ø PV=give-3 snack
'We have pocket money but later on (we) get given snacks.'

Colloquial Minangkabau

(23) a. *Nyo kan di=tanyo a isi jalang-e.*
3 EMPH PV=ask what contents wedding.gift-3
'They were asked what was in the wedding gift.'
   b. *Ø Di=caliak dek urang ma.*
Ø PV=look CAUSE person EMPH
'(The contents of the wedding gift) were examined.'

The actor argument is also frequently not expressed if it is retrievable from the discourse context or if it is underspecified in some way. For example, in (24) and (25) the actor participants are unexpressed but understood to mean 'someone'.

Colloquial Minangkabau

(24) *Patah roda-nyo kalau di=dorong Ø.*
broken wheel-3 TOP PV=push Ø
'Its wheel will break if it is pushed.'

Colloquial Minangkabau

(25) *Duo jalur ka di=buek Ø di sinan tu tu ma.*
two lane FUT PV=make Ø LOC there DEM:dist DEM:dist EMPH
'(They're) going to make it two lanes over there you know.'

3.4 The P2 construction

3.4.1 *The Standard Minangkabau P2 construction*

In Standard Minangkabau, there is another kind of passive voice construction called the *pasif semu* or the P2 construction. Like a *di* passive, the undergoer in a P2 construction performs the pivot function and the actor appears as a full NP. The P2 verb is unmarked yet the construction remains distinctive from a bare active construction because of the word order constraints it exhibits. However, due to the high frequency of bare verbs and the flexible word order constraints of Colloquial Minangkabau, the P2 is not a distinctive construction in this variety.

The P2 construction in Standard Minangkabau closely resembles the Malay/Indonesian P2 construction. In the Malay/Indonesian P2 construction the verb is not marked for voice and the actor-verb word order of an active sentence is retained. However, the undergoer occurs in pre-verbal position and is assigned pivot status (Arka and Manning 2008; Chung 1976; Cole, Hermon and Tjung 2006). A further distinguishing feature of the P2 construction is that no auxiliaries (i.e. negators and TAM adverbials) may intervene between the actor and the verb (Sneddon 1996: 249).

These constraints also apply to the P2 construction in Standard Minangkabau. An example of a canonical Minangkabau P2 clause can found in (26). In this example, the undergoer argument, *buku tu*, 'that book', is in pivot position. Notice also that the verb is unmarked for voice and that the TAM adverbial *alun* cannot intervene between the actor, *den*, and the verb, *baco*, 'read'.

Standard Minangkabau

(26) a. *Buku tu   alun   den Ø-baco.*
          book DEM:dist not.yet 1SG Ø-read
          'I haven't read that book yet.'   (Elicitation)

      b. **Buku tu   den alun   Ø-baco.*
          book DEM:dist 1SG not.yet Ø-read
          'That book I haven't read.'   (Elicitation)

Also consider Example (27). In this case the positioning of the perfective aspectual marker *alah* actually changes the meaning of the sentence. Example (a) is a canonical P2 clause with the first-person pronoun *ambo* filling the actor argument slot. However, in (b), *ambo* becomes a possessor rather than an actor and the verb *masak*, 'cook', has a stative reading. This change in semantic role assignment is a result of the actor-auxiliary inversion, which demonstrates not only that the word order constraints on the P2 construction in Standard Minangkabau are restrictive, but that a change in word order can also have an effect on the meaning of the clause.

Standard Minangkabau

(27) a. *Gulai dagiang alah ambo Ø-masak.*
     curry meat   PRF 1SG   Ø-cook
     'I've cooked a beef curry.'   (Elicitation)

     b. *Gulai dagiang ambo alah masak.*
     curry meat   1SG   PRF cook
     'My beef curry is cooked.'   (Elicitation)

Like the Standard Indonesian P2 construction, in the Standard Minangkabau P2 construction, only the undergoer argument in an embedded P2 clause can be raised to pivot position in the matrix clause (see Example (28)). This demonstrates that the undergoer is the pivot in a P2 clause, differentiating the P2 clause from a bare active construction (cf. Chung 1976).

Standard Minangkabau

(28) a. *Inyo maanggap gulai   dagiang alah ambo Ø-masak.*
     3   AV:believe curry   meat   PRF 1SG   Ø-cook
                 UNDERGOER
     'They believe I've already cooked the beef curry.'   (Elicitation)

     b. *Gulai   dagiang di=anggap-nyo alah ambo Ø-masak.*
     curry   meat   PV=believe-3 PRF 1SG   Ø-cook
     UNDERGOER
     'They believe I've already cooked the beef curry.'   (Elicitation)

     c. **Ambo   di=anggap-nyo gulai   dagiang alah Ø-masak.*
     1SG   PV =believe-3   curry   meat   PRF Ø-cook
     ACTOR
                                        (Elicitation)

3.4.2 *Bare verbs and the P2 construction*

The P2 construction is indistinguishable from the bare active construction in Colloquial Minangkabau because of the flexible word order constraints and the pervasive use of bare verbs in this variety (see Section 4). More specifically we find that the constraints governing the ordering of clausal components in the Standard Minangkabau P2 clause do not apply in Colloquial Minangkabau, and there appear to be no surface or deep syntactic differences between bare active constructions and constructions which more closely resemble the P2 construction. This is not an unusual phenomenon and there is reason to believe that in some non-standard, regional, and contact varieties of Malay/Indonesian, such as Basilectal Jakarta Indonesian, Sarang Lan Malay, Mundung Darat Malay and Kuching Malay, the P2 construction has also lost its distinctiveness (Cole, Hermon and Tjung 2006; Cole, Hermon and Yanti 2007).

In Colloquial Minangkabau, clauses with bare verbs allow both 'auxiliary + actor' and 'actor + auxiliary' word orders. Since a variety of interpretations is available for clauses with bare verbs (see Section 4), the shift in word order does not entail any shift in meaning. For example, consider the sentences in (29) and (30) which show that there are no constraints on the ordering of the actor and the auxiliary adverbials.

Standard Minangkabau

(29)   a.   *Inyo acok   ubah   gaya-nyo.*
            3   often change style-3
            'He often changes his style.'   (Elicitation)

       b.   *Acok inyo ubah   gaya-nyo.*
            often 3   change style-3
            'He often changes his style.'   (Elicitation)

       c.   *Gaya-nyo acok   inyo ubah.*
            style-3   often 3   change
            'He often changes his style.'   (Elicitation)

       d.   *Gaya-nyo inyo acok   ubah.*
            style-3   3   often change
            'He often changes his style.'   (Elicitation)

Standard Minangkabau

(30)   a.   *Aden alah baco buku ko.*
            1SG   PRF read book DEM:prox
            'I have read this book.'   (Elicitation)

       b.   *Alah aden baco buku ko.*
            PRF 1SG read book DEM:prox
            'I have read this book.'   (Elicitation)

       c.   *Buku ko   alah aden baco.*
            book DEM:prox PRF 1SG read
            'I have read this book.'   (Elicitation)

       d.   *Buku ko   aden alah baco.*
            book DEM:prox 1SG PRF read
            'I have read this book.'   (Elicitation)

Similarly in (31), notice that the ordering of the auxiliary and the actor (*alah aden tu*) resembles the ordering of a canonical P2 clause. The clause can also be interpreted as a bare active since the verb can be optionally marked by the active voice marker *maN-* as shown in sentence (b).

Colloquial Minangkabau

(31) a. *Alah aden tu   pa racik   den racik   sado-e*
          PRF   1SG   DEM:dist xx thinly.slice 1SG thinly.slice all-3
          dulu   tu.
          moment DEM:dist
          'I've already sliced it, I sliced it all before.'

      b. *Alah aden tu   maracik   den racik   sado-e*
          PRF   1SG   DEM:dist AV:thinly.slice 1SG thinly.slice all-3
          dulu   tu.
          moment DEM:dist
          'I've already sliced it, I sliced it all before.'   (Elicitation)

There are thus no surface syntactic differences between the bare active and the P2 constructions in Colloquial Minangkabau. Applying raising tests to what appear to be P2 constructions in Colloquial Minangkabau also reveals that there are no underlying syntactic differences between bare active and supposed P2 clauses since both actors and undergoers can be 'raised' (see Example (32)).

Standard Minangkabau

(32) a. *Di=anggap-nyo buku   ko   alah aden baco.*
          PV=believe-3   book DEM:prox PRF 1SG   read
                          UNDERGOER
          'They believe that this book, I read it.'   (Elicitation)

      b. *Buku   ko   di=anggap-nyo alah aden baco.*
          book   DEM:prox PV =believe-3 PRF 1SG   read
          UNDERGOER
          'This book, they believe I read it.'   (Elicitation)

      c. *Aden   di=anggap-nyo buku ko   alah baco.*
          1SG   PV=believe-3 book DEM:prox PRF read
          ACTOR
          'I, they believe, read this book.'   (Elicitation)

4.   Voice in Colloquial Minangkabau

We have already seen that many of the syntactic constraints that define grammatical relations and semantic role assignment in the Standard Minangkabau clause do not apply in Colloquial Minangkabau. In addition to this, in Colloquial Minangkabau, bare verbs are frequently found in contexts where one would expect to find a verb marked for active or passive voice in Standard Minangkabau. Utterances with bare verbs are not *ungrammatical* in this variety, rather they are *underspecified*. This

means that in Colloquial Minangkabau *maN-* and *di-* are optional and their function is primarily semantic or conceptual as opposed to syntactic or pragmatic. Consequently, Colloquial Minangkabau appears to have more in common typologically with 'isolating' Austronesian languages and is thus better characterised as having a 'Sundic-type' voice system rather than an Indonesian-type voice system.

4.1   'Sundic-type' voice

The 'Sundic-type' group is a set of typologically similar languages which appear to be pivotless (Gil 2008, 2015a). In this set, Gil includes Sundanese (cf. Hardjadibrata 1985), Riau Indonesian and other Colloquial Malay/Indonesian varieties such as Jakarta Indonesian, Sulsel Indonesian, Irian Indonesian and Kuala Lumpur Malay (cf. Gil 2002), Colloquial Javanese (cf. Conners 2008), as well as Colloquial Minangkabau. Gil also argues that Mentawai displays some Sundic-type features. All these languages have traditionally been characterised as having Indonesian-type voice systems. However, Gil argues that 'voice' morphology in these languages encodes a semantic distinction rather than a syntactic distinction.

Indonesian-type languages have active and passive voice constructions as well as a *pasif semu* construction (see Section 3). Sundic-type languages, on the other hand, have a three-way distinction between a neutral construction where the verb is unmarked for 'voice' (i.e. it is a bare verb), a generalised active construction where the verb is marked by a 'generalised active' morpheme, and a generalised passive construction in which the verb is marked by a 'generalised passive' morpheme. An additional difference is that unlike Indonesian-type languages, in which the pivot must precede the verb, Sundic-type languages are pivotless, therefore any NP may precede the verb (Gil 2008). The differences between these two kinds of voice systems are summarised in Table 1.

Riau Indonesian, a colloquial variety of Indonesian used for inter-ethnic communication in Riau Province, Sumatra, has been described in detail by Gil (cf. 1994, 2001, 2002, 2006) and perhaps demonstrates the typological properties of the Sundic-type most effectively. In Riau Indonesian, Gil (2002) argues that the prefix *N* and the proclitic *di*, cognate with Standard Malay/Indonesian *meN* and *di* respectively, are generalised verbal semantic markers and do not encode a voice distinction. According to Gil's analysis, *N* marks the fact that the actor participant is referentially salient or conceptually significant in the utterance. Similarly, *di* marks the patient participant as referentially or conceptually important. The Riau Indonesian semantic markers *N* and *di* also do not correspond with any shift in word order or syntactic structure and can even be used to mark the conceptual salience of a null actor or a null patient. Similar arguments have also been made for Jakarta Indonesian (Conners, Bowden & Gil 2015).

Table 1. Indonesian-type voice and Sundic-type voice (Gil 2008)

	CLAUSE STRUCTURE	VOICE TYPE
INDONESIAN-TYPE VOICE	Actor AV-Verb Undergoer	active
	Undergoer PV-Verb Actor	passive (symmetric)
	STANDARD INDONESIAN ALSO HAS:	
	Undergoer PV-Verb PP Actor	passive (asymmetric)
	Undergoer Actor-Verb	*pasif semu*
SUNDIC-TYPE VOICE	NP Verb NP	neutral
	NP general.ACT-Verb NP	generalised active
	NP general.PASS-Verb NP	generalised passive

Example (33) provides a minimal pair to demonstrate the use of *N* and *di* with the same verb root, *simer*, 'polish'.

Riau Indonesian

(33) a. *Mister,   aku nyimer   lagi.*
       white.person 1SG N-polish CNJ.OP
       'I'm going off to shine shoes.'
       [Context: At table with shoeshine boys; speaker takes leave.]

(Gil 2002: 262)

   b. *Aden disimer.*
       1SG   DI-polish
       'I'm polishing them.'
       [Context: Shoeshine boy pointing to potential customer's sandals, addressing other shoeshine boys, who are possible competitors.] (Gil 2002: 249)

In both (33a) and (33b) the actor precedes the verb and in neither example is a patient overtly expressed. Nevertheless, the verb in (33a) is marked by the generalised active marker *N* whereas the verb in (33b) is marked by the generalised patient marker *di*. Since the shift in verb marking does not entail any change in syntactic organisation or grammatical relations, Gil (2002) concludes that *N* and *di* must encode a semantic or conceptual distinction. Indeed, if we look at the context for each of the utterances in (33) we find that in (a) the actor, the boy taking leave, is more conceptually salient, whereas in (b) it is the patient, the customer's sandals, that is the conceptually salient participant. Thus, it is this conceptual shift in focus to the actor or patient that triggers the use of *N* and *di* respectively.

Sundic-type languages like Riau Indonesian also have a 'neutral' construction (see Table 1) in which the verb is unmarked for voice. The sentences in (34) show two Riau Indonesian 'neutral' constructions in which the verb *beli*, 'buy', is unmarked.

Riau Indonesian

(34) a. *Beli aku laser, kan.*
      buy 1SG laser Q
      'I'll buy a laser, right.'
      [Context: Contemplating a shopping trip.]   (Gil 2006: 43)

    b. *Beli nasi goreng aku.*
      buy rice fry   1SG
      'I bought the fried rice.'
      [Context: Group of people decide they want to play cards; somebody tells
      speaker to go out and buy some; speaker objects on the grounds that it's
      somebody else's turn to go out.]   (Gil 2006: 43 – Riau Indonesian)

Like the generalised active and passive constructions, a 'neutral' construction does
not entail any shift in grammatical relations and actor and patient NPs are unre-
stricted in terms of which position in the clause they can appear. So in (34) notice
that the verb is followed by the actor, the first person pronoun *aku*, and then the
patient, *laser*, ' a laser' in (a). However, in (b) the verb is followed by the patient,
nasi goreng, 'fried rice', and then the actor, *aku*.

    The fact that 'voice' in Riau Indonesian does not encode any distinctions in
grammatical relations, coupled with the additional fact that 'neutral' verbs can be
left unmarked, implies that unless significant contextual background is given, the
semantic roles of the participants in the clause will be underspecified. In other
words, since neither word order nor verbal marking can tell us who is the pa-
tient and who is the actor, or indeed who is the pivot, Riau Indonesian clauses
remain rather vague. According to Gil (2001, 2002, 2006, 2007), the vagueness
and underspecification of semantic roles (and many other clausal elements in Riau
Indonesian) does not impede communication in the language and the interpret-
ability of underspecified utterances in the language actually relies on a system of
'Associational Semantics' (see Section 4.3.3). In fact, since Sundic-type voice sys-
tems entail underspecification of semantic roles by their very nature, Gil (2008)
argues that Associational Semantics are a feature of all Sundic-type languages.

4.2 Bare verbs

Colloquial Minangkabau has a construction similar to the 'neutral' construction in
Riau Indonesian, which I call the 'bare verb' construction. Bare verbs must appear
obligatorily in a number of clause types in Standard Minangkabau, for example,
in stative clauses, imperatives, prohibitives and P2 constructions. There is also a
number of lexicalised bare verbs in Minangkabau that are not marked for voice in
either Standard or Colloquial Minangkabau (Crouch 2009: 185–186), such as *lari*,
'run', and *tagak*, 'stand', in Examples (35) and (36).

Colloquial Minangkabau

(35) *Lari lah nyo ka dalam rimbo.*
run PRF 3 to inside jungle
'He ran into the jungle.'

Colloquial Minangkabau

(36) *Makin banyak, kami tagak jo situ.*
increasingly many 1pl stand also there
'There were more and more (of them, but) we kept standing there.'

In a true 'neutral' or 'bare verb' construction, the main verb is unmarked for voice and the actor and undergoer NPs can appear in any position around the verb. In an equivalent sentence in Standard Minangkabau the verb would be marked for voice by either *maN* or *di*. Since Standard Minangkabau is the prestige variety, bare verb constructions are generally considered by Minangkabau speakers to be 'incorrect', 'improper' and 'sloppy' even though they are clearly a systematic feature of Colloquial Minangkabau.

In a bare verb construction, the semantic roles of the participants are indeterminate because word order is not restricted. Whether the verb can be assigned a 'voice' is also often unclear because the verb is unmarked and there is no clearly defined pivot slot in the clause. Nevertheless, hearers are still able to interpret the meanings of bare verbs, understand the referents of the participants in the clause, and assign semantic roles. Context and Associational Semantics play a significant role in the interpretability of bare verb constructions, as will be explained in Section 4.3.

Bare verb constructions are not peculiar to Riau Indonesian or Colloquial Minangkabau. In fact, bare verbs may be a feature of colloquial registers of Austronesian languages more generally. For example, bare verbs can be found in object voice constructions in Madurese where voice marked verbs would be expected instead. Interestingly, most speakers find this kind of bare verb construction unacceptable even though they might frequently use it themselves in conversation, albeit unawares (Davies 2005: 201). Bare verb forms are also used in Spoken Jakarta Indonesian (Wouk 1989, 2004; Conners, Bowden & Gil 2015). These forms are treated by Wouk as unaffixed active verbs that have a slightly different discourse function to voice marked active verbs, whereas Conners et al. argue that bare verbs in Jakarta Indonesian are unspecified for voice. It has also been argued that *N-* in Jakarta Indonesian, traditionally described as the active voice marker, actually encodes aspect and event structure (Conners & Brugman 2013; Ginsberg & Paauw 2010; Hidayat 2011).

Similarly, in colloquial 'low' varieties of Malay, bare forms of active verbs are used with more frequency than voice marked forms (Benjamin 1993: 366–367). And in Tagalog, bare active verbs also occur in naturalistic data (Himmelmann 2008), albeit far less frequently than in Malayic varieties and Colloquial Minangkabau.

However, unlike Madurese where bare verbs are limited to object voice clauses (cf. Davies 2005), or colloquial Malay (cf. Benjamin 1993) and Tagalog (cf. Himmelmann 2008) where bare verbs are treated as unmarked active verbs, bare verbs in Colloquial Minangkabau are neither active nor passive. In fact, the existence of the bare verb construction demonstrates that *maN-* and *di-* are optional and syntactic voice is not a relevant category in Colloquial Minangkabau.

Contextual information is extremely important in helping to decode the semantic roles of participants in a clause with a bare verb. So much contextual knowledge is required to disambiguate bare verbs that outsiders to the conversation are sometimes unable to do so. In Examples (37) and (38), the (a) sentences come from the MPI EVA Minangkabau corpus. The sentences were then taken out of context and my Minangkabau language consultants were asked to assign the actor and undergoer roles to each of the participants in the clause and transform the sentence into active and passive voice. The active and passive alternatives given by the consultants are provided in the (b) and (c) sentences. An active interpretation of these utterances seems to be preferred as a 'default' option. Nevertheless, given the right context, however unusual that context may be, a passive interpretation is still available. This means that the actor and undergoer may appear in any position within a bare verb construction.

Colloquial Minangkabau

(37) a. *Eka tu   masak di   dapua, masak mi   untuk cowok-e.*
          Eka DEM:dist cook LOC kitchen cook noodles for   guy-3
          'Eka is in the kitchen making noodles for her boyfriend.'/
          'Eka is being cooked in the kitchen, cooked by noodles for her boyfriend.'
      b. *Eka tu   mamasak di   dapua, mamasak mi   untuk cowok-e.*
          Eka DEM:dist AV:cook LOC kitchen AV:cook noodles for   guy-3
          'Eka is in the kitchen making noodles for her boyfriend.'   (Elicitation)
      c. *Eka tu   di=masak di   dapua, di=masak mi   untuk cowok-e.*
          Eka DEM:dist PV=cook LOC kitchen PV=cook noodles for   guy-3
          'Eka is being cooked in the kitchen, cooked by noodles for her boyfriend.'
                                                                      (Elicitation)

(38) Colloquial Minangkabau

      a. *Kebetulan ado   urang bali padi.*
          in.fact   exist person buy rice.plant
          'In fact, somebody bought the rice plants.'/
          'In fact, somebody was bought by the rice plants.'
      b. *Kebetulan ado   urang mambali padi.*
          in.fact   exist person AV:buy rice.plant
          'In fact, somebody bought the rice plants.'   (Elicitation)

   c.   *Kebetulan ado   urang di=bali padi.*
       in.fact   exist person PV=buy rice.plant
       'In fact, somebody was bought by the rice plants.'   (Elicitation)

Similarly, in Examples (39) to (44), the (a) sentences, which are bare verb constructions, are taken from the MPI EVA corpus. In these sentences the syntax of the clause is irrelevant when determining the semantic role of the participants. When the examples are taken out of context, the pre-verbal participants in each of the (a) sentences can be assigned either the actor or undergoer role without affecting the grammaticality of the clause. In fact, as the (b) and (c) sentences reveal, assigning the pre-verbal participant the undergoer role and marking the verb for passive voice produces a sentence just as acceptable as assigning the participant the actor role and marking the verb for active voice.

Colloquial Minangkabau

(39) a. *Tu   gigik acek   ndak Mbon?*
         DEM:dist bite   leech NEG TRU:Mambon
         'It was bitten by a leech wasn't it Mbon?'/
         'It bit a leech didn't it Mbon?'

      b. *Tu   di=gigik acek   ndak Mbon?*
         DEM:dist PV=bite leech NEG TRU:Mambon
         'It was bitten by a leech wasn't it Mbon?'   (Elicitation)

      c. *Tu   manggigik acek   ndak Mbon?*
         DEM:dist AV:bite   leech NEG TRU:Mambon
         'It bit a leech didn't it Mbon?'   (Elicitation)

Colloquial Minangkabau

(40) a. *Tu   pisang karek-karek potong-potong.*
         DEM:dist banana RED-chop   RED-cut
         'Then the banana is chopped up, cut up.'/
         'The banana chopped (it) up, cut (it) up.'

      b. *Tu   pisang di=karek-karek di=potong-potong.*
         DEM:dist banana PV=RED-chop PV=RED-cut
         'Then the banana is chopped up, cut up.'   (Elicitation)

      c. *Tu   pisang mangarek-karek mamotong-potong.*
         DEM:dist banana AV:RED-chop   AV:RED-cut
         'The banana chopped (it) up, cut (it) up.'   (Elicitation)

Colloquial Minangkabau

(41) a. *Puskesmas   alun   bukak lai.*
         community.health.centre not.yet open   EMPH
         'The Community Health Centre isn't open yet.'/'
         The Community Health Centre hasn't opened (it) yet.'

b. *Puskesmas   alun   di=bukak lai.*
community.health.centre   not.yet PV=open EMPH
'The Community Health Centre isn't open yet.'   (Elicitation)

c. *Puskesmas   alun   mambukak lai.*
community.health.centre   not.yet AV:open   EMPH
'The Community Health Centre hasn't opened (it) yet.'   (Elicitation)

Colloquial Minangkabau

(42) a. *Lah cuci muko Afif?*
PRF wash face Afif
'Have you washed your face yet Afif?'/
'Has your face been washed yet Afif?'

      b. *Lah mancuci muko Afif?*
PRF AV:wash face   Afif
'Have you washed your face yet Afif?'   (Elicitation)

      c. *Lah di=cuci   muko Afif?*
PRF PV=wash face   Afif
'Has your face been washed yet Afif?'   (Elicitation)

Colloquial Minangkabau

(43) a. *Tu   masak nyo ba-a?*
DEM:dist cook   3   POSS-what
'How did he cook it?'/
'How is it cooked?'

      b. *Tu   mamasak nyo ba-a?*
DEM:dist AV:cook   3   POSS -what
'How did he cook it?'   (Elicitation)

      c. *Tu   di=masak nyo ba-a?*
DEM:dist PV=cook   3   POSS -what
'How is it cooked?'   (Elicitation)

Colloquial Minangkabau

(44) a. *Aden galak den caliak-e ma.*
1SG   laugh 1SG see-3   EMPH
'I laughed to see him.'/
'I laughed and he looked at me.'

      b. *Aden galak den mancaliak-e ma.*
1SG   laugh 1SG AV-see-3   EMPH
'I laughed to see him.'   (Elicitation)

      c. *Aden galak den di=caliak-e ma.*
1SG   laugh 1SG PV=see-3   EMPH
'I laughed and he looked at me.'   (Elicitation)

The canonical passive voice clause in Standard Minangkabau adheres to a rigid clause structure, with the actor participant appearing either as a post-verb enclitic or as a full NP optionally marked by *dek*. In a Colloquial Minangkabau bare verb construction, *dek* is also used to mark the actor participant.

Examples (45), (46) and (47) come from two folk tales about *Kak Kancia*, 'Brother Mousedeer'. *Kak Kancia* is the central character and referred to in these examples by the third person pronoun *nyo*. Even though the verbs *ambuangan*, 'throw', *sasakan*, 'surround', and *tangkok*, 'catch' are unmarked for voice, without any additional contextual information we can safely assign the undergoer role to *nyo*, 'he', and the actor role to the other NP because it is marked by the *dek* preposition.

Colloquial Minangkabau

(45) *Nyo ambuang-an lah   dek   gajah   jo   balalai.*
    3   throw-APPL   EMPH CAUSE elephant with trunk
    'Elephant threw him out with his trunk.'

Colloquial Minangkabau

(46) *Nyo sasak-an juo lah   dek   baruak   ko.*
    3   tight- APPL also EMPH CAUSE short.tailed.macaque DEM:prox
    'And he was surrounded by the short-tailed macaques.'

Colloquial Minangkabau

(47) *Singkek carito, lah nyo tangkok dek   pak   tani.*
    short   story PRF 3   catch   CAUSE father farm
    'To cut the story short, he was caught by the farmer.'

In Examples (48) and (49), there is only one available interpretation as the post-verbal '*dek* + NP' clearly marks the actor participant.

Colloquial Minangkabau

(48) *Tapi nyo agiah-e dek   urang cek-e.*
    but 3   give-3 CAUSE person talk-3
    'Someone gave it to him, he said.'

Colloquial Minangkabau

(49) *Nyo jawek dek   si   Malin ko.*
    3   answer CAUSE PERS Malin DEM:prox
    'Malin answered him.'

In Colloquial Minangkabau bare verb constructions, the actor participant can also appear as a post-verbal enclitic. The bare verb construction in (50) is interesting as both participants appear after the verb. However, because it is encoded as a post-verbal enclitic, the actor role must be assigned to *nyo*, 'he'.

Colloquial Minangkabau

(50) *Tanam-nyo sawit.*
plant-3 palm.oil.tree
'He planted palm oil trees.'

In Standard Minangkabau the applicative *an* plays a clearly defined role as a valency changing device (Crouch 2009: 170–175). However, in Colloquial Minangkabau, *an*, like *maN* and *di*, appears to be optional. For example, (51) shows that the applicative *an* is not always required to create transitive causative verbs from intransitive verb roots in Colloquial Minangkabau. Sentence (a) shows how the verb *malarian*, 'kidnap', is used. The verb is transitive and causative and is created by affixing *an* to the active intransitive verb root *lari*, 'run'. Sentence (b) shows that the 'kidnap' interpretation of the verb is also possible even if the verb is not marked by the applicative. We would expect that the sentence in (b) would mean 'yes the child ran', but it is clear from the context that the speaker intends the 'kidnap' meaning of the verb.[3]

Colloquial Minangkabau

(51) a. *Malari-an anak surang.*
AV:run-APPL child ONE:person
'He kidnapped his own child.'
b. *Yo malari anak yo.*
yes AV:run child EMPH
'Yes he kidnapped his child.'

In addition, in Example (52), we see that a 'dative' construction is possible without the use of the applicative. The verb *mambao*, 'bring', is transitive. In this example it is interpreted as ditransitive. The recipient participant *den*, 'me', appears in core argument position even though an applicative has not been used to license a recipient participant. This sentence would be considered ungrammatical in Standard Minangkabau as the applicativised form *mambaoan*, 'bring (to)' would be required to allow this particular syntactic arrangement.

Colloquial Minangkabau

(52) *Sabanta lai Yaya mambao den nasi goreang.*
ONE:moment EMPH TRU:Satria AV:bring 1SG cooked.rice fry
'In a moment Yaya will bring me fried rice.'

---

3. Example (51b) could of course also be an example of a sporadic haplology since [an], the first syllable of *anak*, 'child', has the same phonological form as the applicative.

Similarly in (53), *bao*, 'bring', should be interpreted as a ditransitive verb since as well as licensing the actor *nyo*, 'she', the verb also licenses an undergoer *samba*, 'side dishes', and a benefactive participant *sia*, 'who'. In Standard Minangkabau we would expect the verb to be marked for voice and the valency shown on the verb by the applicative *an*.

Colloquial Minangkabau

(53) *O   kayak kapatang nyo bao   samba   tu   untuak sia?*
FILL like   yesterday 3   bring side.dishes DEM:dist for   who
'So yesterday, who did she bring the side dishes for?'

4.3 Bare verbs: Problems and explanations

Bare verbs create problems for the analysis of the Minangkabau voice system because their very existence implies that active and passive voice marking, and indeed applicative marking, are optional and not required to mark information about semantic roles or argument structure.[4] If voice marking is not a syntactic requirement in Colloquial Minangkabau, then we must ask ourselves what is the function of voice marking in this variety? Bare verb constructions are also problematic because the semantic roles of the participants in the clause are not explicitly marked, raising additional questions about how speakers are able to interpret such clauses.

4.3.1   *What is the function of voice marking in Colloquial Minangkabau?*
An important point to make about the bare verb construction in Colloquial Minangkabau is that it is only the voice markers *maN* and *di* and the applicative *an* which appear to be optional. Minangkabau has a range of other verbal affixes, including locative applicative *-i* and the primarily lexico-semantic affixes *ta*, *ba*, and *pa* (Crouch 2009). These semantic affixes are not optional in Colloquial Minangkabau and the subtleties of meaning they impart are not available if the affix is not used.

For example, consider the verb *aja* in (54). The verb is marked by the active voice marker *maN-* and the reflexive marker *ba-* in (b) and (c) respectively. Even though these two morphological and semantic derivations are available for the verb, only the active interpretation, and not the reflexive interpretation, is available for the bare verb in (a).

---

4. Bare verbs also raise questions about the nature of categoriality in Minangkabau since they can function in both predicative and non-predicative constructions. For further discussion see Crouch (2009).

Standard Minangkabau

(54) a. *Aden mulai aja   baso   Minang duo minggu nan lalu.*
     1SG   start teach language Minang two weeks REL pass
     'I started teaching Minangkabau two weeks ago.'   (Elicitation)

    b. *Aden mulai maaja   baso   Minang duo minggu nan lalu.*
     1SG   start AV:teach language Minang two weeks REL pass
     'I started teaching Minangkabau two weeks ago.'   (Elicitation)

    c. *Aden mulai baraja   baso   Minang duo minggu nan lalu.*
     1SG   start REFL-teach language Minang two weeks REL pass
     'I started studying Minangkabau two weeks ago.'   (Elicitation)

Similarly in (55), the bare verb *baka*, 'burn', only has an active interpretation available. To obtain an involuntary reading of the verb, the involuntary marker *ta* must be used.

Standard Minangkabau

(55) a. *Inyo baka lauak tu.*
     3   burn fish   DEM:dist
     'He roasted the fish.'   (Elicitation)

    b. *Inyo mambaka lauak tu.*
     3   AV:burn fish   DEM:dist
     'He roasted the fish.'   (Elicitation)

    c. *Inyo ta-baka   lauak tu.*
     3   INV-burn fish   DEM:dist
     'He accidentally burned the fish.'   (Elicitation)

These examples show that the lexical/derivational affixes *ba* and *ta* are required to mark semantic information on the verb but the *maN-* prefix is not; a bare verb root can therefore only be assigned active semantics. Additionally, as demonstrated in Section 4.2, when there is no contextual information to determine the semantic roles of participants in a bare verb construction, although either an active or passive 'voice' interpretation is possible, native speakers have a preference for the active interpretation, i.e. assigning the actor role to the pre-verbal NP.

But why is the passive voice marker *di* also optional? The most common place to find a bare passive verb is in a construction where the actor participant is marked by the preposition *dek*. Because there is no ambiguity as to semantic roles in this case, perhaps the *di-* marking on the verb is redundant. One would think then, that passive and active voice marking is only obligatory when semantic roles require disambiguation. Yet, as the examples in Section 4.2 show, even in ambiguous cases voice marking is not a requirement.

So do bare verbs in Colloquial Minangkabau *mean* the same as voice marked verbs? As we have seen, speakers are generally able to assign semantic roles to

participants in a clause with a bare verb, even bare verbs with no applicatives, which suggests that the role of voice marking in Colloquial Minangkabau is not syntactic. In Standard Minangkabau, *maN-* has a clearly delineated semantic function and *di-* also has some associated semantic and aspectual properties as a result of its syntactic role in individuating the undergoer participant (Crouch 2009). Thus, in Colloquial Minangkabau, the opposition between bare verbs, *maN-*, and *di-* must be primarily motivated by these semantic and conceptual factors in a similar way to the opposition between the active and passive voice markers *N-* and *di-* in Riau Indonesian. In Colloquial Minangkabau, the lack of voice marking on bare verbs has no syntactic consequences, whereas verbs that are marked for voice show that the actor or undergoer is particularly conceptually salient, or show that the event has certain aspectual properties.

4.3.2   *How are semantic roles determined?*

Discourse pragmatics provide the clues that allow hearers to disambiguate the semantic roles of a bare verb's arguments in Colloquial Minangkabau. If one of the arguments in the clause refers to a participant in the conversation, is the protagonist of the narrative, has had more previous mentions than another participant, or if one of the participants is definite and the other is not, then that participant will be more referentially salient than the other. Establishing which is an actor and which is an undergoer is more difficult. In some bare verb constructions the actor is marked by *dek*, in others it is not. In these cases, contextual background provides the necessary information.

As Examples (37) to (44) demonstrate, either participant in a bare verb construction can be assigned the actor or undergoer role, and either of the participant NPs can appear before or after the verb. Only knowledge of the context can provide the clues to work out the intended semantic roles of the bare verb's arguments. But perhaps in these cases the speaker does not *intend* anything, therefore the hearer has no need to disambiguate in the first place. Perhaps bare verb constructions do not represent some kind of defective or underspecified version of a voice marked verb construction. Instead, they are used in cases where voice and semantic roles are not just underspecified, but unimportant. This is because ambiguity is a perfectly acceptable feature of Colloquial Minangkabau discourse. Gil (2002, 2006) explores similar ideas in his discussion of Riau Indonesian's voice system. He argues that no morphosyntactic devices exist in the language to disambiguate semantic roles and that semantic roles are therefore typically vague and underspecified. Gil's (2001, 2007) theory of 'Associational Semantics' may explain why semantic roles can be underspecified not only in Riau Indonesian, but in Colloquial Minangkabau as well.

4.3.3 *Associational semantics*

Gil (1994, 2000, 2001, 2007, 2008) argues that Riau Indonesian is a monocategorial language with some isolating morphology. He also suggests that it is a typologically 'simplistic' language. Monocategoriality and isolating morphology also mean that constructions in this language are often underspecified. For this reason, Gil (2007) proposes that Riau Indonesian is a highly 'Associational' language, i.e. the interpretability of underspecified utterances in this language relies on Associational Semantics. Gil (2008) argues that Minangkabau, as well as Mentawai and Sundanese, are also Associational languages. It is these typological traits, Gil argues, that form the basis of the Sundic-type voice system (see Section 4.1).

According to Gil (2001, 2007), if a language is Associational there will be a large number of available interpretations for any utterance that is underspecified. The interpretations rely on the fact that the entities and events referred to in the utterance are semantically associated with each other in a variety of possible ways. For example, consider the Riau Indonesian sentence in (56) which is highly underspecified. Not only is *makan*, 'eat', not specified for tense and aspect, but *ayam* 'chicken' is not marked for definiteness. The semantic role of *ayam* 'chicken' is also unspecified since there is no voice marking on the verb. This means that *ayam* could be interpreted as an agent, in which case the interpretation of the sentence would be 'the chicken is eating'. However, *ayam* could also be a patient ('someone is eating the chicken'), a benefactor ('someone is eating for the chicken'), or even a comitative participant ('someone is eating with the chicken'). The ontological type of the utterance is also unspecified therefore we cannot be sure if *ayam makan* is actually referring to an event. Some of the other available interpretations of ontological type might include an entity ('the chicken that is eating'; 'chicken food'), a location ('where the chicken is eating'), or indeed a time ('when the chicken is eating') (Gil 2007).

(56) *Ayam   makan.*
     chicken eat
     'The chicken is eating… etc.'　　　　　　(Gil 2007: 73 – Riau Indonesian)

Gil (2007, 2008) has devised an experiment to test the Associationality of a range of languages, first to establish what the typological restrictions on Associational Semantics are, and second, to find out whether Associationality is a salient feature of grammatically 'simplistic' languages like Riau Indonesian as well as creole languages in general. In the experiment, subjects are presented with a sentence in their language and two pictures. Subjects are then asked to evaluate whether the situation entailed by the test sentence is accurately portrayed in any of the pictures. Subjects may choose only one of the pictures, both pictures or neither of the pictures. One of the pictures shows the test interpretation. If subjects choose this picture, then

this demonstrates that an Associational interpretation is available in their language. A sample of some of the test sentences from Gil's (2007) Associationality experiment with Minangkabau speakers and English speakers can be found in Table 2.

The 'Bare Peripheral' sentences show the juxtaposition of an event and an entity which do not belong in the same semantic frame. Gil argues that, in an Associational language, an interpretation can be made from the juxtaposed elements that fits the situation portrayed in the test picture, whereas in a non-Associational language, the interpretation will not fit. Consider sentence (1) in Table 2. In an Associational language, the juxtaposed elements CLOWN DRINK BOOK refer to an event which is associated with a clown, drinking and a book. Therefore, the interpretation represented in the test picture, *clown drinking while reading book*, will be available. However, in a non-Associational language the interpretation represented in the test picture will not be available since BOOK will have to be interpreted as part of the argument structure, or semantic frame, of DRINK, which is semantically odd (Gil 2007).

Table 2. Test Sentences from the Associationality Experiment (Gil 2007: 96, 105)

A. BARE PERIPHERAL SENTENCES:

MINANGKABAU TEST SENTENCES	ENGLISH TEST SENTENCES	TEST PICTURE SHOWS:
1. *Badut minum buku* clown drink book	'The clown is drinking the book'	*Clown drinking while reading book*
2. *Badut bali sanang* clown buy happy	'The clown is buying happiness'	*Clown buying fruit with happy face*
3. *Kopi galak* coffee laugh	'The coffee is laughing'	*Person spilling coffee, onlooker laughing*
4. *Pitih sanang* money happy	'The money is happy'	*Man holding money with happy face*

B. BARE PATIENT PRECEDING SENTENCES:

MINANGKABAU TEST SENTENCES	ENGLISH TEST SENTENCES	TEST PICTURE SHOWS:
5. *Buruang makan* bird eat	'The bird is eating'	*Cat eating a bird*
6. *Harimau takuik* tiger afraid	'The tigers are afraid'	*People fearing tigers*
7. *Mancik kaja kuciang* mouse chase cat	'The mouse is chasing the cat'	*Cat chasing mouse*
8. *Oto tundo padusi* car push woman	'The car is pushing the woman'	*Woman pushing car*

The 'Bare Patient Preceding' sentences contain two elements: an event preceded by an entity which is understood to be what Gil terms the 'patient', but what I refer to as the 'undergoer'. In an SVO language, the preceding entity will be interpreted

as the agent/actor therefore the situation represented in the test picture will not be available. However, in an Associational language the test interpretation will be available since all the sentence entails is an 'event X associated with entity Y' (Gil 2007). I was able to confirm with my language consultants that the test sentences in Table 2 were grammatical and while there was some variation as to which Associational interpretations were available, my consultants consistently permitted Associational interpretations for sentences 7 and 8. These are bare patient preceding sentences containing two participants, which implies, context permitting, that the roles of agent and patient (or actor and undergoer) are able to be assigned to either participant when there is no morphological marking.

Table 3. Results for Minangkabau and English. (Gil 2007: 86)

LANGUAGE	NUMBER OF SUBJECTS	% AVAILABLE ASSOCIATIONAL INTERPRETATION:	
		BARE PERIPHERAL	BARE PATIENT PRECEDING
MINANGKABAU	30	74	57
ENGLISH	32	7	4

The results of Gil's (2007) experiment (see Table 3) show that the percentage of available Associational interpretations of the test sentences was much higher for Minangkabau than it was for English. These findings suggest that underspecified utterances in Colloquial Minangkabau are interpretable because the compositional Associationality of the language allows for a range of possible meanings to be drawn from the individual elements in the utterance. The Associational nature of Minangkabau thus in part accounts for the presence of bare verbs in the colloquial register; morphological underspecification for voice is acceptable since a number of possible interpretations are available regardless.

5. Conclusion

As a result of examining the linguistic structures revealed in the MPI EVA corpus of naturalistic Minangkabau data, I conclude that Standard Minangkabau and Colloquial Minangkabau differ not only in terms of their function and use, but also in terms of their form. The two varieties are so distinct that they also require different typological characterisations in their voice systems: Standard Minangkabau is an Indonesian-type language whereas Colloquial Minangkabau is a Sundic-type language. Colloquial Minangkabau is rife with underspecification and indeterminacy due to the conceptual nature of 'voice' in this variety as well as its flexible word order constraints and the pervasive use of bare verbs. It was suggested that

Colloquial Minangkabau can thus be described as an 'Associational' language, meaning that the language allows for a wide range of possible interpretations of an utterance. This is in keeping with the typological characterisation of Colloquial Minangkabau as a Sundic-type language since 'Associationality' appears to be a feature of isolating Austronesian languages and Sundic-type languages more generally (cf. Gil 2001, 2007, 2008).

Acknowledgements

References

Adelaar, K. Alexander. 1992a. *Proto-Malayic: The Reconstruction of its Phonology and Parts of its Lexicon and Morphology*. Canberra: Pacific Linguistics.

Adelaar, K. Alexander. 1992b. The relevance of Salako for Proto-Malayic and for Old Malay epigraphy. *Bijdragen tot de Taal-, Land- en Volkenkunde* 148: 381–408. https://doi.org/10.1163/22134379-90003143

Adelaar, K. Alexander. 1995. Minangkabau. In *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies, Part 1: Fascile 1*, Darrell T. Tyron, Malcolm D. Ross, Charles E. Grimes, Adrian Clynes & K. Alexander Adelaar (eds), 433–441. Berlin: Mouton de Gruyter.

Adelaar, K. Alexander. 2005. Much ado about *di-*. *Bijdragen tot de Taal-, Land- en Volkenkunde* 161: 127–142.

Adelaar, K. Alexander. 2008. Yet Another Explanation of the Passive Prefix *di-*. Paper presented at the *Twelfth International Symposium on Malay/Indonesian Linguistics*, 26th-27th June, Leiden, The Netherlands.

Adnani, Dahnil. 1971. Minangkabau Sentences and Their Constituents: An Exploration in Semantically Based Grammatical Analysis. PhD dissertation, Georgetown University.

Ansjar, Mohammad. 1971. An Introduction to Minangkabau Morphology. MA thesis, IKIP Malang.

Anwar, Khaidir. 1992. *Semantik Bahasa Minang*. Padang: Yayasan Pengkajian Kebudayaan Minang.

Arifin, Syamsir et al. 1979. *Kata Tugas Bahasa Minangkabau* (Minangkabau Function Words). Padang: Pusat Pembihaan dan Pengembangan Bahasa – Departemen Pendidikan dan Kebudayaan.

Arka, I. Wayan. 2003. Voice systems in the Austronesian languages of Nusantara: Typology, symmetricality and undergoer orientation. *Linguistik Indonesia* 21: 113–139.

Arka, I. Wayan & Manning, Christopher D. 2008. Voice and grammatical relations in Indonesian: A new perspective. In *Voice and Grammatical Relations in Austronesian Languages*, Peter Austin & Simon Musgrave (eds), 45–69. Stanford CA: CSLI.

Be, Kim Hoa Nio. 1978/1979. *Morfologi dan Sintaksis Bahasa Minangkabau* (Minangkabau Morphology and Syntax). Padang: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan Padang, IKIP Padang.

Be, Kim Hoa Nio. 1984. *Sistem Morfologi Kata Benda dan Kata Sifat Bahasa Minangkabau* (The Morphology of Nouns and Adjectives in Minangkabau). Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Benjamin, Geoffrey. 1993. Grammar and polity: The cultural and political background to Standard Malay. In *The Role of Theory in Language Description*, William Foley (ed.), 341–392. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110872835.341

Biber, Douglas. 2000. Investigating language use through corpus-based analyses of association patterns. In *Usage Based Models of Language*, Michael Barlow & Suzanne Kemmer (eds), 287–313. Stanford CA: CSLI.

Biber, Douglas, Conrad, Susan & Reppen, Randi. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511804489

Blust, Robert. 2013. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Brodkin, Dan & Fortin, Catherine. 2017. Minangkabau -i: A locative, transitivizing, iterative, adversative suffix. *Proceedings of the Linguistic Society of America* 42(2): 1–12.

Bybee, Joan & Hopper, Paul. 2001. Introduction to frequency and the emergence of linguistic structure. In *Frequency and the Emergence of Linguistic Structure*, Joan Bybee & Paul Hopper (eds), 1–24. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.01byb

Chung, Sandra. 1976. On the subject of two passives in Indonesian. In *Subject and Topic*, Charles N. Li (ed.), 57–98. New York NY: Academic Press.

Cole, Peter, Hermon, Gabriella & Tjung, Yassir. 2006. Is there *pasif semu* in Jakartan Indonesian? *Oceanic Linguistics* 45: 64–90. https://doi.org/10.1353/ol.2006.0009

Cole, Peter, Gabriella Hermon & Yanti. 2007. Voice in Malay/Indonesian. *Lingua* 118: 15–53.

Conners, Thomas. 2008. Tengger Javanese Morphosyntax. PhD dissertation, Yale University.

Conners, Thomas J. & Brugman, Claudia. 2013. The event semantic role of the nasal prefix in Jakarta Indonesian. In *Proceedings of the Twelfth International Conference on Austronesian Linguistics*, I. Wayan Arka and N. L. K. Mas Indrawati (eds), 213–225. Canberra: Pacific Linguistics.

Conners, Thomas, Bowden, John & Gil, David. 2015. Valency classes in Jakarta Indonesian. In *Valency Classes in the World's Languages*, Andrej Malchukov & Bernard Comrie (eds), 941–986. Berlin: Mouton De Gruyter.

Crouch, Sophie. 2009. Voice and Verb Morphology in Minangkabau, a Language of West Sumatra, Indonesia. MA thesis, University of Western Australia.

Dardjowidjojo, Soenjono. 1978. *The Sentence Patterns of Indonesian*. Honolulu HI: University of Hawai'i Press.

Davies, William D. 2005. The richness of Madurese voice. In *The Many Faces of Austronesian Voice Systems*, I. Wayan Arka & Malcolm D. Ross (eds), 197–220. Canberra: Pacific Linguistics.

Drakard, Jane. 1999. *A Kingdom of Words: Language and Power in Sumatra*. Oxford: OUP.

Fadlul Rahman, Fitri, Kurniati, Santi, Prima Putri, Yessy & Gil, David. 2013. Word-internal language mixing: Borrowing, code-switching or register-switching? Paper presented at *Seventeenth International Symposium on Malay/Indonesian Linguistics*, 8th-9th June, Padang, Sumatra Barat, Indonesia.

Foley, William & Van Valin , Jr., Robert. 1984. *Functional Syntax and Universal Grammar*. Cambridge: CUP.

Foley, William & Van Valin , Jr., Robert. 1985. Information packaging in the clause. In *Language Typology and Syntactic Description*, Timothy Shopen (ed.), 282–364. Cambridge: CUP.

Fortin, Catherine R. 2001. A Descriptive Study of Minangkabau, with an Emphasis on Verbal Morphology. MA thesis, University of Pittsburgh.

Fortin, Catherine R. 2004. Syntactic and semantic valence: Morphosyntactic evidence from Minangkabau. *BLS* 29: 137–147. https://doi.org/10.3765/bls.v29i1.1019

Francis, Gil. 1993. A corpus-driven approach to grammar: Principles, methods and examples. In *Corpus Linguistics: Readings in a Widening Discipline*, Geoffrey Sampson & Diana McCarthy (eds), 198–211. London: Continuum.

Genta, Sakti. 1999. Turn-taking in the Minangkabau Language: A Case Study of a Television Chat Show. MA thesis, La Trobe University.

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17: 179–200. https://doi.org/10.1017/S0332586500003000

Gil, David. 2000. Syntactic categories, cross-linguistic variation and Universal Grammar. In *Approaches to the Typology of Word Classes*, Petra M. Vogel & Bernard Comrie (eds), 173–216. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110806120.173

Gil, David. 2001. Creoles, complexity and Riau Indonesian. *Linguistic Typology* 5: 325–371.

Gil, David. 2002. The prefixes *di-* and *N-* in Malay/Indonesian dialects. In *The History and Typology of Western Austronesian Voice Systems*, Fay Wouk & Malcolm D. Ross (eds), 241–283. Canberra: Pacific Linguistics.

Gil, David. 2006. Intonation and thematic roles in Indonesian. In *Topic and Focus, Cross Linguistic Perspectives on Meaning and Intonation*, Chungmin M. Lee, Matthew Gordon & Daniel Buring (eds), 41–68. Dordrecht: Springer.

Gil, David. 2007. Creoles, complexity and associational semantics. In *Deconstructing Creole: New Horizons in Language Creation* [Typological Studies in Language 73], Umberto Ansaldo & Stephen J. Matthews (eds), 67–108. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.73.06gil

Gil, David. 2008. So where are all those Indonesian-type voice systems? An exploration of clause structure in Mentawai. Paper presented at *International Workshop on Minority Languages in the Malay/Indonesian Speaking World*, 28th June, Leiden, The Netherlands.

Gil, David. 2015a. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia, The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: Mouton de Gruyter.

Gil, David. 2015b. *Malayic languages of Sumatra*. A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and Universitas Bung Hatta, Padang.

Ginsberg, Elisabeth & Paauw, Scott. 2010. Verbal aspect in Colloquial Jakarta Indonesian. Paper presented at *ISLOJ 2*: Lombok, Indonesia.

Gordon , Jr., Raymond G. (ed.). 2005. *Ethnologue: Languages of the World*, 15th edn. Dallas TX: SIL International.

Hardjadibrata, R. Rabinranat. 1985. *Sundanese: A Syntactical Analysis*. Canberra: Pacific Linguistics.

Hidayat, Lanny. 2011. The analysis of the semantic function of the prefix *N-* in Basilectal Jakarta Indonesian. In *Proceedings of the Workshop on Tense, Aspect, Mood and Evidentiality in Indonesian Languages*, Asako Shiohara, Anthony Jukes, Atsuko Utsumi, Antonia Soriente (eds), 173–182. Tokyo: ILCAA.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological overview. In *The Austronesian Languages of Asia and Madagascar*, Karl Alexander Adelaar & Nikolaus P. Himmelmann (eds), 110–181. London: Routledge.

Himmelmann, Nikolaus P. 2008. Lexical categories and voice in Tagalog. In *Voice and Grammatical Relations in Austronesian Languages*, Peter K. Austin & Simon Musgrave (eds), 247–293. Stanford CA: CSLI.

Keller, Frank. 1999. Review of *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*, Carson T. Schütze. *Journal of Logic, Language and Information* 8: 114–121. https://doi.org/10.1023/A:1008323501411

Klamer, Marian. 1996. *Voice in Austronesian* [NUSA 39]. Jakarta: Universitas Atma Jaya.

Medan, Tamsin. n.d. *Dialek-dialek Minangkabau di Daerah Minangkabau/Sumatera Barat (suatu pemerian dialektologis)*. (Minangkabau Dialects in the Minangkabau Region/West Sumatra (a dialectological description)). Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Mosel, Ulrike. 2006. Fieldwork and community language work. In *Essentials of Language Documentation*, Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds), 67–86. Berlin: Mouton de Gruyter.

Moussay, Gérard. 1995. *Dictionnaire Minangkabau, Indonesien, Français*. Paris: Harmattan, Association Archipel.

Moussay, Gérard. 1998. *Tata Bahasa Minangkabau* (Minangkabau Grammar). Jakarta: Kepustakaan Populer Gramedia.

Musgrave, Simon. 2000. Pronouns and morphology: Undergoer subject clauses in Indonesian. In *Yearbook of Morphology*, Geert Booij & Jaap van Marle (eds), 155–186. Amsterdam: Kluwer.

Nadra. 2006. *Rekonstruksi Bahasa Minangkabau* (A Reconstruction of Minangkabau). Padang: Andalas University Press.

Nadra, Reniwati & Yades, Efri. 2008. Daerah asal dan arah migrasi orang Minangkabau di Provinsi Jambi berdasarkan kajian variasi dialectal. (The origin and direction of the migration of Minangkabau people in Jambi Province based on a study of dialectal variation). *Makara, Sosial Humanoria* 12: 1–8.

Nerbonne, John (ed.). 1998. *Linguistic Databases*. Stanford CA: CSLI.

Noviatri, Reniwati & Asnan, Gusti. 2017. Affixes of Minangkabau in the origin and Rantau area: Study of morphological variation. *Jurnal Arbitrer* (The Arbitrary Journal) 4: 86–92. https://doi.org/10.25077/ar.4.2.86-92.2017

Puspawati. 1997. *Aspek Semantis Modalitas Bahasa Minangkabau: Suatu Kajian Deskriptif: Laporan Penelitian* (Semantic Aspects of Modality in Minangkabau: A Descriptive Study: Research Results). Padang: Lembaga Penelitian dan Pengabdian Masyarakat.

Ramadian. 1992. *Adverbia Penanda Modalitas Bahasa Minangkabau* (Adverbial Modality Markers in Minangkabau). Padang: Angkasa Raya.

Ross, Malcolm D. 2002. The history and transitivity of Western Austronesian voice and voice-marking. In *The History and Typology of Western Austronesian Voice Systems*, Fay Wouk & Malcolm D. Ross (eds), 17–63. Canberra: Pacific Linguistics.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Chicago IL: University of Chicago Press.

Shibatani, Masoyoshi. 2006. On the conceptual framework for voice phenomena. *Linguistics* 44: 217–269. https://doi.org/10.1515/LING.2006.009

Sneddon, James Neil. 1996. *Indonesian: A Comprehensive Grammar*. London: Routledge.

Tanner, N. 1972. Minangkabau Disputes. PhD dissertation, University of California.

Thompson, Sandra & Hopper, Paul J. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. In *Frequency and the Emergence of Linguistic Structure* [Typological Studies in Language 45], Joan Bybee & Paul Hopper (eds), 27–60. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.03tho

van der Toorn, Johannes L. 1899. *Minangkabausche Spraakkunst*. S-Gravenhage: Martinus Nijhoff. https://doi.org/10.1007/978-94-011-9312-2

Voorhoeve, Petrus. 1955. *Critical Survey of Studies on the Language of Sumatra*. ʻs-Gravenhage: Martinus Nijhoff. https://doi.org/10.1007/978-94-015-0522-2

Williams, Gerald E. 1961. Colloquial Minangkabau: A Description of Phonological and Morphological Structure. PhD dissertation, University of Chicago.

Wouk, Fay. 1989. The Impact of Discourse on Grammar: Verb Morphology in Spoken Jakarta Indonesian. PhD dissertation, University of California.

Wouk, Fay. 1999. Sasak is different: A discourse perspective on voice. *Oceanic Linguistics* 38: 91–114. https://doi.org/10.2307/3623394

Wouk, Fay. 2004. Transitive bare verb stems in Spoken Jakarta Indonesian. *Nusa* 53: 29–46.

CHAPTER 5

Javanese undressed
'Peripheral' dialects in typological perspective

Thomas J. Conners
University of Maryland

This chapter makes two claims about Javanese, one concerning its internal
dialect variation, and one concerning its place in mainland Southeast Asian
(MSEA) typology. First, Javanese exhibits extreme dialect variation, with many
features of these variants not appearing in descriptions of Javanese, which mostly
concern the Central variety. Second, the existence of these features changes
the position of Javanese in the continuum of isolating-to-synthetic languages.
Relevant features from six dialects of Javanese show that the Central variety –
that of Yogyakarta and Solo – inadequately characterises Javanese as a whole;
rather, the geographically and socially 'peripheral' dialects more strongly tend
toward isolating morphology. Consequently, Javanese is less of an outlier in the
MSEA Sprachbund than is generally acknowledged. Historical evidence shows
that the Central variety is innovative with respect to Javanese overall.

Keywords: Javanese, internal dialect variation, typology, peripheral dialects,
isolating morphology

1. Introduction

Javanese is the giant elephant in the room at the Austronesian party, often presented
as a monolithic language with a funny speech level system and elaborate verbal
morphology that has already been thoroughly studied, documented, and catego-
rised typologically and as such need not be discussed further. Yet it is surprising
how little we know about the world's eleventh largest language,[1] and how far off
base the truth is from what we presume about it.

---

[1] The CIA World Factbook (2016/2017) lists Javanese as the eleventh largest language in terms
of native speakers; *Ethnologue* (2018) lists it as the tenth largest. It is the majority language spoken
in the provinces of Central Java and East Java, and the Special District of Yogyakarta. It is the
largest minority language spoken in Banten, West Java, and Lampung provinces

This paper makes two observations about Javanese, one with respect to its internal dialect variation, and one with respect to its place in mainland Southeast Asian (MSEA) typology. First, there is a great deal of variation in Javanese, with many dialects sharing a number of features that are not recognised in common descriptions of Javanese (e.g. Keeler 1984 and Uhlenbeck 1983). Second, when viewed in the context of these dialects, Javanese is less of an outlier in the MSEA typological Sprachbund than is generally described.

The first goal of this paper therefore is to demonstrate that 'Javanese', contrary to its common descriptions, encompasses a range of generally under-described varieties and variation, the extremes of which are not mutually intelligible. I argue here for a first order, categorical distinction between a Central Javanese variety which is the one most often described or presupposed in documentation and other descriptive materials, and what I call 'peripheral' varieties. In this paper I take five exemplars of these peripheral varieties – varieties chosen simply to demonstrate how they differ, each, and taken collectively, from the Central variety, and how, on many parameters, they resemble one another. The existence of these features changes the position of Javanese in the continuum of isolating-to-synthetic languages. The geographically and socially peripheral dialects more strongly tend toward isolating morphology. Consequently, Javanese is less of an outlier in the MSEA Sprachbund than generally acknowledged.

1.1 Javanese variation

In this section I review some of the features of the collection I am calling 'peripheral' varieties which differentiate these varieties from the Central one.

Examples (1a–f) demonstrate some significant lexical differences that can be found among the five peripheral varieties – Banten, Banyumasan, Pesisir Lor, Tengger, and Osing, in contrast with one another as well as with Central Javanese:[2]

(1) a. *Aku arep teko karo kanca-ne.*   Central
     b. *Kite pen teke karo batur-e.*   Banten
     c. *Nyong pang teka karo batir-e.*   Banyumasan
     d. *Inyong pan teka karo kanca-ne.*   Pesisir Lor
     e. *Eyang kate teka karo rewang-e.*   Tengger

---

2. Examples in this work make use of glosses given in the Leipzig Glossing conventions. Additional glosses are: AGFOC 'agent focus', APPL1 'applicative marker 1'; APPL2 'applicative marker 2', ASS 'associative marker', PATFOC 'patient focus'.

    f.   *Isun kate teko ambe konco-ne.*   Osing
        1SG want come with friend-ASS[3]
        'I am going to come with a friend.'

Lexical variation exemplifies significant dialect differences and will be addressed in much more detail as relevant to the discussion of diachrony in Section 4.3. The typological implications of intra-Javanese variation lie in their morphological and morphosyntactic properties.

Peripheral Javanese dialects also differ substantially in their phonological, morphological, morphosyntactic, and pragmatic properties. Such phenomena that have often been described for Javanese as a whole in fact refer largely to two dialects, those of the 'exemplary centres' Yogyakarta and Solo (Surakarta) (Errington 1985, 1988). They represent the Central variety, hereafter referred to as Central Javanese or CJ. However, many of these distinctive 'Javanese' features, such as the speech level system and the fully articulated verbal morphosyntax, are found only in those Central dialects and do not characterise Javanese writ large. Moreover, there is greater similarity among geographically discontiguous dialects outside of these centrally located 'exemplary centres' than between them and the Central varieties. The majority of Javanese varieties have never been documented or described, though some recent works on these varieties are referenced below.

I propose that the Central dialect is more innovative with respect to many changes than the collection of peripheral dialects, and I argue for this in terms of phonological, morphological, and pragmatic features.

A straightforward example is the considerable dialect variation in the verbal paradigm. Central Javanese has a complex set of verbal affixes that encode, among other functions, person, voice, valency, and mood, resulting in up to 44 forms for a single verb, as exemplified in Table 1.[4]

Although Austronesian languages are rightly known for their complex voice and applicative morphology, this level of complexity in the verbal system is exceptional

---

3. *-ne* 'ASS' is an associative marker, which functions to relate the marked word to some other element either prominent in the same clause or salient in the discourse. The translation is an artifact of the metalanguage.

4. In Table 1, the first column lists the four moods encoded on the verb in CJ and the second column shows the voice form for each mood. In the case of the indicative mood, there are two types of patient focus construction, with PatFoc II being an accidental or adversative passive. The various mood and voice combinations are signalled through a range of prefixes, suffixes or bare marking on the root. These are shown in the third and fourth columns. The label *neutral* in the fourth column is used to contrast these suffixes with the two applicative suffixes in fifth and sixth columns. The neutral suffixes are not considered part of the applicative system and are used to indicate basic mood and voice distinctions.

Table 1. Verbal morphology in Central Javanese

Mood	Voice	Prefix	Neutral	Applicative I	Applicative II
Indicative	Agent Focus	N-	Ø	-i	-aké
	Patient Focus I	tak-/kok-/di-			
	Patient Focus II	ke-		-an	Ø
Imperative	Agent Focus	N-	-a	-ana	-na
	Patient Focus I	Ø	-en		
Propositive	Agent Focus	(aku) tak N-	Ø	-i	-aké
	Patient Focus I	tak-	-é	-ané	-né
Subjunctive	Agent Focus	N-	-a	-ana	-na
	Patient Focus I	tak-/kok-/di-	-en	-na	

among the languages of Indonesia. However, most dialects of Javanese show no-where near this level of complexity. Compare for example the CJ verbal system with the Tengger verbal system in Table 4. I argue in Section 4.3, based on historical and comparative evidence, that it is the Central variety (Vander Klok & Conners 2019) which has undergone a significant degree of innovation – hence, differentiation from the peripheral varieties. Further, based on synchronic evidence, I argue that the peripheral dialects collectively are more exemplary of Javanese on the whole.

1.2   Javanese dialects in typological perspective in the MSEA Sprachbund

Having established that there is a linguistically based first-order distinction between Central and peripheral varieties of Javanese, I will consider the features on which this distinction is based to review the position of Javanese within the larger typology of mainland Southeast Asian languages, specifically with respect to its isolating morphosyntactic character.

The MSEA linguistic area is a Sprachbund which includes languages from five families spoken in mainland Southeast Asia (Austroasiatic, Hmong-Mien, Tai-Kadai, Tibeto-Burman, and Austronesian). This area is characterised by many shared features including, saliently, the following:[5]

– sentence final particles (Dryer 2011)
– prominence of pronoun imposters (Conners, Brugman & Adams 2016), i.e. open class anaphors/pronoun avoidance (Flannery 2009)
– grammatical reduplication (Goddard 2005)
– SVO word order (Benedict 1994)

---

5.   For a summary of general features see Comrie (2009) and Enfield & Comrie (2015).

– collapse of morphosyntactic distinction between types of nominal attribution (Gil 2011)
– numeral classifiers (Aikhenvald 2000; Senft 2008)
– tone (Maddieson 2011)
– limited inflectional and derivational morphology (Goddard 2005)
– periphrastic constructions to express voice and valency possibilities (Yap Iwasaki 2007)

The final two features are the most salient and pervasive features defining the MSEA linguistic area as a region of isolating languages.

Javanese is typically characterised as sharing the first five of these features, but lacking the final four (for example, no variety of Javanese has tone, but all varieties have sentence-final particles and grammatical reduplication).

Because it lacks the last four features, Javanese is not traditionally included within the MSEA linguistic area, and, because it lacks the last two, it is not included in the group of isolating languages.[6]

Contrary to this common notion, I argue that Javanese is better represented by the characterisation of the peripheral dialects, which are much more isolating in their morphology and morphosyntax than the better-described Central variety. Javanese dialects, therefore, fall on different points on a typological cline, with most varieties of Javanese closer to the isolating end, though not as extreme as some other MSEA languages.[7] The peripheral varieties show that Javanese shares more critical features with the isolating MSEA Sprachbund than is commonly acknowledged.

This line of argument further allows Javanese, when viewed from the peripheral dialects, to be placed within the larger Mekong-Mamberamo (MM) linguistic area (Gil 2015). This area includes the MSEA languages described above and extends to include most of the languages of Indonesia and the island of New Guinea. Gil characterises this area by the presence of the following 17 features:

1. passing gesture
2. repeated dental clicks expressing amazement
3. conventionalised greeting with 'where'
4. 'eye day' > 'sun' lexicalisation

---

6. Many authors refer to the 'mainland' Southeast Asian linguistic area, to the exclusion of insular languages such as Javanese. See for example Campbell (2017), Enfield (2005), Bisang (1996), Siebenhütter (2018), and Migliazza (1996).

7. As noted by Gil ('What Does It Mean to Be an Isolating Language', this volume), many varieties of Malay/Indonesian are more isolating than many MSEA languages, not least because they have comparatively little compounding compared to most MSEA languages. Javanese varieties similarly have little compounding, though that feature is not further addressed in the current paper.

5. d/t place-of-articulation asymmetry
6. numeral classifiers
7. verby adjectives
8. basic SVO word order
9. iamitive perfects
10. 'give' causatives
11. low differentiation of adnominal attributive constructions
12. weakly developed grammatical voice
13. isolating word structure
14. short words
15. low grammatical-morpheme density
16. optional thematic-role flagging
17. optional tense-aspect-mood marking

The perspective that the 'peripheral' varieties provides is that Javanese shares critical features with the MSEA Sprachbund.

Siebenhütter (2018) provides a map, reproduced here in Figure 1, of the MSEA linguistic area. Of note here are three distinct regions. The first region, which largely comprises languages spoken in Thailand, Laos, Vietnam and Cambodia, encompassing members from five distinct language families, represents the core
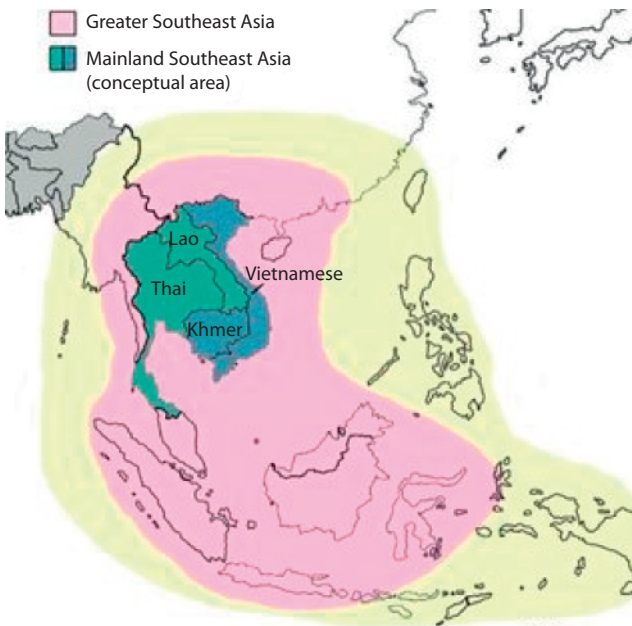
Figure 1. Southeast Asian linguistic area, from Siebenhütter (2018)

MSEA area. The second region includes the languages of southern China, Hainan, Malaysia, and most of western Indonesia. This 'Greater Southeast Asia' area encompasses the Javanese varieties under discussion here. A yet larger area, geographically defined, includes the Philippines, Taiwan, and the languages of eastern Indonesia, Papua, and Timor. Many of the languages in this broader area are included in Gil's MM linguistic grouping.

The remainder of this paper is structured as follows. In Section 2, I discuss the data collection methods. In Section 3, the central and the peripheral dialects are described in more detail. The descriptions given in Section 3 include those features which differentiate Central Javanese from the peripheral dialects, and which help to better place the peripheral dialects within the broader MSEA linguistic area.

The features discussed can be clustered into those that are relevant for Javanese dialectology and those that are relevant for regional typology. The former set of features includes lexical (borrowed and innovative vocabulary fields), phonological (vowel harmony and vowel raising) and pragmatic features (specifically, the existence and complexity of the speech level system). The latter set includes periphrastic passive and causative constructions, more extensive use of clitics as opposed to affixes, and complexity of verbal morphosyntax.

The relevant facts having been given in Section 3, section 4 provides three key arguments. First, I discuss how the data support the first order distinction between CJ and the peripheral dialects. Second, I discuss the implications of this revised view for the position of Javanese in both the MSEA and MM linguistic areas. Finally, I present a sample of lexical evidence to further support the claim that CJ is in fact a very innovative dialect, vis-à-vis the peripheral dialects, providing at least a partial explanation for the current situation. Section 5 summarises the arguments and concludes the chapter.

2. Fieldwork, data collection, method

The evidence presented here is based on original fieldwork and comes from a geographically widespread range of Javanese dialects: Banten, Banyumasan, Tengger, Pesisir Lor, and Osing. The named varieties extend from the extreme west to the extreme east of the island of Java.

The fieldwork was conducted over many years. The Tengger data was collected during an extended stay in Ngadas village in 2002–2003, and some shorter subsequent trips. The Osing data was collected in 2006 on a field trip to Banyuwangi; the Banten data was collected on two field trips to Cilegon in 2006–2007. The Banyumasan data was collected on a number of field trips between 2004 and 2011, and the Pesisir Lor data was collected on an extended field trip in 2006. For all

data other than Tengger and Banyumasan, research assistants helped with the data collection and transcription. The data collected in all cases was primarily spontaneous naturalistic speech, which was transcribed and entered into a corpus using Filemaker Pro. Additional elicitation sessions were held both in the field and at the Jakarta Field Station of the Max Planck Institute for Evolutionary Anthropology.[8]

Figure 2 shows a map of Java and Madura. The main languages (other than Jakarta Indonesian) are labelled. The dialects of Javanese discussed in the present paper are shaded across their approximate area of use.
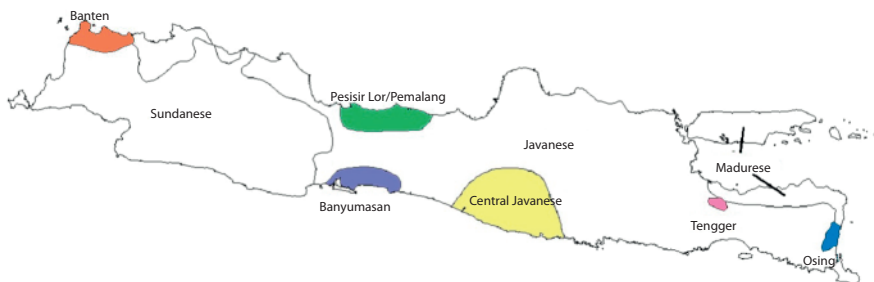
Figure 2. Map of Javanese dialects mentioned (Source: Author)

Until very recently, most work on Javanese has focused on the Central dialects, as spoken in and around the sultans' palaces in Yogyakarta and Surakarta/Solo. These are the 'exemplary centres' as conceived by Errington (1985, 1998). Almost all Dutch colonial work focused on these varieties (see e.g. Uhlenbeck 1978). Even today almost all grammatical descriptions, published dictionaries, and pedagogical materials focus on the Central dialect (Robson 1992; Robson & Wibisono 2002; Horne 1961, 1974; Ogloblin 2005, among others). Some recent exceptions include work on Cirebon Javanese (Ewing 2005); Paciran Javanese (Vander Klok 2012, 2015, 2018); Peranakan Javanese (Cole et al. 2000, 2003a, 2003b, 2008, and others);[9] Tengger Javanese (Conners 2008), Banyumasan (Conners & Brugman 2013), Osing (Wittke 2019), Surinamese Javanese (Villerius 2017, 2019), and various eastern varieties of Javanese (Hoogervorst 2010).[10] Still other dialects, which represent perhaps 70 million of the 90 million speakers, remain largely undescribed.

---

8. The primary materials, Conners, Thomas J., and Singgih Sugianto. 2011. Javanese Database. Jakarta Field Station, Max Planck Institute for Evolutionary Anthropology, are stored as part of The Language Archive at the Max Planck Institute for Psycholinguistics at <https://hdl.handle.net/1839/697bec61-af98–4df8-b38b-a029054aeb81> and are available for download.

9. The language described in those works is a mixed language spoken by Chinese in those cities, with elements of Javanese, Indonesian, and several mostly southern Min varieties of Chinese.

10. See further references in Vander Klok & Conners (2019).

3. Selected features in MSEA and Javanese dialects

In this section, I present in greater detail those features that are relevant to the dialectology of Javanese and those that are relevant to the typological position of Javanese. For coherence, these features are described and presented first for Central Javanese and then, dialect by dialect, for each of the peripheral varieties.

Here I invoke and will maintain below a distinction used by Vander Klok & Conners (2019: 69):

> Pertaining to terms concerning language, some use 'dialect' to refer to a variety of a language that is characterized by different grammatical features, while 'variety' is associated with a group according to some external factor, perhaps geographical or social (e.g., Wardhaugh 2015). For instance, in reference to the external factor of prestige, it may be more appropriate to say 'prestige variety' instead of 'prestige dialect'

In keeping with this terminological practice, 'Central' and 'peripheral' constitute a first-order distinction from the perspectives of both dialect and variety, as the peripheral dialects are peripheral geographically as well as culturally: all five varieties are spoken on a coast, mountain border, or some other boundary, away from the traditional centres of power and culture. Having introduced Central Javanese, the discussion of peripheral dialects will proceed geographically from west to east.[11]

Phonological and pragmatic evidence are presented first: this set of features is frequently grouped together and referenced to show the distinctiveness of 'Javanese' in the Austronesian typological context. Examination of these features in the peripheral dialects shows that they pattern more closely with one another than with the Central variety, and this collection of features motivates the first-order distinction between the central and peripheral dialects. Characterising 'Javanese' in terms of the central dialects alone therefore renders a picture inaccurate on both dialectological and typological grounds.

Following this I outline morphological and morphosyntactic features for each dialect; these are the features relevant for the discussion of Javanese within the broader MSEA linguistic area.

3.1 Central Javanese

Central Javanese is spoken by some 15 million people in the courtly cities of Yogya and Solo, which today fall within the political divisions of the Special Administrative District of Yogyakarta and the surrounding area of Central Java.

---

11. Where relevant, evidence demonstrating innovations within CJ *vis-à-vis* the peripheral dialects will be given.

3.1.1   *Lexical, phonological, and pragmatic features of Central Javanese*

This variety is characterised most prominently by a highly elaborate and articulated series of speech levels: different lexical and morphological items that are indexed according to the relative status of speaker and interlocutor.

Table 2 provides lexical equivalents from the five speech levels in CJ. *Ngoko* is the low, or neutral, speech level. It is the first acquired and most frequently used, particularly among peers and social equals. *Madyo* through *Krama Inggil* represent increasingly polite and stratified levels.[12]

Table 2. Some speech level distinctions in Central Javanese

Ngoko	Madyo	Krama	Krama Andhap	Krama Inggil	Gloss
wong		*tiyang*			'person'
iki	*niki*	*punika*			'this'
kandha	*sanjang*	*criyos*	*matur*	*ngendika*	'say'
mata		*mripat**		*paningal*	'eye'
aku		*kula*	*dalem*		1st person
kowe	*samang*	*sampeyan*		*penjenengan*	2nd person
		ratu			'king'

* With most body part terms and other lexemes potentially used to refer to humans, the *Krama* form is used as the neutral term in both *Ngoko* and *Krama* speech. The specifically *Ngoko* term, here *mata* 'eye', would be offensive with human reference, and is therefore generally only used with reference to animals or other non-humans.

While other register systems are used symmetrically between interlocutors, participants in the speech level system engage in asymmetric exchanges. So, a child will use *Krama* with or in reference to, say, a parent, and *Krama Inggil* with or in reference to a grandparent, whereas each of those would use *Ngoko* to the child. Example (2) shows three levels expressing the same semantic material and speech act, differing in speech level items.

<div align="right">Central Javanese</div>

(2)	a.	*Punika punapa inggih kagungan panjenengan?*	Krama
	b.	*Niki napa nggeh gadhahan sampeyan?*	Madyo
	c.	*Iki apa yo dhuwek--mu?*	Ngoko
		this INTERR yes possession 2	
		'Is this yours?'	Errington (1998)

This system has been well described elsewhere, particularly in Errington (1985, 1998) and Poedjosoedarmo (1968). In varieties spoken outside these centres the

---

12. An empty cell here means there is no distinct form for that level. *Ngoko*, *Madyo*, and *Krama* are 'base forms', and lexical items from *Krama Andhap* and *Krama Inggil* can be mixed in with any of those.

system is less common, less elaborate, and less frequently used (Errington 1998; Oetomo 1990), and it is becoming even less so (Smith-Hefner 2009; Conners 2010; Zentz 2015).[13]

The speech level system is probably the single most salient feature ascribed to Javanese; yet even this most definitional feature of Javanese is a recent innovation that most clearly characterises CJ.[14]

Javanese has a long and rich literary tradition which has resulted in a comparatively large lexicon, with significant borrowings from Sanskrit, Arabic, Chinese, Dutch, and Malay (Robson & Wibisono 2002; Zoetmulder 1983). The Central variety includes many lexical items from poetry and from *wayang*, the elaborate shadow puppet stories based on the Indian epics of the *Mahabharata* and the *Ramayana* (and a source of common cultural reference). On the whole CJ shows a great deal of innovation and lexical differentiation from peripheral dialects.

Phonologically, CJ is characterised by a vowel raising and vowel harmony system, where an open /a/ in word-final position raises to an [ɔ], as in (3a); this change then spreads harmonically to /a/ in a preceding open syllable, as seen in (3b). The harmony does not spread beyond one additional syllable, and is blocked by particular clusters, as seen in (3c) and (3d) respectively.

(3) a. *teka*   /teka/   [təkɔ]   'come'   Central Javanese
     b. *mata*   /mata/   [mɔtɔ]   'eye'
     c. *rabasa* /rabasa/ [rabɔsɔ] / *[rɔbɔsɔ] 'ravage'
     d. *warna* /warna/ [warnɔ] / *[wɔrnɔ]   'colour'

As will become important in later discussion, suffixation in CJ blocks vowel raising, as can be seen in the examples in (4):

(4) a. *teka-ne*   /teka-ne/ [təka-ne]   Central Javanese
         come-ASS
         'the coming'
     b. *mata-ku* /mata-ku/ [mata-ku]
         eye-1
         'my eye'

---

13. Increasing bilingualism in the more 'egalitarian' Indonesian is also putting pressure on the continued health of the speech level system.

14. Using evidence from Balinese, Clynes (1994) shows convincingly that the speech level system is an earlier innovation than previously thought. Nevertheless, it was most fully articulated in the 15th and 16th centuries out of the highly stratified courts of the Mataram period, crucially centred in Yogyakarta and Surakarta, where the CJ dialect is found, and not in the East Javanese centred Majapahit era, which had ended by the close of the 15th century.

It is relevant to note here that this pattern of vowel raising and vowel harmony is a relatively recent innovation in the CJ dialects. This can be seen through comparable forms from both Old Javanese (Examples (5a) and (5b)) and Middle Javanese ((5c) and (5d)):

(5)   a.   *dasa* [dasa] 'ten'   Hunter (1999)
     b.   *desa* [desa] 'village'
     c.   *dasa* [dɔsɔ] 'ten'
     d.   *desa* [desɔ] 'village'

3.1.2   *Morphological and morphosyntactic features in Central Javanese*

In isolating languages, mood, tense, aspect, and voice markers tend to be expressed via modals, auxiliaries, particles, and other independent word-like elements. In the most extreme cases, verb forms do not inflect for any of these features, or for agreement. There is often no distinction even between finite and non-finite forms of the verb. In all varieties of Javanese, there is similarly no morpholexical distinction between the finite and the nonfinite forms of predicates, nor do they inflect for mood, tense, or aspect (though see Conners & Brugman (2013) on Banyumasan Javanese). Predicates in Javanese also show no agreement inflection for person or number. They do, however, encode voice distinctions, and can host a series of applicative affixes. In Central Javanese, there are two distinct applicative paradigms, each of which inflects for four moods and two voices. CJ further uses affixation to signal two or three types of voice alternation: actor focus and two types of patient focus, one neutral and one accidental or non-volitional. The resulting paradigms are summarised in Table 1 above.

The verbal system of Old Javanese (OJ) does not show this affixation pattern. While OJ made use of applicatives and other suffixes, the patterns were nowhere near this complex. Old Javanese, similarly to other Austronesian languages such as the Philippine languages, marked a number of voice alternations primarily through a series of verbal prefixes and infixes (see Hunter 1999 for details).

Isolating languages strongly tend to form various types of predicate constructions periphrastically. Central Javanese has no periphrastic verbal constructions: all primary alternations are marked morphologically on the verb. This is true of causative constructions, which are marked by means of the applicative affixes shown in Table 2. It is also true of voice alternations, where in patient focus constructions, first and second person agents are marked with proclitics, and third person agents are expressed through a combination of voice affixation and optional use of oblique adjuncts. Agent focus constructions are signalled by nasal prefixation on the verb. As mentioned above, the verbal system of Central Javanese encodes a complex range of mood, voice, and argument alternations. In comparison to other western Indonesian languages, this system is strikingly full and articulated.

Except for the agents in the patient focus constructions, evidence of cliticisation in CJ is lacking. Specifically, phonological evidence from the interaction between vowel raising and affixation, and reduplication, clearly demonstrates that postverbal elements in this variety are in fact affixal (cf. the effects of adding a suffix in blocking vowel raising and harmony in (3) & (4) above). Further, there is no collapse of nominal attribution.

In CJ, causative constructions are formed morphologically through the addition of one of a number of applicative suffixes, as seen with the use of *--ake* (marking the agent) and *-i* (marking the beneficiary) in Example (6).

(6) *Sutawijaya ng-asor-ke*   *lan ng-uasa-i*   *Pajang*
Sutawijaya AGFOC-humble-APPL1 and AGFOC-power-APPL2 Pajang
'Sutawijaya humbled and ruled over Pajang.'   Central Javanese

Voice constructions are formed morphologically, through the use of first and second-person agent proclitics or third-person voice affixation, as seen in the examples in (7):

(7) a. *Buku kuwi wis tak-waca.*   Central Javanese
    book that PFCT 1-read
    'I already read that book.'
   b. *Buku kuwi wis kok-waca.*
    book that PFCT 2-read
    'You already read that book.'
   c. *Buku kuwi wis di-waca*   *(dening) Gunnawan.*
    book that PFCT PATFOC-read (by)   Gunnawan
    'That book was already read by Gunnawan.'

In CJ there are no periphrastic equivalents to the causative and voice constructions given above.

Another defining feature of isolating languages is the correlation between morpheme and word and the predominance of cliticisation vs. affixation. In a purely isolating language, all morphemes are independent words.[15] In comparison to affixes, clitics show a greater degree of independent wordhood, though they still show some type of dependency; phonological, morphological, or syntactic. All dialects of Javanese allow for some degree of cliticisation; this can be seen in (7a) with a first person agent in a voice alternation construction. All dialects of Javanese have a corresponding construction, *modulo* variation in the lexical items used.

---

15. See Gil ('What Does It Mean to Be an Isolating Language', this volume), for an important discussion on the usefulness of the traditional morphological typology, particularly with respect to the cross-linguistic difficulties in defining "word".

Further, all varieties of Javanese demonstrate a clear use of affixation. An example is the nasal prefix previously discussed, which exhibits the phonological behaviour of an affix rather than a clitic, i.e., place assimilation to the following segment. Such affixation in Javanese, however, is highly restricted.

Evidence from vowel raising and reduplication shows that in the peripheral varieties of Javanese, there is a greater use of encliticisation as opposed to suffixation. In CJ, as mentioned above, an open /a/ in a final syllable will raise to [ɔ]. This raising, however, is blocked by the presence of a suffix, as demonstrated in Examples (4a) and (4b) above, providing clear evidence that the suffix forms a phonological word with the root, as it affects its internal morphophonological behaviour. This pattern is not replicated in a number of the peripheral varieties, as will be demonstrated below.

Another feature that has been identified as characterising the MSEA linguistic area is the collapse of nominal attribution. That is, MSEA languages exhibit only weak morphological differentiation between adjectival modification, genitive or possessive modification, and relative clause modification: they simply use parataxis. It has been argued that Standard Indonesian, as an example of a western Austronesian language, moderately differentiates these three constructions (see Gil 2011). In Central Javanese, possession is marked through an associative enclitic on the modified noun (8a); adjectival modification is similarly marked through an associative enclitic on the modified noun (8b); and relative clauses are formed with an overt relativiser. Examples (8a–b) show also that simple parataxis between a possessor expression and the possessee expression is ungrammatical. Finally, CJ relative clauses must include the associative marker on the head, and the relative clause must be introduced by an overt relativiser, as seen in the grammaticality of (8c): compare this with the ungrammaticality of (8d), without the relativiser. Central Javanese therefore moderately differentiates these three functions.

(8) a. *buku-ne Singgih. *buku Singgih* Central Javanese
         book-ASS Singgih
         'Singgih's book.'
     b. *buku-ne abang. *buku abang*
         book-ASS red
         'red book'
     c. *buku-ne sing aku wis   maca*
         book-ASS REL 1   PFCT AGFOC-read
         'the book that I read'
     d. *buku-ne aku wis maca*

In the peripheral varieties, however, marking of adjectives and relative clauses is collapsed, and possession can be expressed through parataxis. The CJ pattern of moderate differentiation is not found in any of the peripheral dialects, although the patterns of collapse differ among them.

Here I have introduced a set of features that characterise the Central Javanese dialect. Innovative lexical features, vowel harmony, vowel raising, and the speech level system are properties that distinguish Javanese from other Austronesian languages. Below they will be used to distinguish the Central dialect from peripheral dialects. Morphological complexity in the verbal system, morphological formation of causative, applicative, and voice alternation constructions, and the relevance of the notion 'wordhood' with respect to cliticisation, are properties which distinguish CJ from languages of MSEA. Below they will be used to show how the peripheral dialects pattern with the MSEA languages.

3.2 Banten

Banten Javanese is the only Javanese variety under exploration here that is spoken in an area discontiguous with other Javanese speech communities. Spoken in the northern part of Banten Province, on the western edge of the island of Java, it is surrounded by a Sundanese speaking area, and shows strong evidence of contact. There are reports of between three million (Sigit & Sulistiyono 2017) and five hundred thousand (Simons & Fennig 2018) speakers of Banten Javanese.

3.2.1 *Lexical, phonological, and pragmatic features in Banten*

This dialect is characterised by the generalisation of open, final /a/ becoming [ə] word finally, perhaps under influence from Betawi. This generalisation is robust, having been extended to lexemes borrowed from Indonesian as in (9b) and found in a different set of conditions than the vowel raising rule of CJ (9c):[16]

(9) a. *nana* /nana/ [nanə] 'neg. exist' Banten
 b. *bebelanja* /bebelandʒa/ [bəbəlandʒə] 'shop'
 c. *ora* /ora/ [orə] 'neg' cf. CJ [ora]

There is no vowel harmony pattern with this mutation, as shown in (9a).

The speech level system exists in Banten in highly restricted social contexts. While there is evidence of influence in Banten of the adoption of the speech level system from the Central Javanese varieties, its application in Banten is only partial, incomplete, and often conflicting. For example, the Central Javanese high or *Krama* speech level applicative marker *-aken* has been generalised in Banten to all levels.

Lexically, Banten is highly divergent from other varieties of Javanese, the result of significant contact with the surrounding language of Sundanese and the nearby Malay variety of Jakarta Indonesian. See Section 4.3 for more discussion of Banten lexicon.

---

16. There is a small set of lexical exceptions to the rule in CJ. Example (9c) is one of them.

3.2.2   *Morphological and morphosyntactic features in Banten*

Banten has a periphrastic causative construction. Whether a periphrastic passive exists is unknown. Its verbal system is simpler than that of CJ, though not to the extent seen in the other peripheral dialects. In Example (10), Banten shows a prefix (*di-*) only to add the (discourse-available) instrument argument, where CJ has the same prefix plus the applicative suffix *-ake*. That *-ake* is a suffix can be seen in example (11) below, where the root *tuku* 'buy' undergoes phonological change when affixed: [tuku] > [tukɔʔ]. No phonological change is seen on the stem with the addition of *di-*, and so its status as a prefix is less clear.

Banten

(10)   *Lamun sing bagi-an   jabe   di-sogok   gati tah*   cf. CJ
      if   rel   part-AN[17] outside PATFOC-poke.into hard PTCL   di-sogok-ake
      'But it's hard to poke it into (the gutter) from the outside.'

It has also been noted above that the applicative ending *-aken*, which marks the formal or *Krama* causative in CJ, has been regularised to both formal and informal speech levels in Banten Javanese, seen in (11).

(11)   *Kite arep nukok-aken   sate bae!*
      1SG will AGFOC-buy-CAUS sate only
      'I'm just going to buy them sate!'

This applicative occurs in imperative and voice constructions, showing a great deal of syncretism in contrast to the CJ system.

3.3   Banyumasan

Banyumasan is spoken in the southwest corner of Central Java Province. The dialect described here is characteristic of the cities in and around Cilacap, Purwokerto, Purbalingga, and Banjarnegara. It does not include the varieties spoken in Kebumen, Brebes, or Slawi. Banyumasan has perhaps 15 million speakers. Unlike the other varieties described here which are separated from Central Javanese by significant geographical boundaries, there is no major geographical boundary between Banyumasan and Central Javanese. However, in all the relevant features, Banyumasan patterns with the other peripheral dialects and not Central Javanese.

---

17. The suffix *-an* has a wide range of functions in Javanese, none of which is germane to the current discussion. For ease of exposition, therefore, it is glossed simply as '-AN' throughout.

3.3.1   *Lexical, phonological, and pragmatic features of Banyumasan*

Phonologically, Banyumasan is characterised by a lack of vowel raising and vowel harmony. Similarly to the Pesisir Lor varieties discussed below, other than in a few systems such as the pronominal system, Banyumasan maintains some lexical forms from Old or Middle Javanese no longer in use in other varieties, and it has innovated or borrowed lexemes (see Section 4.3). The speech level system exists in Banyumasan, though it is reduced compared to the highly articulated form found in the Central variety.

3.3.2   *Morphological and morphosyntactic features of Banyumasan*

The Banyumasan dialect shows a periphrastic causative construction but no periphrastic voice construction, similar to that shown below for Tengger. It also shows a very restricted verbal system. There are no distinct applicative forms for neutral (12a), imperative (12b), propositive (12c) and subjunctive (12d) sentences:

Banyumasan

(12) a. *Sira esih manjing-na montor-e?*   cf. CJ *ng-lebok-ake*
          2   still   enter-APPL1 car-ASS
          'Are you still putting your car in?'

      b. *M-bukak-i   bae lemari-ne   nggone nyong.* cf. CJ *m-bukak-ana*
          AGFOC-open-APPL2 only cupboard-ASS place   1
          'Just open the cupboard at my place.'

      c. *Nggon kiye tak-jagong-i.*   cf. CJ *jagong-ane*
          place   this 1-sit-APPL2
          'I'll just sit here.'

      d. *Maen-e nyong nuduh-na   ngarep pasar.*   cf. CJ *nuduh-na*
          good-ASS 1   AGFOC-show-APPL1 front   market
          'It's best if I show (him) the front of the market!'

Banyumasan also shows a collapse of nominal attribution, as shown in (13).

(13) a. *buku Singgih*   Banyumasan
          book Singgih
          'Singgih's book'

      b. *buku abang*
          book red
          'red book'

      c. *buku nyong m-aca wingi*
          book 1   N-read yesterday
          'the book I read yesterday'[18]

---

18. Note that this is not a patient focus construction, as indicated by the presence of the *N-* marker on the verb. The corresponding patient focus in Banyumasan would be *buku tak-waca wingi*.

Example (13a) shows that possessive attribution is expressed in Banyumasan through parataxis. Similarly, in (13b) with adjectival modification, the head is unmarked. From (13c) we can see that relativisation is grammatical with a null relativiser.

These all contrast with CJ, where the head in all three of these types of nominal attribution is obligatorily marked by the associative marker, as shown above in Example (8).

In sum, all three nominal attribution functions use the same paratactic construction in Banyumasan, in contrast with CJ, but coinciding with the pattern found throughout the MSEA linguistic area.

3.4   Pesisir Lor

The Pesisir Lor (PL) classification refers to a broader dialect area than any of the others under consideration here, and potentially encompasses almost any of the dialects spoken on the north coast of Java from Surabaya in the east to Cirebon in the west. For the present paper, I define this variety as that spoken in and around the cities of Pemalang, Tegal, and Pekalongan, in the northwest coast of Central Java Province. There are perhaps 10 million speakers of this variety.

3.4.1   *Lexical, phonological, and pragmatic features of Pesisir Lor (PL)*
Phonologically, PL is characterised by a general lack of vowel raising and vowel harmony; however, there is sporadic mutation of /a/ in open, final syllables to [ə], as in (14c). No harmony pattern is associated with this mutation.

(14)   a.   *ana*   [ana]   'exist'   Pesisir Lor
       b.   *apa*   [apa]   'what'
       c.   *lima* [limə] 'five'   cf. CJ [limɔ]

The speech level system is present in the PL variety; however, it is not as robust or as fully articulated as in the Central variety. Lexically, the pronominal system is slightly archaic and divergent, but it is relatively neutral in terms of lexical retentions or innovations when compared to other varieties.

3.4.2   *Morphological and morphosyntactic features of Pesisir Lor*
The PL dialects have a periphrastic causative construction. They do not, however, have a periphrastic voice construction. In line with other peripheral dialects, their verbal system is simpler than that of CJ.

(15) a. *Ujung-ujung di-dol-na   jas   kuwe.*   Pesisir Lor
         REDUP-end   PATFOC-sell-APPL1 coat that
         'In the end, she sold the coat.'

     b. *Umpama jas   kuwe di-dol-na…*
         Suppose jacket that PATFOC-sell-APPL1
         'If that jacket were to be sold…'

     c. *Kon tak-golet-i.*
         2SG 1-seek-APPL2
         'I will be the one to look for you!'

     d. *Wingi   wis   tak-golet-i.*
         yesterday PFCT 1-seek-APPL2
         'Yesterday I already looked (for it).'

(15a) shows a PL example in the indicative, with the form of the verb *di-dol-na*.
That same verb form is used in (15b) in a subjunctive clause. In CJ, these verbs have
different forms: the indicative clause is marked with *-ake* and the subjunctive clause
is marked with *-na*. Examples (15c) and (15d) show a collapse of distinction in the
verb forms in the propositive and indicative patient focus constructions; both use
the *-i* applicative marker. In the corresponding constructions in CJ, these clauses
would be marked with *-ane* and *-i* respectively. The entire verbal system of all PL
varieties is significantly simpler than the CJ system.

    The phonological evidence for encliticisation in PL is weaker than in some
other dialects; the vowel mutation (final /a/ to [ə]) is only sporadic, and there is no
vowel harmony.

3.5   Tengger

The Tengger language is spoken by an estimated 90,000–600,000 speakers on the mid
and upper elevations of the Bromo Semeru massif in East Java (Simons & Fennig
2018; Badan Pusat Statistik 2010). The Tengger people are predominantly Hindu
Javanese, and for centuries have remained largely isolated from contact with low-
landers, which accounts for the dialect's significant divergence from other varieties.

3.5.1   *Lexical, phonological, and pragmatic features of Tengger*
The Tengger dialect is distinguished by a complete lack of vowel raising, as can be
seen in Example (16); it also lacks vowel harmony.

(16) a. *manja* [mandʒa] 'to plant'   Tengger
     b. *gaga*   [gaga]   'dry field'

Tengger also shows lexical retentions from Middle Javanese not found in other varieties of Javanese. Further, there are a number of lexemes which neither occur in other varieties nor can be traced back to either earlier forms of Javanese or other proto-Malayo-Polynesian etymons, suggesting strongly the presence of a non-Austronesian substrate influence at some time. The pronominal system diverges significantly from other Javanese varieties not only in the retention of many elements from Middle and Old Javanese, but also in having innovated a gender distinction in the first-person pronouns, as shown in Table 3:

Table 3. Tengger pronouns

	Masculine/Feminine	**Familiar**	**Distant**
1st person	*(r)eyang/isun*	–	–
2nd person	–	*sira*	*rika*

Tengger has no native speech level system.

3.5.2   *Morphological and morphosyntactic features of Tengger*

In addition to morphological causatives and passives, Tengger has both periphrastic causative and periphrastic voice constructions. The voice construction is shown in Example (17):

(17) a. *Eyang   kenek   antem.*   Tengger
          1MASC PATFOC hit
          'I got hit.'
      b. *Isun   kenek   tending.*
          1FEM PATFOC kick
          'I got kicked.'

As signalled in the translations, there is an adversative connotation to this voice construction, whereas the morphological voice alternation is pragmatically neutral. In CJ, both voice alternations are formed morphologically, as shown in (18):

(18) a. *Aku di-antem.*   Central Javanese
          1   PATFOC-hit
          'I was hit.'
      b. *Aku k-antem-an.*
          1   PATFOC2-hit-AN
          'I got hit.'

Example (18a) gives the neutral voice alternation formed with the prefix *di-*. In (18b), however, the adversative or accidental passive in CJ is formed using the affixes *ke-* and *-an*.

Tengger also has a periphrastic causative construction, as in (19):

(19) *Eyang   kon   umbah klambi-ne.*   Tengger
     1MASC CAUS wash   clothes-ASS
     'I had the clothes washed [by someone else].'

Note the lack of morphology on the lexical verb, *umbah*. There is strict adjacency between the causative marker and the verb; the construction is monoclausal. Note how this construction differs from the CJ causative construction, shown in (20a–b). In (20a), the causative is formed morphologically through the use of the applicative marker. Example (20b), however, is not monoclausal: there is not strict adjacency between the *kon* and the complement verb. The recipient of the order can naturally appear between the two.

(20) a. *Aku ng-umbah-ake   klambi-ne.*   CJ
       1   AGFOC-wash-CAUS clothes-ASS
       'I had the clothes washed.'
   b. *Aku kon   (Siti) ng-umbah-i   klambi-ne.*
       1   order (Siti) AGFOC-wash-APPL2 clothes-ASS
       'I ordered (Siti) to wash the clothes.'

Beyond the periphrastic construction, the verbal system in Tengger is among the simplest of the peripheral dialects, as shown in Table 4.[19]

Table 4. Verbal morphology in Tengger Javanese

Mood	Voice		Prefix	Neutral	Applicative I	Applicative II
Indicative	Agent focus		N-			-na
	Patient focus I		di-	Ø	-i	
	Patient focus II		ke-			Ø
Propositive		Agent focus	tak N-			-na
Optative		Agent focus	N-	-a		
	Patient focus I					
		2nd person	Ø	-en	-i	-na
		3rd person	di-	-a	Ø	Ø
		unmarked		Ø	-i	-na

---

19. The subjunctive marker *-a* has a much wider use and range of functions in Tengger than in other dialects. See Conners (2008) for a full account.

Tengger has a series of person-marked proclitics and enclitics used in voice, propositive, and possessive constructions. However, it lacks the vowel raising and vowel harmony that I am here using as evidence of greater use of enclitics as opposed to suffixes in other varieties, so I remain neutral on the issue of Tengger enclitics. Tengger also shows a complete collapse of nominal attribution.

3.6   Osing

Osing Javanese is spoken in Banyuwangi Province, in the far eastern edge of the island of Java. Located immediately across the strait from Bali, it is heavily influenced by contact with Balinese.[20] There are some 500,000 speakers of Osing. There has been a recent move among Osing speakers to preserve their variety and it is now used in local radio programs and other local media, as described in Arps (2009) and Wittke (2019).

3.6.1   *Lexical, phonological and pragmatic features of Osing*

Osing is distinguished from other dialects discussed here in having generalised to all environments the vowel raising and vowel harmony pattern described for Central Javanese, as shown in (21):[21]

(21)   a.   *konco* /kontʃo/   [kɔntʃɔ]   'friend' cf. CJ [kɔntʃɔ]   Osing
       b.   *moto* /moto/   [mɔtɔ]   'eye'   cf. CJ [mɔtɔ]

Lexically, Osing retains many elements from Middle Javanese that have been lost in more central varieties. This is especially clear in its pronominal and deictic systems; the pronouns are given in (22).

(22)   a.   *i(ng)sun*   1st person   Osing
       b.   *(h)iro, siro, riko* 2nd person

These forms are all found in Middle Javanese, and several occur in Old Javanese as well. Second person pronouns in Osing are marked for familiarity with the speaker and a degree of politeness, but this is independent of the speech level

---

20. In fact, the name of the dialect, *Osing*, is taken from the verbal negator which was borrowed from Balinese (the native Javanese form being *ora*).

21. There are several cases where vowel raising does not apply in Osing, such as *keluarga* [kəluarga] 'family'; however, this is most likely a more recent borrowing from Indonesian. The fact that the addition of the associative marker and other suffixes does not block vowel raising in most cases in Osing suggests that these forms have undergone reanalysis whereby the raised vowel is now the underlying one.

system described elsewhere. Osing's speech level system is rudimentary, and where there are speakers with greater control over some form of the speech level system, it tends to have been acquired not natively, but rather upon interaction with outside people, or experience in outside areas (Wittke 2019).

3.6.2 *Morphological and morphosyntactic features of Osing*

The features of Osing under discussion include: a periphrastic voice alternation (but not a periphrastic causative); its simpler verbal system, in comparison to that of Central Javanese; a collapse of nominal attribution; and the strongest distinction of the peripheral dialects in its categorisation of certain elements as enclitics, as opposed to suffixes. All dialects have person marked proclitics which function in voice constructions and person marked enclitics which function in possessive and object constructions. What is of note here is the weak bond between the root and other markers, such as the associative marker *-(n)e*, the applicative markers, and the polyfunctional *-an* ending.

In addition to a morphologically formed voice construction, Osing also makes use of a periphrastic voice construction, shown in (23):

(23) *Konco-ne ison iku kenek bagi-bagi.*   Osing
      friend-ASS 1   that PATFOC REDUP-part
      'My friends got separated.'

This construction contrasts with the morphological voice alternation in conveying an accidental or unintentional connotation.

The verbal system of Osing is much simpler than the pattern displayed in CJ. Most of the applicative functions are collapsed into two applicative endings, *-i* and *-no*, as shown in (24):

(24) a. *Umpane podho-podho nyegurno,*   *dheke*   Osing
         bait-ASS REDUP-same AGFOC-plunge-APPL1 3
         ole, ison osing ok.
         get 1   NEG EMPH
         'We put in the bait together, he got some fish, I didn't'
     b. *(Siro) umpan-e nyegur-no*   *ning banyu-ne!*
         (2)   bait-ASS AGFOC-plunge-APPL to   water-ASS
         '(You) throw the bait into the water!'

The two sentences in (24) differ in their mood: the first is a declarative and the second is an imperative. In CJ, different applicative suffixes (*-ake* and *-na*, respectively) would be used on the same verb root, reflecting this difference. However, in Osing, the same applicative marker is used, exemplifying fewer distinctions marked on the verb. The pattern in Osing is similar to that seen in Tengger, as shown in Table 4.

Osing also shows a strong tendency toward cliticisation over affixation. As noted above, Osing has generalised the vowel raising phenomenon found in other varieties of Javanese to most conditions. However, it has not regularised it to all domains. Unlike in other varieties which have vowel raising and vowel harmony, the presence of an affix does not block this in Osing. This suggests that affixes in Osing are less bound to stems and show greater independent wordhood. In example (4), we saw that the associative marker does not block the vowel raising pattern for the word *kanca*. That is, we have the form [kɔntʃɔ-ne] attested instead of the expected [kantʃa-ne]. This pattern repeats with all other 'suffixal' markers, as shown in example (25):

(25) a. *moto-mu* [mɔtɔ-mu] 'eye-2 = your eyes' *[mata-mu]   Osing
      b. *ono-an-e* [ɔnɔʔ-an-e] 'exist-AN-ASS'   *[anane]
      c. *ono*   [ɔnɔʔ]   'exist'   *[anak]

Osing has a series of morphemes which are unambiguously proclitic, and which are used in voice constructions and also in the propositive construction. It also has a series of possessive and object clitics for first and second person, as exemplified in (25a). Interestingly, we see in (25b) that in Osing, not only does the *-an* ending, which can derive nouns from verbs, not block vowel raising, but there is also an epenthetic glottal which blocks the merger of a root final vowel with the vowel of a vowel initial affix. Again, this is clear evidence that the enclitic markers in Osing do not form a phonological word with the root, in contrast with similar contexts in CJ. This then is strong evidence for the greater reliance on clitics in Osing.[22]

4.   Discussion

4.1   Central vs. peripheral varieties of Javanese

In order to highlight the variation that exists in Javanese dialects, I have presented a set of features that, taken together, are frequently cited to characterise the entire 'language' Javanese. I have chosen precisely these features because they show the unexpected points of variation between the dialects, justifying the first-order distinction into Central and peripheral that I have described here. Further, I have highlighted points of variation on most significant linguistic levels – phonological, lexical, morphosyntactic, and pragmatic. Additional areas of dialect variation exist across these components of the grammatical system.

---

22. Data for the differentiation of attributive constructions were not collected at the time of data collection and remains for future work.

The variation of Javanese dialects is of course not limited to these features, and not all dialect variation aligns with the central/peripheral first-order distinction. In fact, there may be as much variation as alignment among the peripheral dialects. For example, a general phonological process common to many varieties of Javanese is the devoicing of word-final oral stops along with the place-of-articulation change of word-final /k/ to [ʔ]. However, Banyumasan does not display either of these characteristic processes.

Dialects also vary widely in their personal and demonstrative pronominal paradigms, to the extent that individual dialects are frequently referred to by some subset of these differentiating items. Tengger, for example, is sometimes referred to as *basa eyang-isun*, or 'I-I language'; Similarly, Pesisir Lor varieties are referred to as *basa inyong*, also 'I language'.[23]

Thus, it is self-evident that the peripheral dialects do not constitute a linguistically based dialect grouping, except in contrast with the Central dialect – the latter however can be justified on linguistic grounds. The main point of this chapter is that the Central dialect, while coherent and noted as the prestige variety, does not represent the language, as a whole or even in the majority.

4.2 Javanese in typological perspective

The MSEA and MM linguistic areas are defined in terms of sets of linguistic features (largely overlapping) that are shared among many of their respective member languages. Among these features, and particularly relevant for the current volume, is their highly isolating morphological typology. While many languages in the MSEA and MM areas display highly isolating morphology, including Tai-Kadai, Hmong-Mien, Tibeto-Burman, Austroasiatic, and Papuan languages, it has generally been held to be less true for the Austronesian languages of Southeast Asia. I have now shown that this characterisation does not hold for the Javanese language in general. It has been shown that the five peripheral dialects of Javanese explored here better represent the general situation of Javanese than does CJ. Of course, not all of the peripheral dialects will share the full range of the features and phenomena as described above, but they are all more isolating than Central Javanese.

Each dialect was described above in terms of two sets of features: one relevant for a discussion of dialects, and the other relevant for a discussion of broader typology. Of the set of features that define the MSEA and MM areas, as given above,

---

23. The variation in the personal pronoun systems may be attributable to the fact that each variety may have only two actual pronominal forms – first and second – unmarked for number or gender (the exception, noted above, being Tengger).

I have described the Javanese dialects only in terms of those features which inform the identification of a language as isolating or non-isolating, given the salience of this feature for both linguistic areas. These features, a subset of the list given in Section 1, are repeated here:

–   collapse of morphosyntactic distinction between types of nominal attribution
–   limited inflectional and derivational morphology
–   periphrastic constructions to express voice and valency possibilities
–   greater use of clitics as opposed to affixes

Central Javanese contravenes all of these diagnostic features. It is this, along with its lack of other relevant features, such as nominal classifiers and tone, that has marked Javanese as falling outside of the MSEA grouping. From this perspective, Central Javanese also falls outside of the MM linguistic area. However, most of the peripheral varieties described here conform to some or all of these features, thereby placing them within both typological groups to some degree.

For example, Tengger shows a collapse in nominal attribution, a much simpler verbal paradigm, and periphrastic constructions for both causatives and voice alternations, placing it comfortably within the Sprachbund. Membership within a Sprachbund implies significant contact with speakers of potentially unrelated languages over time periods sufficient for linguistic features to become shared. It is therefore necessary to appropriately identify potential contact preconditions for the rise of these points of variation.

In other respects, however, CJ and indeed many other Western Austronesian languages conform to the broad typology. All varieties of Javanese treat pronouns as an open class, allowing kin terms, deictics, and proper names to function anaphorically (Conners, Brugman & Adams 2016; Adams & Conners 2019). In its verbal system, tense, aspect and mood are all expressed through independent modals and auxiliaries: this is where Javanese shares properties with other Western Austronesian languages.

In their phonology, all varieties of Javanese are different from other Western Austronesian languages but similar to the general MSEA Sprachbund in having an incipient tone system: the distinction in phonation types between modal and breathy voice, which is phonemic in Javanese,[24] is often a precursor of tonogenesis, as seen for example in the Chamic languages (Matisoff 1973; Thurgood 1999).

---

24. It is likely that some varieties of Javanese are losing the phonemic contrast between breathy and modal voice in vowels and are introducing a phonemic voicing distinction on obstruents. This would be under heavy influence from Indonesian; more research is required.

I have just described three features that all Javanese dialects share, which show their similarity to other MSEA languages, in contradistinction to Western Austronesian. In the next section, I demonstrate that the CJ variety here described is actually the outlier, in terms of a broader regional typology, and that the peripheral dialects conform much more closely to the typological norm of the language area.

4.3 Javanese in diachronic perspective

There are a number of linguistic properties which characterise most of the peripheral dialects of Javanese, when viewed in contrast to the Central dialect. For example, the vowel raising and vowel harmony patterns discussed above. Four of the five peripheral dialects show no vowel raising; Banten displays an independently developed pattern of vowel centralisation. CJ's innovative vowel harmony also does not appear in four of the five varieties. The one dialect that does show vowel raising and harmony, Osing, has generalised the pattern to all cases. What is common among all five of the peripheral dialects is that these phenomena that are characteristic of the Central dialect do not exist. These distinguishing features of CJ have, in the literature, erroneously been attributed to Javanese as a whole. However, importantly, the CJ dialect has been the innovative dialect. In placing Javanese within the MM linguistic area, it is important to understand why it was traditionally excluded from the MSEA linguistic area. 'Javanese' was exemplified by CJ. As I have gone to some length to demonstrate above, this characterisation of Javanese is not accurate.

Further to this argument, I present here a detailed discussion of the comparative lexicons of CJ and the peripheral dialects to further evidence the innovative nature of CJ, at this point bringing in earlier stages of the language.

In terms of lexicon, the peripheral varieties display greater conservatism in many systems, including pronominal systems.[25] In each of the tables below, comparative forms are given for each of the modern dialects, as well as the equivalent forms in Old Javanese, and, where available, a reconstructed proto-form. In some cases, the proto-form can be reconstructed to Proto-Western-Malayo-Polynesian (PWMP), Proto-Malayo-Polynesian (PMP), or all the way to Proto-Austronesian (PAN). For ease of reference in the following tables, forms with antecedents in OJ are shaded in dark grey; forms with antecedents in PAN/PMP are shaded in light grey.

---

25. Note that there is no singular/plural distinction in Javanese pronouns. Deictic pronouns often serve as 3rd person personal pronouns.

Table 5 shows that in the pronominal paradigm, CJ has maintained the PAN first person personal pronoun, but has innovated the second person form. This contrasts with Osing and Tengger, which maintain both the PAN second/third person forms and the OJ first person forms. Banten maintains both PAN forms, though there has been a reanalysis of PAN *ita from second to first person (which is seen elsewhere in PMP).

Table 5. Dialect variation in personal pronouns

	1st person	**2nd person**
Proto	PAN *aku	PAN *iten/*ita; PAN *si ia
Old Javanese	*ingsun, isun; aku*	*sira* (2nd/3rd person); *rika* (3rd person)
CJ	*aku*	*kowe*
Banten	*kite, kule*	*sire*
Banyumasan	*nyong*	*sira, rika*
Pesisir Lor	*e/inyong*	*kon, kowe*
Tengger	*eyang* (m) *isun* (f)	*sira, rika*
Osing	*i(ng)sun*	*(h)iro, siro, riko*

The pattern of lexical innovation of the CJ variety found in the pronominal system extends to other semantic domains, including vocabulary for body parts, agriculture, functional items and more. Some examples are provided in Tables 6–8:

Table 6. Dialect variation in body part terms

Gloss	**'Nose/mouth'**	**'Eye'**	**'Head'**
Proto	PMP *ijuN	*PMP mata; PAN *maCa	PAN *qulu
Old Javanese	*(h)irung*	*mata*	*ĕṇḍas* (*śirah* [SKT])
CJ	*irung*	*mripat*	*sirah*
Banyumasan	*cungur*	*mata*	*ndhas*
Banten	*cungur*	*mate*	*endhas*
Pesisir Lor	*irung*	*matə*	*(ə)ndas*
Tengger	*congor**	*mata*	*ndhas*
Osing	*irong*	*moto*	*ndias*

* Possibly related to Malay *moncong* 'snout' <PWMP *mu(n)cuN

Table 7. Dialect variation in agricultural terms

Gloss	'To plant'	'To hoe'	'Field'
Proto	PAN *t(um)anem	PMP *mula 'to plant'; PWMP *bacuk	
Old Javanese	*atandur/anandur*	pacul; olah 'to move'?	tĕgal; *gagā*
CJ	*nandur*	*gebuk, bacuk*	*tegal**
Banten	*nandur*	*macul*	*tegal*
Banyumasan	*nandur*	*macul*	*kebun/tegal*
Pesisir Lor	*nandur*	*macul*	*tegalan*
Tengger	*manja***	*molah*	*gaga*
Osing	*nandiur*	*macol*	*tegalan*

* This root is possibly from PWMP *tegaN 'dry', through irregular derivation. The cognate in Tagalog *tigáŋ* 'extremely dry' refers mostly to soil (thanks to D. Kaufman p.c. for pointing this out.)

** D. Kaufman (p.c.) points out, more tenuously, that this might be a borrowing from Sanskrit *mañjari* 'f. cluster of blossoms, flower.'

Table 8. Dialect variation in other vocabulary

Gloss	'Near'	'Money'	'What'	'Not exist'
Proto			PMP *apa	PWMP *la(N)ka 'rare'
Old Javanese	*parĕk*	*pisis*	*paran; apa*	(*langka* not in Zoetmulder);* *tan* 'not' + *hana* 'exist'; *ora < tanora* 'not'
CJ	*cedhak*	*dhuit***	*opo*	*ora ono*
Banten	*parek*	*picis*	*ape*	*lake, nane*
Banyumasan	*perek*	*dhuwit*	*apa*	*langka*
Pesisir Lor	*parek*	*dhuit*	*apa*	*langka*
Tengger	*parek*	*picis*	*paran*	*nana*
Osing	*parek*	*picis*	*paran*	*hing ono*

* OJ *nanā* 'broken, destroyed'. OJ *nanā 2* 'varied, various [SKT]'

** Perhaps from *duitj* Old Dutch coin worth 1/160th of a guilder (D. Kaufman, p.c.).

Taken together, the examples in Tables 5–8 identify specific lexical items in several semantic domains where the peripheral dialects tend towards conservatism, compared to the Central dialect. In this very small sample of twelve lexical items, half of the CJ forms are innovative, that is, have no antecedent in PAN or Old Javanese. The maximum number of such innovations in any other dialect is three. These items demonstrate the point made in Section 3 that substantial lexical variation exists between CJ and the peripheral dialects.

5.   Conclusion

I have argued for two main points in this paper. First, that Javanese evidences a wide range of dialect variation, and that what has often been described as 'Javanese' more appropriately characterises the Central Javanese dialect. Second, when viewed from the perspective of the peripheral dialects – here exemplified by Osing, Tengger, Pesisir Lor, Banyumasan and Banten – Javanese can be better situated with respect to surrounding linguistic areas, in particular the MSEA and MM Sprachbunds. As a final point, comparative evidence suggests that a strong direction of differentiation was via innovation in the Central dialect.

Historically, then, we can speculate that Javanese was earlier a more typical isolating language. Eventually, a single variety became well established, and emerged as a prestige variety, exerting a strong influence on surrounding dialects. This prestige variety underwent a series of changes that resulted in a far more complex verbal paradigm, non-periphrastic verbal constructions, more frequent use of affixation, and other features associated with non-isolating languages. These changes then spread out radially from this central variety. As reported by Gil (Gil 'What Does It Mean to Be an Isolating Language', this volume), a similar process seems also to underpin the development of Malay and modern Indonesian.

To substantiate these claims, I marshalled data from six dialects with respect to two sets of features: those that differentiate CJ from the other five, and those that, I have argued, better contextualise Javanese within the surrounding linguistic areas.

The MSEA linguistic area, as its name suggests, generally does not include languages of island Southeast Asia, particularly those spoken on the islands nearest to that area, Indonesia. This, at least, is the traditional view, and it is supported by the linguistic properties of Standard Indonesian as well as Standard Javanese. However, when non-Standard dialects are taken into account, that linguistic area can be expanded to include many of the languages of Indonesia and even into Papua – precisely the hypothesis advanced by Gil (2015 and 'What Does It Mean to Be an Isolating Language', this volume).

Acknowledgements

References

Adams, Nikki & Conners, Thomas J. 2020. Imposters and their implications for third person feature specification. *Journal of Linguistics* Vol. 58.2: 537–567.

Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices* [Oxford Studies in Typology and Linguistic Theory]. Oxford: OUP.

Arps, Bernard. 2009. Osing kids and the banners of Blambangan: Ethnolinguistic identity and the regional past as ambient themes in an East Javanese town. *Wacana* 11(1): 1–38. https://doi.org/10.17510/wjhi.v11i1.142

Badan Pusat Statistic. 2010. Kewargaan, suku bangsa, agama, dan bahasa sehari-hari penduduk Indonesian: Hasil sensus penduduk 2010. Jakarta, Indonesia.

Benedict, Paul K. 1994. On VO vs OV in Southeast Asia. *Linguistics of the Tibeto-Burman Area* 17(1): 173–174.

Bisang, Walter. 1996. Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in South-east Asian languages. *Studies in Language* 20: 519–597. https://doi.org/10.1075/sl.20.3.03bis

Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In *The Cambridge Handbook of Areal Linguistics*, Raymond Hickey (ed.), 19–39. Cambridge: CUP. https://doi.org/10.1017/9781107279872.003

Clynes, Adrian. 1994. Old Javanese influence in Balinese: Balinese speech styles. In *Language Contact and Change in the Austronesian World*, Tom E. Dutton & Darrell T. Tryon (eds), 141–179. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110883091.141

Cole, Peter, Hara, Yurie & Yap, Ngee Thai. 2008. Auxiliary fronting in Peranakan Javanese. *Linguistics* 44: 1–43. https://doi.org/10.1017/S002222670700494X

Cole, Peter, Hermon, Gabriella, Kim, Chonghyuck, Sim, Chang-Yong, & Tsai, Yaping. 2003a. Anaphoric Expressions in Javanese. In *Proceedings of AFLA VIII: The Eighth Meeting of the Austronesian Formal Linguistics Association* [MITWPL 44], Andrea Rakowski & Norvin Richards (eds). Cambrige MA: Dept. of Linguistics & Philosophy, MIT. <https://stuff.mit.edu/afs/athena/activity/m/mitwpl/pre2006site/WPL-volumes-files/WPL44.html> (1 June 2020).

Cole, Peter, Hermon, Gabriella, Inoha, Kozue, & Tjung, Yassir. 2003b. A constraint on WH in situ in Javanese. In *Proceedings of AFLA VIII: The Eighth Meeting of the Austronesian Formal Linguistics Association* [MITWPL 44], Andrea Rakowski & Norvin Richards (eds). Cambridge MA: Dept. of Linguistics & Philosophy, MIT. <https://stuff.mit.edu/afs/athena/activity/m/mitwpl/pre2006site/WPL-volumes-files/WPL44.html> (1 June 2020).

Cole, Peter, Jonczyk, Elizabeth & Lilley, Jason. 2000. A note on extraction from object position in Javanese and other Javanic languages. In *Proceedings of the Sixth Meeting of the Austronesian Formal Linguistics Association* [Toronto Working Papers in Linguistics, Vol. 16 No. 2], 87–93. Toronto: University of Toronto.

Comrie, Bernard. 2009. *Areal Typology of Southeast Asian Languages: Evidence from the World Atlas of Language Structures*. UC Los Angeles: UCLA Center for Southeast Asian Studies. <http://www.escholarship.org/uc/item/8wn0t0g0:> (1 June 2020).

Conners, Thomas. 2008. Tengger Javanese. PhD Dissertation, Yale University.

Conners, Thomas J. 2010. Death of linguistic subsystems: Javanese speech levels. Paper presented at *Humanities of the Lesser Known*, Lund University, Lund, Sweden, 10–11 September 2010.

Conners, Thomas & Brugman, Claudia. 2013. The event semantic role of the nasal prefix in Jakarta Indonesian. In *Proceedings of the Twelfth International Conference on Austronesian Linguistics*, Vol. 2, I. Wayan Arka & N. L.K Mas Indrawati (eds), 213–225. Canberra: Pacific Linguistics.

Conners, Thomas, Brugman, Claudia & Adams, Nikki. 2016. Reference tracking and noncanonical referring expressions in Indonesian and other discourse pro-drop languages. *NUSA: Linguistic studies of languages in and around Indonesia* 60(1): 58–88.

Dryer, Matthew S. 2011. Position of polar question particles. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), Chapter 92. Munich: Max Planck Digital Library. <http://wals.info/chapter/60> (1 June 2020).

Enfield, Nicholas J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 31(1): 181–206. https://doi.org/10.1146/annurev.anthro.34.081804.120406

Enfield, Nicholas J. & Comrie, Bernard (eds). 2015. *Languages of Mainland Southeast Asia. The State of the Art*. Berlin: Mouton de Gruyter.

Errington, J. Joseph. 1985. *Language and Social Change in Java. Linguistic Reflexes of Modernization in a Traditional Royal Polity*. Athens OH: Ohio University Press.

Errington, J. Joseph. 1988. *Structure and Style in Javanese. A Semiotic View of Linguistic Etiquette*. Philadelphia PA: University of Pennsylvania Press.

Errington, J. Joseph. 1998. *Shifting Languages: Interaction and Identity in Javanese Indonesia*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511612480

Ewing, Michael C. 2005. *Grammar and Inference in Conversation: Identifying Clause Structure in Spoken Javanese* [Studies in Discourse and Grammar 18]. Amsterdam: John Benjamins. https://doi.org/10.1075/sidag.18

Flannery, Greg. 2009. Open and closed systems of self-reference and addressee-reference in Indonesian and English: A broad typological distinction. In *Selected Papers from the 2009 Conference of the Australian Linguistic Society*, Yvonne Tries & Rik De Busser (eds). Sydney: ALS.

Gil, David. 2011. Genitives, adjectives and relative clauses. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), Chapter 60. Munich: Max Planck Digital Library. <http://wals.info/chapter/60> (1 June 2020).

Gil, David. 2015. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia. The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: Mouton de Gruyter.

Goddard, Cliff. 2005. *The Languages of East and Southeast Asia: An Introduction*. Oxford: OUP.

Hoogervorst, Tom. 2010. East Javanese Dialects. MA thesis, Leiden University.

Horne, Elinor. 1961. *Beginning Javanese* [Yale Linguistic Series 3]. New Haven CT: Yale University Press.

Horne, Elinor. 1974. *Javanese-English Dictionary* [Yale Linguistic Series]. New Haven CT: Yale University Press.

Hunter, Thomas. 1999. *A grammar of Old Javanese*. Ms, Udayana University.

Keeler, Ward. 1984. *Javanese: A Cultural Approach* [Ohio University Monographs in International Studies, Southeast Asia Series 69]. Athens OH: Ohio University.

Maddieson, Ian. 2011. Tone. *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), Chapter 13. Munich: Max Planck Digital Library. <http://wals.info/chapter/13> (1 June 2020).

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In *Southern California Occasional Papers in Linguistics No. 1*, Larry M. Hyman (ed), 72–95. Los Angeles CA: University of Southern California, Dept. of Linguistics.

Migliazza, Brian. 1996. Mainland Southeast Asia: A unique linguistic area. *Notes on Linguistics* 75: 17–25.

Oetomo, Dede. 1990. The Bahasa Indonesia of the middle class. *Prisma: Indonesian Journal of Social and Economic Affairs* 50: 68–79.

Ogloblin, Alexander K. 2005. Javanese. In *The Austronesian Languages of Asia and Madagascar*, K. Alexander Adelaar & Nicholaus P. Himmelmann (eds), 590–624. Chapter 21. London: Routledge.

Poedjosoedarmo, Soepomo. 1968. Javanese speech levels. *Indonesia* 6: 54–81.
https://doi.org/10.2307/3350711

Robson, Stuart. 1992. *Javanese Grammar for Students*. Melbourne: Monash Asia Institute.

Robson, Stuart & Wibisono, Singgih. 2002. *Javanese-English Dictionary*. Hong Kong: Periplus.

Senft, Gunter (ed). 2008. *Systems of Nominal Classification* [Language, Culture & Cognition 4]. Cambridge: CUP.

Siebenhütter, Stefanie. 2018. Study of linguistic areas: Evidence from cultural words, semantic maps, and spatial reference in Southeast Asia. In *Handbook of the Changing World Language Map*, Stanley D. Brunn & Roland Kehrein (eds), 27–43. Cham: Springer.
https://doi.org/10.1007/978-3-319-73400-2_83-1

Sigit, Haris Triono & Sulistiyono. 2017. Application for learning Javanese language of Banten using Knutt Morris Pratt Algorithm. *International Journal of Advanced Research in Computer Science* Vol. 8.8: 495–498.

Simons, Gary F. & Fennig, Charles D. (eds). 2018. *Ethnologue: Languages of the world*, 21st edn. Dallas TX: SIL International.

Smith-Hefner, Nancy J. 2009. Language shift, gender, and ideologies of modernity in Central Java, Indonesia. *Journal of Linguistic Anthropology* 19(1): 57–77.
https://doi.org/10.1111/j.1548-1395.2009.01019.x

The World Factbook 2016–17. Washington DC: Central Intelligence Agency. <https://www.cia.gov/library/publications/the-world-factbook/index.html> (1 June 2020).

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change* [Oceanic Linguistics Special Publication 28]. Honolulu HI: University of Hawai'i Press.

Uhlenbeck, Eugenius M. 1978. *Studies in Javanese Morphology*. The Hague: Martinus Nijhoff.

Uhlenbeck, Eugenius M. 1983. *Javanese Linguistics*. Dordrecht: Foris.

Vander Klok, Jozina. 2012. Tense, aspect, and modal markers in Paciran Javanese. PhD dissertation, McGill University.

Vander Klok, Jozina. 2015. The dichotomy of auxiliaries in Javanese: evidence from two dialects. *Australian Journal of Linguistics* 35(2): 142–167.
https://doi.org/10.1080/07268602.2015.1004999

Vander Klok, Jozina. 2018. Types of polar questions in Javanese. *NUSA* 63: 1–64.

Vander Klok, Jozina & Conners, Thomas J. 2019. Using questionnaires as a tool for comparative linguistic field research: Two case studies on Javanese. In *Language Documentation & Conservation* [Special Publication 16, Methodological Tools for Linguistic Description and Typology], Aimée Lahaussois & Marine Vuillermet (eds), 62–96. Mano'a HI: University of Hawaii Press. <https://scholarspace.manoa.hawaii.edu/handle/10125/24858> (1 June 2020).

Villerius, Sophie. 2017. Developments in Surinamese Javanese. In *Boundaries and Bridges: Language Contact in Multilingual Ecologies*, Kofi Yakpo & Pieter C. Muysken (eds), 151–178. Berlin: Walter de Gruyter. https://doi.org/10.1515/9781614514886-006

Villerius, Sophie. 2019. Development of Surinamese Javanese: Language Contact and Change in a Multilingual Context. PhD dissertation, Radboud University, Nijmegen.

Wardhaugh, Ronald. 2015. *An Introduction to Sociolinguistics*, 7th edn. Oxford: Blackwell.

Wittke, Jonas. 2019. Osing Javanese. PhD dissertation, Rice University.

Yap, Foong Ha & Iwasaki, Shoichi. 2007. The emergence of 'give' passives in East and Southeast Asian languages. In *SEALS VIII: Papers from the 8th Annual Meeting of the Southeast Asian Linguistics Society*, Mark Alves, Paul Sidwell & David Gil (eds), 193–208. Canberra: Pacific Linguistics.

Zoetmulder, Petrus J. 1983. *Old Javanese-English Dictionary*. Amsterdam: Springer.

Zentz, Lauren. 2015. Moving languages: Syncretism and shift in Central Java. In *Tilburg Papers in Culture Studies No. 141*, Zane Goebel, Deborah Cole & Howard Manns, (eds), 51–63. Tilburg: Tilburg University.

Are the Central Flores languages really typologically unusual?

Alexander Elias
University of California, Berkeley

The languages of Central Flores (Austronesian) are typologically distinct from their nearby relatives. They have elaborate numeral classifier systems, quinary numeral systems, and lack all bound morphology. McWhorter (2019) proposes that their isolating typology is due to imperfect language acquisition of a Sulawesi language, brought to Flores by settlers from Sulawesi in the relatively recent past. I propose an alternative interpretation, which better accounts for the other typological features found in Central Flores: the Central Flores languages are isolating because they have a strong substrate influence from a now-extinct isolating language which belonged to the Mekong-Mamberamo linguistic area (Gil 2015). This explanation better accounts for the typological profile of Central Flores and is a more plausible contact scenario.

Keywords: Central Flores languages, Eastern Indonesia, isolating languages, Mekong-Mamberamo linguistic area, substrate influence

1. Introduction

The Central Flores languages (Austronesian: Central Malayo-Polynesian) are a group of serialising SVO languages with obligatory numeral classifier systems, spoken on the island of Flores, one of the Lesser Sunda Islands in the east of Indonesia. These languages, which are almost completely lacking in bound morphology, include Lio, Ende, Nage, Keo, Ngadha and Rongga. Taken in their local context, this typological profile is unusual: other Austronesian languages of eastern Indonesia generally have some bound morphology and non-obligatory numeral classifier systems. However, in a broader view, the Central Flores languages are typologically similar to many of the isolating languages of Mainland Southeast Asia and Western New Guinea, many of which are also isolating, serialising SVO languages with obligatory numeral classifier systems.

The question of how the Central Flores languages became isolating is addressed by McWhorter (2019): he claims that Central Flores languages lost their morphology because they were acquired by a large number of adult speakers in the relatively recent past, perhaps arriving from Sulawesi. Under this account, the Central Flores languages lost their morphology due to imperfect learning and simplification by adult learners, and are an instance of a more general process which is exemplified by creole languages.

However, an explanation based on simplification alone cannot account for the other typological parallels between Central Flores, Mainland Southeast Asia and Western New Guinea, particularly the presence of complex numeral classifier systems. Gil (2015) has proposed a Mekong-Mamberamo (MM) linguistic area spanning Mainland Southeast Asia, the Indonesian Archipelago and Western New Guinea, based on 17 shared typological features. The Austronesian family is a relative newcomer to the Mekong-Mamberamo linguistic area, and has displaced the genealogically diverse MM-type languages in most parts of the Indonesian Archipelago. However, these pre-Austronesian languages have left their mark on the modern Austronesian languages of Indonesia to varying degrees. If the Mekong-Mamberamo hypothesis is correct, it provides a more economical explanation for the typology of the Central Flores languages.

In this paper, I argue that McWhorter's proposed scenario of relatively recent contact with Sulawesi is implausible, and the typology of Central Flores languages is better explained as a reflection of a substrate language with Mekong-Mamberamo typology. This substrate is stronger in Central Flores than in East and West Flores, reflecting differing contact conditions between the Austronesian settlers and the pre-Austronesian population at the time of the original Austronesian settlement of Flores between 2500–1500 BCE (Bellwood 1997).

The structure of this paper is as follows. The next section gives the theoretical background (Section 2): I will introduce the Mekong-Mamberamo proposal which will guide the rest of the paper (Section 2.1) and outline McWhorter's view on isolating languages (Section 2.2), McWhorter's stance on language complexity (Section 2.3) and his historical scenario for the development of the Central Flores languages (Section 2.4). Section 3 is a typological overview of the languages of West Flores (Section 3.1), East Flores (Section 3.2) and an introduction to the languages of Central Flores (Section 3.3). Section 4 makes up the bulk of the paper: in this section, I examine the list of Mekong-Mamberamo features and illustrate their presence or absence in the languages of Flores. I then introduce additional relevant data about the numeral systems of Central Flores (Section 5) before offering my own historical interpretation of the data (Section 6) and finishing with a conclusion (Section 7).

2.   Theoretical background

In this section, I will lay the theoretical groundwork needed to interpret the data presented in Section 4. First, I briefly describe the Mekong-Mamberamo linguistic area proposal (put forth in Gil 2015) and introduce the features which he identifies as typical of the area. These features will be defined and explained more fully in Section 4 when I address their presence or absence in the languages of Central Flores. After introducing the Mekong-Mamberamo proposal, I outline the theoretical framework from which McWhorter (2019) approaches the question of Central Flores typology. I outline his thinking on how languages become simplified by adult language acquisition (Section 2.2), the criteria he proposes to evaluate linguistic complexity (Section 2.3) and his historical proposal for the Central Flores languages (Section 2.4).

2.1   The Mekong-Mamberamo language area

Based on typological similarities between the languages of Mainland Southeast Asia, the Indonesian Archipelago and Western New Guinea, Gil (2015) proposes the existence of the Mekong-Mamberamo linguistic area defined by the 17 typological features listed in Table 1.

Table 1. List of Mekong-Mamberamo typological features (Gil 2015: 267)

1	passing gesture
2	repeated dental clicks expressing amazement
3	conventionalised greeting with 'where'
4	'eye day' → 'sun' lexicalisation
5	d/t place-of-articulation asymmetry
6	numeral classifiers
7	verby adjectives
8	basic SVO constituent order
9	iamitive perfects
10	'give' causatives
11	low differentiation of adnominal attributive constructions
12	weakly developed grammatical voice
13	isolating word structure
14	short words
15	low grammatical-morpheme density
16	optional thematic-role flagging
17	optional TAM marking

Gil (2015) proposes that the typological similarities between Mekong-Mamberamo languages reflect an ancient pattern of cultural contact across the area, leading to linguistic convergence. The Mekong-Mamberamo area is proposed to be of great antiquity, pre-dating the arrival of the Austronesians in the Indonesian Archipelago around 2500–1500 BCE (Bellwood 1997). When the Austronesians arrived in the Mekong-Mamberamo linguistic area, they brought with them a distinctly non-Mekong-Mamberamo type language: verb-initial, with copious morphology and a well-developed system of grammatical voice. This typological profile describes many of the modern Austronesian languages of Taiwan and the Philippines, and is the typological profile inherited from Proto-Austronesian (Blust 2013).

However, the Austronesian speakers who went south to the Indonesian Archipelago and spread east and west encountered speakers of Mekong-Mamberamo type languages where they settled. Eventually, almost all pre-Austronesian Mekong-Mamberamo type languages of the Indonesian Archipelago were displaced by Austronesian languages, but in the process they were restructured to fit the Mekong-Mamberamo typological profile to varying degrees. The degree to which any particular Austronesian language was restructured must have depended on a number of factors, including the ratio of settlers to local population on an island, the intensity and nature of the contact between them, and the social relationships between the settlers and the locals.

Thus, the Mekong-Mamberamo typological features display a saddle-shaped geographical distribution in many cases: they are most common in Mainland Southeast Asia and Western New Guinea area, with a patchier distribution across the Indonesian Archipelago. This is because of the incomplete restructuring of many Austronesian languages, which displaced the pre-Austronesian languages without fully conforming to the Mekong-Mamberamo typological profile yet.

The argument which I will develop in this paper is that the Central Flores languages are an example of a particularly heavily restructured group of Austronesian languages which have conformed almost totally to the Mekong-Mamberamo profile. Thus, they appear typologically unusual relative to many other Austronesian languages, but appear typologically well-behaved when seen as part of the Mekong-Mamberamo linguistic area. The typological differences between Central Flores and other parts of Flores reflect differences in the circumstances of contact between the Austronesian settlers and the pre-existing population (see Section 6). Various other lines of evidence, such as the Central Flores numeral systems (see Section 5) back up this scenario.

2.2   McWhorter's view of isolating languages

The account of Central Flores typology outlined above, and which I will argue for in this paper, stands in contrast to McWhorter (2019), who also seeks to account for the isolating typology of the Central Flores languages. He proposes that a group of settlers from Sulawesi, perhaps speaking a language ancestral to Tukang Besi, arrived in Flores in the relatively recent past (i.e., once Flores was already inhabited by Austronesian speakers). Their language was adopted by the pre-existing Austronesian speakers of Central Flores and was acquired by large numbers of adults. The process of imperfect adult language acquisition at that time resulted in the loss of bound morphology as adult speakers simplified the grammar. Thus, in McWhorter's view, the Central Flores languages emerged from the simplification of Tukang Besi or some other language of Sulawesi, driven by imperfect adult language acquisition. Central Flores isolating typology is the result of general cognitive processes at play whenever a language is imposed on a group of adult speakers (with creole languages best exemplifying this process of simplification).

McWhorter has developed an argument in a series of publications (2001, 2007, 2008, 2011, 2016, 2019) that highly isolating languages do not come about in situations of unbroken language transmission, but must always be the product of an episode of intense contact leading to imperfect adult language acquisition. In McWhorter's words, "isolating typology *signals* heavy adult acquisition in a language's past, rather than merely *suggesting* it." (McWhorter 2019: 193, emphasis in original). McWhorter's claim is as follows: because there is no other diachronic mechanism by which languages achieve such a totally isolating morphosyntactic profile, all 'radically' isolating languages must logically have undergone an episode of intense contact in the past which led to their current typological profile.

This is related to the distinction between 'esoteric' and 'exoteric' languages drawn by Thurston (1987): 'esoteric' languages are used only for communication within a small and tightly-knit group, where adult acquisition of the language is rare, while 'exoteric' languages are used for intergroup communication and as such, are commonly learned by adults. The process of adult learning strips away user-unfriendly opacities such as suppletion, irregularities and complex morphophonological alternations, and leaves an 'exoteric' language with less overall complexity than its 'esoteric' sisters.

In keeping with this view, McWhorter claims that abundant affixal morphology is the 'natural state' of language when it is transmitted uninterrupted between generations, given the vast learning capacity of infants. As the argument goes, it is inevitable that irregularities and opacities will accrue in a language which is learned only by infants because they have no strong need to restore systematicity. On the

other hand, adult learners will seek to extend regularities and reduce opacities, because their language learning capacity is severely limited compared to that of infants.

The implication of this line of argument is that isolating languages do not stay isolating for long under regular conditions of intergenerational transmission. This serves as a kind of linguistic timer: when faced with an isolating language, one must not only posit a contact event, but it must be of rather recent date.

In this paper, I seek to show instead that the predictive strength of McWhorter's hypothesis (i.e., simple languages only ever arise due to imperfect adult language learning) leads him to propose an unsound historical scenario of recent contact with Sulawesi to account for the typology of the Central Flores languages. McWhorter's hypothesis predicts that any changes occurring as a result of imperfect adult language learning tend towards simplification as he defines it. In this case, it is difficult for McWhorter's explanation to account for the development of an elaborate system of classifiers in the Central Flores languages, which is more complex than that of most Austronesian languages.

2.3   McWhorter's definition of linguistic complexity

In order to formalise his argument about the relative complexity of languages, McWhorter has attempted to measure linguistic complexity – a notoriously difficult task – along three axes (2007: 21–35):

1. Overspecification: "Languages differ in the degree to which they overtly and obligatorily mark semantic distinctions" (McWhorter 2007: 21). A language is more complex to the extent that it requires overt marking of person and number, noun class, definiteness, evidentiality, clusivity, tense, aspect, mood, etc…
2. Structural Elaboration: "An aspect of one grammar may differ from that aspect in another's in terms of the number of rules (in phonology and syntax) or foundational elements (in terms of phonemic inventory) required to generate surface forms" (McWhorter 2007: 29). A language is more complex to the extent that there are more unpredictable morphophonemic alternations, a larger phonemic inventory, more inflectional classes, word order alternations, etc…
3. Irregularity: "Grammars differ in the degree to which they are festooned with irregularity and suppletion" (McWhorter 2007: 33). A language is more complex to the extent that its noun class system has arbitrary assignment, various unpredictable plural marking strategies, suppletion in its conjugational system, etc…

The purpose of this paper is not to dispute the fact that the Central Flores languages are relatively simple by McWhorter's metric of complexity (for a refutation of McWhorter's claim that certain languages of Timor are unusually simple by his own metric, see Schapper, this volume). They do indeed stand out in their local context as unusually isolating and devoid of opacities and irregularities.

However, many scholars would dispute McWhorter's complexity criteria, and much ink has been spilled trying to argue for and against various interpretations of linguistic complexity. To take one example, Fenk-Oczlon and Fenk (2008: 56) point to the fact that creole languages, often having simple morphology and phonology, tend towards a high level of polysemy and homophony. The Central Flores languages certainly tend towards polysemy, such as between intransitive and transitive uses of verbs (see the end of Section 4.7). Under McWhorter's definition, that contributes to overall simplicity because valency changes are not overtly marked on verbs. However, Fenk-Oczlon and Fenk (2009) point out that this massively increases the semantic complexity of the language, since each polysemous lexical item must still be associated with the proper range of possible constructions somehow.

2.4 McWhorter's proposed historical scenario

In keeping with his views on the origins and development of isolating languages, McWhorter seeks to explain the typology of Central Flores by reference to either (1) a relatively recent migration from Sulawesi to Flores by speakers of a language similar or ancestral to Tukang Besi, or (2) contact with *Homo floresiensis*, a species of small hominid recently described from a handful of skeletons found in a cave in northwestern Flores dated to around 12,000 years ago (Brown et al. 2004). This second option is rather fanciful, and even assuming that Austronesian speakers co-existed with *Homo floresiensis* at some point, this would require that the imperfectly acquired speech of *Homo floresiensis* then became the dominant language of the entire community, even as they were pushed to extinction by modern humans. In my opinion, it is safe to lay the *Homo floresiensis* idea to rest, but the first scenario deserves a more careful look.

In this context, 'relatively recent contact' means that the contact occurred well after the initial contact between the incoming Austronesians and the earlier non-Austronesian ('Papuan') population of Flores, which may be placed between 2500–1500 BCE (Bellwood 1997). The occurrence of this initial contact is uncontroversial, because Flores was certainly inhabited at the time of the Austronesian settlement. However, McWhorter believes that little or nothing can be recovered about the earlier non-Austronesian languages of Flores, and that any proposals about them will ultimately lead to a scientific dead end:

> One might propose that the central Flores languages became isolating in contact with now-extinct Papuan languages that were also isolating. This is reasonable – but a scientific dead end … These hypothetical isolating Papuan languages of Flores could only remain, therefore, an unverifiable surmise, whereas this paper is an attempt to assign a more systematic and refutable explanation to the facts.
>
> (McWhorter 2011: 252)

Thus, to explain the isolating typology of Central Flores, McWhorter departs from the consensus to propose a second episode of contact, where Austronesian speakers from Sulawesi migrated to Flores, then shifted to the local languages but left traces of their imperfect adult language acquisition in the isolating typology of the modern Central Flores languages. However, as McWhorter himself points out, "lexical and grammatical data in support of this scenario are lacking" (McWhorter 2019: 195). Indeed, there is a conspicuous absence of parallels between the languages of Central Flores and Sulawesi in lexicon, grammar and phonology. This absence is all the more conspicuous because the newcomers from Sulawesi would almost certainly have been in a socially dominant position over the local population, given that they were economically and technologically advanced enough to launch an overseas expedition.

In addition, numerous rulers in the region have traditionally established their claim to legitimacy by reference to foreign origin, including the Sika-speaking kingdom of East Flores (cf. Lewis 2010 *The Stranger-kings of Sikka*). The theme of an immigrant 'stranger-king' or 'xenarch' arriving from overseas and establishing a dynasty is prevalent throughout the eastern Lesser Sunda Islands (Lewis 2010), so there is precedent for the notion that at least some groups of outside settlers enjoyed a high level of prestige in the area. Following the predictions of Thomason and Kaufman (1988), we would expect to find many lexical traces of the dominant group – the situation would be analogous to the Norman conquest of England by socially dominant but numerically inferior French speakers, leading to the shift of French speakers to English but with heavy lexical influence of French on the resulting English language. The alternative, that the arrivals from Sulawesi became integrated as the equals or the inferiors of the local population, is less plausible.

The evidence adduced by McWhorter in favour of this contact with Sulawesi is rather circumstantial, from history and folklore, and is hardly the smoking gun which allows us to draw a direct link between Sulawesi and Flores:

> The Gowa empire of southwestern Sulawesi controlled the Manggarai region of Flores from 1658 to 1750, and many Manggarai trace their ancestry to migrations from Gowa on the southwestern leg of Sulawesi (Erb 1999: 85–86). One of the ancestor stories of the Nage involves invaders from Gowa as well (Forth 1998: 230) and their cosmology traces them in general to either Sulawesi or 'Bugis bonerate'. Manggarai and Nage people also trace ancestry to what they term the Minangkabau (Erb 1999: 85; Forth 1998: 81) but Van Bekkum (1944) documented the alternate term 'Bonengkabau', suggesting that 'Minangkabau' may be a folk distortion of an actual descent from the more geographically plausible region of the Gulf of Bone between the southwestern and southeastern legs of Sulawesi.   (McWhorter 2019: 194)

A lexical line of evidence, originally put forth by Hull (1998) while proposing a migration from Sulawesi to Timor, is taken up in McWhorter (2019) and adapted to the Flores context. This argument states that the rate of cognate matches between

Sulawesi and Flores is very high, and that the forms of the cognates are so similar that they cannot have been separated for 3000 years without a fresh injection of Sulawesi lexicon into the Flores languages. In other words, pairs of languages separated for that long should undergo more evolution from their common source than is actually observed. In an earlier work, he cites pairs from Tukang Besi and languages of Flores such as the those listed in Table 2, reproduced from McWhorter (2011: 241) with a few minor errors in the Ende, Rongga and Ngadha data corrected. The Proto-Malayo-Polynesian forms on the right have been added by me, drawn from the Austronesian Comparative Dictionary (Blust and Trussel 2019):

Table 2. Lexical similarities between Tukang Besi and Flores are retentions, not innovations

Gloss	PMP	T. Besi	Sika	Ende	Ngadha	Rongga	Keo
'come'	*maRi	*mai*	*mai*	*mai*	*mai*	*mai*	*maʔi*
'dead'	*matay	*mate*	*mate*	*mata*	*mata*	*mata*	*mata*
'fish'	*hikan	*ika*	*iaŋ*	*ʔika*	*ika*	*ika*	*ʔika*
'fowl'	*manuk	*manu*	*manu*	*manu*	*manu*	*manu*	*manu*
'liver'	*qatay	*ate*	*wateŋ*	*ʔate*	*ate*	*ate*	*ʔate*
'pig'	*babuy	*wawu*	*wawi*	*wawi*	*wawi*	*wawi*	*wawi*
'rain'	*quzan	*usa-*	*uran*	*ʔura*	*uza*	*nuɭa*	*ʔura*
'stone'	*batu	*watu*	*watu*	*watu*	*watu*	*watu*	*watu*

This argument is flawed because it rests on shared retentions to support claims about historical relatedness. The forms cited by McWhorter are minimally changed from Proto-Malayo-Polynesian, and hundreds of additional examples of languages with similar forms could be adduced from across the Austronesian family. Shared retentions can never be taken as subgrouping evidence in orthodox comparative linguistics; conservative languages are similar because of their relation to their common ancestor, not to each other. In order for this evidence to support a link between Tukang Besi and Flores, it would be necessary for McWhorter to show that the lexical similarities are in fact innovations, which they are not.

Another argument which McWhorter marshals in support of a recent contact hypothesis is the clinality of isolating languages in Flores: he states that the languages of Flores become less isolating in a cline to the west and east. He suggests that this shows that there was total loss of affixation at one place (the landing site of the invaders from Sulawesi, presumably) which radiated outwards to neighbouring languages with less and less intensity. This is in fact not the case: the isolating languages of Central Flores form a well-defined clade, and there is ample evidence that the isolating profile of these languages can be reconstructed to their common ancestor, Proto-Central Flores (Elias 2018). Within the Central Flores-speaking area, isolating morphology is the rule, but the borders of this area to the west and east are sharp, not a gradient as McWhorter suggests. Although it is true that Sika is somewhat less

complex than Lamaholot, it still retains a system of verbal conjugation and other morphological complexities which put it in a separate class from the Central Flores languages typologically. The pattern indicates that the current distribution of isolating morphology in Flores is not because of diffusion through contact, but rather because of common descent from a single, highly isolating Proto-Central Flores ancestor. This shows up very clearly in the linguistic data as an easily reconstructible node at Proto-Central Flores with well-defined bundles of isoglosses delimiting the boundaries of Central Flores (see Section 3.3 for a list of innovations).

3.   Introduction to the languages of Flores

The island of Flores is part of the Lesser Sunda Islands chain, located in the east of Indonesia (Nusa Tenggara Timur Province). Flores has a population of nearly two million as of the 2010 Indonesian census, and these people all speak Austronesian languages of the Central Malayo-Polynesian (CMP) group.

Linguistically, Flores can be divided into three approximately equal sections: West, East and Central Flores. West Flores is dominated by Manggarai (with a few poorly known languages similar to Manggarai spoken on the peripheries) while East Flores is populated by speakers of Sika and Lamaholot. Across Central Flores stretches the Central Flores Linkage: Lio, Ende, Nage, Keo, Ngadha and Rongga. Figure 1 shows a map of the languages of Flores.

Blust (2008) finds some evidence that the languages of Central Flores subgroup with West Flores (Manggarai) and the languages of nearby Sumba and Hawu, in a primary branch of CMP dubbed 'Flores-Sumba-Hawu', while East Flores (Sika, Lamaholot on East Flores plus Kedang, Alorese on neighbouring islands) belongs to a separate branch of CMP dubbed 'Flores-Lembata'. Fricke (2019: 229) presents evidence that Flores-Sumba-Hawu and Flores-Lembata form a higher-order subgroup along with Bima, a group called 'Bima-Lembata'.
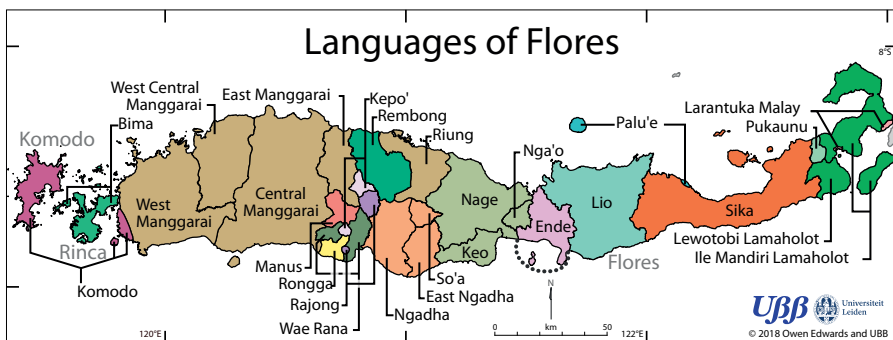

Figure 1.   Map of the Flores languages (created by Owen Edwards, reproduced with permission)

3.1 Languages of West Flores: Manggarai

The western third of Flores is dominated by Manggarai, the largest CMP language by number of speakers. There are a few other poorly known languages in West Flores (Kepo', Rembong, Riung, Manus, Rajong, Wae Rana) which appear to be close to Manggarai based on lexical inspection but remain largely uninvestigated. There is also the better-known Komodo language spoken in the very west of Flores and on neighbouring Komodo Island. For the purposes of this paper, Manggarai will represent all of West Flores, because Komodo is known to be very similar, and the other languages are too scantily known to draw conclusions from. The bulk of the work on these languages has been carried out by Verheijen, a Dutch linguist, including his monumental Manggarai dictionary (Verheijen 1967) and sketch grammar of Komodo (Verheijen 1982).

The Central Manggarai variety described by Semiun (2017) has a modest amount of morphology. It has a set of enclitics which cross-reference the subject of the verb, shown in Table 3. These enclitics need not attach to the verb, but typically attach to the last word of the clause, hence their status as clitics.

Table 3. Manggarai subject-marking clitics

Person	Form	Gloss
1SG	aku haŋ=k	'I eat'
2SG	hau haŋ=h	'you eat'
3SG	hia haŋ=i	'he/she eats'
1PL.INC	ite haŋ=t	'we (inc.) eat'
1PL.EXC	ami haŋ=km	'we (exc.) eat'
2PL	meu haŋ=m	'you (pl.) eat'
3PL	ise haŋ=s	'they eat'

There is a second set of possessive enclitics indicating the possessor, and which differ in some cases from the subject-marking enclitics, as shown in Table 4.

Table 4. Manggarai possessive clitics

Person	Form	Gloss
1SG	mbaru=k	'my house'
2SG	mbaru=m	'your house'
3SG	mbaru=n	'his/her house'
1PL.INC	mbaru=t	'our (inc.) house'
1PL.EXC	mbaru=km	'our (exc.) house'
2PL	mbaru=s	'your (pl.) house'
3PL	mbaru=d	'their house'

Manggarai uses a decimal number system, with familiar Austronesian numerals: *tʃa* 'one', *sua* 'two', *təlu* 'three', *pat* 'four', *lima* 'five', *ənəm* 'six', *pitu* 'seven', *alo* 'eight', *tʃiok* 'nine', *pulu* 'ten' (Verheijen 1967).

Manggarai has numeral classifiers, but these are not obligatory when using numerals:

Manggarai

(1) *pulu wuŋkut*
    ten   knuckle
    'ten knuckles'   (Verheijen 1967: 8)

3.2   Languages of East Flores: Sika, Lamaholot

East Flores has two languages, Sika and Lamaholot, which belong together in the Flores-Lembata subgroup of CMP. These languages show a significant amount of morphological complexity, both inflectional and derivational.

The data presented on Sika is drawn from Fricke (2013), which describes aspects of the grammar of Hewa, an eastern variety of Sika. There is also a Sika dictionary available (Pareira & Lewis 1998).

A subset of Sika verbs is conjugated with an initial consonant cross-referencing the subject (see Table 5). The full set of conjugations appears only with vowel-initial verb roots. In consonant-initial verb roots, the initial consonant undergoes alternations which are mostly predictable from the voicing of the subject marker found in vowel-initial words. There is one irregular verb /ʔa/ 'to eat'. This is not the full extent of verbal conjugation in Sika, but it gives a flavour of the type of alternations encountered.

Table 5. Sika (Hewa) verb conjugation

Person	Pronoun	/inu/ 'to drink'	/pano/ 'to go'	/ʔali/ 'to dig'	/ʔa/ 'to eat'
1SG	*aʔu*	*ʔ-inu*	*pano*	*ʔali*	*ʔoa*
2SG	*ʔau*	*m-inu*	*bano*	*gali*	*goa*
3SG	*nimu*	*n-inu*	*bano*	*gali*	*ga*
1PL.INC	*ʔita*	*t-inu*	*pano*	*ʔali*	*ʔea*
1PL.EXC	*ʔami*	*m-inu*	*bano*	*gali*	*gea*
2PL	*miu*	*m-inu*	*bano*	*gali*	*gea*
3PL	*rimu*	*r-inu*	*pano*	*ʔali*	*ʔa*

In possessive constructions, a morpheme /-n/ is added to the second member of the construction. This can be the possessor, as in the pronominal possessive construction [Noun + Pronoun]:

Sika (Hewa)

(2) *me   nimu-n*
     child 3SG-POSS
     'his/her child'   (Fricke 2013: 39)

The same possessive morpheme /-n/ can appear on the possessum, in the nominal possessive construction [Possessor + Noun]:

Sika (Hewa)

(3) *duʔa   ʔia   me-n*
     woman   DEM   child-POSS
     'that woman's child'   (Fricke 2013: 40)

Sika has a decimal numeral system, with mainly inherited Austronesian numerals: *ha* 'one', *rua* 'two', *təlu* 'three', *hutu* 'four', *lima* 'five', *əna* 'six', *pitu* 'seven', *walu* 'eight', *hiwa* 'nine', *pulu* 'ten'. The only numeral here which is not an inherited Austronesian form is *hutu* 'four', which is likely a loan from the Lio *sutu* 'four' with a regular change of /s/ to /h/ (Fricke 2019: 367).

Sika has numeral classifiers, but these are optional when using numerals. Thus, in the Sika construction *ʔita rua-t* 'the two of us', no numeral classifier is needed.

Sika (Hewa)

(4) *ʔita   rua-t*
     1PL.INC two-ATT
     'the two of us'   (Fricke 2013: 47)

The equivalent construction in a Central Flores language would be ungrammatical without a numeral classifier, as in the Lio example (5):

Lio

(5) **kita   rua*
     1PL.INC two
     Failed reading: '*the two of us'   (Ibu_Ferdy_Frogstory)

Lio

(6) *kita   imu   rua*
     1PL.INC CLASS.HUM two
     'the two of us'   (Ibu_Ferdy_Frogstory)

The other language of East Flores, Lamaholot, shows considerable internal diversity and is also spoken on the neighbouring islands of Solor, Adonara and Lembata. A number of dialects of Lamaholot, both on Flores and on neighbouring islands, have received significant linguistic attention. There is a dictionary of the Lewolema

dialect (Pampus 1999), a PhD thesis describing the Lewotobi dialect (Nagaya 2011), a PhD thesis describing the Central Lembata dialect (Fricke 2019), a description of the morphology of the Lamalera dialect (Keraf 1978), a grammar of the Lewoingu dialect (Nishiyama & Kelen 2007), and a sketch grammar of the Solor dialect (Arndt 1937). The data presented in this section are drawn from Nagaya (2011) on the Lewotobi dialect, spoken in East Flores.

Lamaholot is the most morphologically complex language of Flores, with subject-marking prefixes on verbs, subject-marking enclitics, and possessive marking, among other morphology. Lewotobi Lamaholot uses subject-marking enclitics, shown in Table 6. These do not necessarily attach to the verb, hence their status as a clitic; but this need not concern us here.

Table 6. Lewotobi Lamaholot subject-marking clitics

Person	Form	Gloss
1SG	*go lega=əʔ*	'I walk'
2SG	*mo lega=ko*	'you walk'
3SG	*na lega=aʔ*	'he/she walks'
1PL.INC	*tite lega=kə*	'we (inc.) walk'
1PL.EXC	*kame lega=kə*	'we (exc.) walk'
2PL	*mio lega=kə*	'you (pl.) walk'
3PL	*ra lega=ka*	'they walk'

In addition to the subject-marking enclitics, there is a subset of vowel-initial verbs which take subject-marking prefixes (see Table 7). These often redundantly mark the subject in conjunction with the subject-marking enclitics (which here show an /n-/ initial form due to the preceding nasal vowel).

Table 7. Lewotobi Lamaholot verb conjugation

Person	Form	Gloss
1SG	*go k-enũ=nəʔ*	'I drink'
2SG	*mo m-enũ=no*	'you drink'
3SG	*na n-enũ=naʔ*	'he/she drinks'
1PL.INC	*tite t-enũ=nə*	'we (inc.) drink'
1PL.EXC	*kame m-enũ=nə*	'we (exc.) drink'
2PL	*mio m-enũ=nə*	'you (pl.) drink'
3PL	*ra r-enũ=na*	'they drink'

Lewotobi Lamaholot makes an alienability distinction in possessive constructions. In Lewotobi Lamaholot, a morpheme /-N/ surfaces on the second member in an inalienable possessive construction, and is realised as nasalisation on the final vowel.

Lamaholot (Lewotobi)

(7) *ika leĩ*
     ika lei-N
     Ika foot-POSS
     'Ika's foot'   (Nagaya 2011: 33)

In an alienably possessed construction, the second member is marked by an enclitic /=kɔ̃/ instead:

Lamaholot (Lewotobi)

(8) *ika doi=kɔ̃*
     Ika money=POSS
     'Ika's money'   (Nagaya 2011: 33)

In addition to this inflectional morphology, Lewotobi Lamaholot has a number of derivational affixes (not all of which are still productive) which make it significantly more morphologically complex than the other languages of Flores. For instance, there is a process of 'nasal substitution' in which a verb can be nominalised by replacing the initial consonant with a nasal (and possibly prefixing another consonant as well; see Table 8).

Table 8. Nominalisation by nasal substitution in Lewotobi Lamaholot

Verb	**Gloss**	**Noun**	**Gloss**
pətə	'to cut'	*mətə*	'cutting board'
bitu	'to fish with rod'	*mnitu*	'fishing rod'
dira	'to use a fan'	*mnira*	'fan'
giʔa	'to scratch'	*kniʔa*	'match'

The Lewotobi Lamaholot numeral system is mainly an inherited Austronesian decimal system: *toʔu* 'one', *rua* 'two', *təlo* 'three', *pa* 'four', *lema* 'five', *namu* 'six', *pito* 'seven', *buto* 'eight', *hiwa* 'nine', *pulo* 'ten'. The most interesting numeral here is *buto* 'eight', which is plausibly related to Proto-Central Flores *wutu 'four'. A semantic and formal parallel is found in Lamaholot's relative Kedang *butu rai* 'eight', where *rai* means 'many' (Fricke 2019: 367–368).

   Numeral classifiers are not obligatory when using numerals in Lewotobi Lamaholot:

Lamaholot (Lewotobi)

(9) *gula   rua*
     candy two
     'two pieces of candy'   (Nagaya 2011: 159)

3.3   Languages of Central Flores: Lio, Ende, Nage, Keo, Ngadha, Rongga

The closely related languages of Central Flores form a linkage across the central third of Flores and are very similar in their morphosyntactic structure. The differences between the modern Central Flores languages consist mainly of lexical differences and regular sound correspondences between phonemes. There is strong evidence that they form a clade, descending from Proto-Central Flores (Elias 2018). This evidence includes the loss of bound morphology and all coda consonants, the restructuring of the decimal numeral system into a mixed-base quinary-decimal system (see Section 5), as well as a hefty amount of innovative basic vocabulary (PCF *kobe 'night', *mbeʔo 'to know', *toro 'red', *ndate 'heavy', *kleu 'betel nut', *lidu 'sky', *koe 'to dig', *loŋgo 'back', *pawe 'good', *teʔe 'mat, *weʔe 'near') and a few shared semantic shifts (PMP *beRŋi 'night' becomes 'when?' and PMP *laku 'civet' becomes 'dog') (Elias 2018: 123–125).

The level of description of the Central Flores languages is uneven. Keo is described in a PhD thesis (Baird 2002) and Rongga has a grammar (Arka 2016) and dictionary (Arka et al. 2011). For Ngadha, there is a rather outdated grammar (Arndt 1933b) and dictionary (Arndt 1961). There is a dictionary of Lio (Arndt 1933a), as well as an undergraduate thesis on it (Levi 1978). Finally, there is an unpublished dictionary of Ende available only in electronic format (Aoki & Nakagawa 1993), as well as a description of Ende phonology (McDonnell 2009). Nage has not received much attention from linguists specifically, but Nage culture and folk classification has been the subject of numerous publications by the anthropologist Gregory Forth (Forth 1998, 2004, 2008, 2009, 2016).

There is a group of language varieties transitional between Ende and Keo referred to as Nga'o (shown on the language map, Figure 1) which may be divergent enough to be classified as a separate language, although it has been referred to as a dialect of Ende by previous researchers (Aoki & Nakagawa 1993).

Palu'e, spoken on an island of the same name off the northern coast of Flores, is very similar to the Central Flores languages and deserves a mention. It is described by a dictionary (Donohue 2003) and an analysis of the voice system (Donohue 2005). However, it falls outside of the scope of some of the key innovations that define the Central Flores clade: for instance, it retains a final /-n/ morpheme marking the genitive, so it has not undergone total loss of morphology and final consonants. It retains a typical Austronesian decimal numeral system as well, and has not developed the distinctive quinary-decimal numeral system of the Central Flores languages. Because it does not participate in these key Central Flores innovations, it falls outside of the Central Flores languages as I define them (Elias 2018) and is their closest external relative.

My own fieldwork has dealt mainly with Lio, and my MA thesis provides an analysis of Lio phonology, along with a comparative analysis of the Central Flores languages and reconstruction of aspects of Proto-Central Flores (Elias 2018). Note that the name Lio (/lio/) is often erroneously spelled Li'o, implying the presence of a glottal stop (*/liʔo/). There is no glottal stop in the name of the language, and the confusion arises from the counterintuitive convention used in Arndt's Lio dictionary (Arndt 1933a), where the *absence* of a glottal stop is indicated by an apostrophe.

The data presented in this paper to illustrate the typology of the Central Flores languages will be drawn from my own field data on Lio, as well as the other two Central Flores languages with full grammars: Keo (Baird 2002) and Rongga (Arka 2016). This provides a good geographic sampling of the Central Flores linkage (Lio is in the east, Rongga is in the west, Keo is in the middle). Given the varying level of documentation of the other languages, I will restrict myself to presenting data from these three languages. Given the close similarities between the Central Flores languages, I do not expect that additional data from Ende, Ngadha and Nage would substantially change the argument laid out here.

4.   Central Flores languages have typical Mekong-Mamberamo typology

In this section, I will show that the Central Flores languages in my sample (Lio, Keo, Rongga) show nearly all of the Mekong-Mamberamo features proposed by Gil (2015).

4.1   The passing gesture

The passing gesture, used when a person needs to pass through someone's personal space, is as follows: "while walking, the gesturer bends the top half of the body forward, and … extends the right forearm forward, with the hand oriented vertically, palm facing inward, as though forging a path through an imaginary thicket" (Gil 2015: 270).

I have observed Lio speakers employ the passing gesture when passing between two interlocutors in a conversation or passing through a crowded room. I do not have information on whether speakers of Rongga or Keo use the passing gesture, so I will mark them with '?' in the feature table below, but it is highly likely that they also employ this gesture.

	Lio	Keo	Rongga
The passing gesture:	+	?	?

4.2 Repeated dental clicks expressing amazement

In the Mekong-Mamberamo area, repeated dental clicks are used as a paralinguistic expression of amazement, usually with a positive affect. In contrast, most English speakers use repeated dental clicks to express disapproval, written as 'tsk tsk' in the USA and 'tut tut' in the UK.

Lio speakers do indeed use repeated dental clicks to express amazement, based on first-hand observation. I do not have information on whether speaker of Rongga or Keo use repeated dental clicks to express amazement, so I will mark them with '?' in the feature table, but it is highly likely that they also use them.

	Lio	Keo	Rongga
Repeated dental clicks expressing amazement:	+	?	?

4.3 Conventionalised greeting with "where?"

In the Mekong-Mamberamo area, conventionalised greetings tend to be formed with the question word 'where', as in Indonesian *Mau ke mana*, literally 'Where are you going?'. This is true for common conventionalised greetings both in Lio and in Keo:

Lio

(10) *mbana əmba*
    go   where
    'Where are you going?'   (Elias 2017 fieldnotes)

Keo

(11) *kau nuka ena ʔemba*
    2SG go.up LOC where
    'Where are you going?'   (Baird 2002: 440)

I do not have information on conventional greetings in Rongga, so I will mark it with '?' in the feature table.

	Lio	Keo	Rongga
Conventionalised greeting with 'where?':	+	+	?

4.4 'eye day' to 'sun' lexicalisation

The concept of 'sun' in the Mekong-Mamberamo area is often lexicalised as a collocation meaning something like 'eye of the day', as in Indonesian *mata hari* 'sun (lit. eye of day)'. This holds true in Lio, Rongga and Keo, and the form *mata lədʒa 'sun' is reconstructible to Proto-Central Flores (Elias 2018):

Proto-Central Flores

(12) *mata lədʒa*
  eye day
  'sun'   (Elias 2018)

Lio

(13) *mata lədʒa*
  eye day
  'sun'   (Elias 2017 fieldnotes)

Rongga

(14) *mata ləɹa*
  eye day
  'sun'   (Arka et al. 2011: 128)

Keo

(15) *mata dəra*
  eye day
  'sun'   (Baird 2002: 559)

All three of these languages have 'eye day' to 'sun' lexicalisation, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
'eye day' to 'sun' lexicalisation	+	+	+

4.5 d/t place of articulation asymmetry

An asymmetry in the place of articulation of the coronal stops /t/ and /d/ has been noted in many Mekong-Mamberamo languages: /t/ is dental (and more laminal), while /d/ is alveolar (and more apical).

  This mismatch in place of articulation is present in Lio (Elias 2018):

Lio

(16) /ata/
[ˈʔa.t̪ʰa]
'person' (Elias 2018: 19)

Lio

(17) /ada/
/ˈʔa.da]
'tradition, custom' (Elias 2017 fieldnotes)

There is no mention of a mismatch in place of articulation for Keo in Baird (2002), which lists both /t/ and /d/ as alveolar apical stops. There is also no mention of a mismatch for Rongga in Arka (2016) which lists both /t/ and /d/ as alveolar stops. I will mark them as '?' in the table of features, because this is a rather low-level phonetic feature that could easily be omitted in a grammar, so I am not certain of its absence in Keo and Rongga.

	Lio	Keo	Rongga
d/t place of articulation mismatch:	+	?	?

4.6 Numeral classifiers

A feature of Mekong-Mamberamo languages is the presence of a system of numeral classifiers, such as those famous from the Mainland Southeast Asian languages. Numeral classifiers are independent morphemes which occur when numerals modify nouns in an NP, and the choice of classifier depends on the noun.

There is a large and obligatory set of numeral classifiers in the Central Flores languages. Any time a numeral is used, whether attributively or predicatively, a classifier is obligatory.

Lio

(18) *saʔo* *əsa* *təlu*
house CLASS.GEN three
'three houses' (Ibu_Ferdy_History)

Lio

(19) *saʔo* *təlu*
house three
Failed reading: '*three houses' (Ibu_Ferdy_History)

Lio

(20) *ana kai kolo   sutu*
child 3SG CLASS.HUM four
'She has four children. (lit. Her children are four)'   (Elias 2017 fieldnotes)

Lio

(21) **ana   kai sutu*
child 3SG four
Failed reading: '*She has four children. (lit. Her children are four)'
(Elias 2017 fieldnotes)

The general classifier is Lio, Rongga *əsa*, Keo *Ɂəsa*. When counting humans, the appropriate classifier is Lio *kolo*, Rongga *mori*, Keo *ŋgaɁe*, and when counting animals it is Lio, Rongga *eko*, Keo *Ɂeko*.

In addition to these three common classifiers, there are hundreds of other classifiers which sort by size, shape, texture, function, and many other categories. The classifier Lio, Keo, Rongga *puɁu* ('trunk') is used for large, cylindrical objects such as trees, while Lio, Keo, Rongga *toko* ('bone') is used for smaller cylindrical objects like tubers and sticks. Some classifiers are very abstract: Lio has a classifier *wuŋa*, used for things which can potentially be used as weapons (machetes, bows, spears, digging sticks), and there is a dedicated classifier in Lio, Keo *papa* ('side') for things which naturally come in pairs: spouses, legs, arms, left and right sides of an object.

While the singular forms with the proclitic *sa=* display the expected Austronesian order (numeral-classifier), higher numerals in the Central Flores languages show an inverted order (classifier-numeral).

Lio

(22) *dʒata ria   sa=eko*
eagle large SG=CLASS.ANIMAL
'one large eagle'   (Ibu_Ferdy_Frogstory)

Lio

(23) *laki kolo   təlu*
chief CLASS.HUM three
'three chiefs'   (Ibu_Ferdy_History)

Lio

(24) *uwi   kadʒu toko   rua*
tuber wood CLASS.STICK two
'two cassavas'   (Positional_Elicitation)

Keo

(25) *saʔo ha=ʔesa*
house SG=CLASS.GEN
'one house'   (Baird 2002: 182)

Keo

(26) *dako ʔeko   rua ena wəwa*
dog CLASS.ANIMAL two LOC yard
'There are two dogs in the yard.'   (Baird 2002: 140)

Keo

(27) *ʔata   ŋgaʔe   dima ka dəra ndia*
person CLASS.HUM five eat day here
'Five people ate lunch here.'   (Baird 2002: 187)

Keo

(28) *nio   puʔu   dima rua*
coconut CLASS.TREE five two
'seven coconut trees'   (Baird 2002: 146)

Rongga

(29) *sapi kami   eko   lima əsa*
cow 1PL.EXC CLASS.ANIMAL five one
'We have six cows. (lit. Our cows are six.)'   (Arka et al. 2011: 36)

Rongga

(30) *mbo ito   əsa   ɻua ndau*
house small CLASS.GEN two dem
'those two small houses'   (Arka et al. 2011: xvii)

Rongga

(31) *koɗe fai   mori   ɻua*
person female CLASS.HUM two
'those two women'   (Arka et al. 2011: 7)

Rongga

(32) *dʒaʔo maŋa kamba eko   wutu*
1SG   have buffalo CLASS.ANIMAL four
'I have four buffaloes.'   (Arka et al. 2011: 95)

All three of these languages have obligatory numeral classifiers, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Numeral classifiers:	+	+	+

4.7 Verby adjectives

Mekong-Mamberamo languages tend to have 'verby adjectives': adjectives show similar morphosyntactic behaviour to verbs. This stands in contrast to languages where adjectives take either nominal marking, or their own special marking.

The Central Flores languages have verby adjectives. Due to the Central Flores languages' paucity of bound morphology, determining word classes is tricky and relies on distributional criteria. Nouns, adjectives and verbs can all serve as the head of a predicate in these languages. Baird (2002: 132) does not posit an adjective class separate from verbs for Keo, although Arka does for Rongga (2016: 118).

In the Central Flores languages, adjectives can be defined as a sub-class of verbs whose distinguishing characteristic is that they can modify an NP attributively without a relativiser. So, the Lio phrase *ae (eo) kəta ina* 'that hot water' can be expressed with or without the relativiser *eo*, since *kəta* 'to be cold' is in the sub-class of adjectival verbs:

Lio

(33) *ae*   *(eo)*   *kəta* *ina*
      water (REL) cold DEM
      'that cold water'   (Elias 2017 fieldnotes)

However, in order to express 'that running person', where the noun 'person' is attributively modified by the non-adjectival verb 'run', the relativiser *eo* is required. Omitting it leads to ungrammaticality:

Lio

(34) *ata*   *eo*   *paru* *ina*
      person REL run DEM
      'that running person'   (Elias 2017 fieldnotes)

Lio

(35) **ata*   *paru* *ina*
      person run DEM
      Failed reading: '*that running person'   (Elias 2017 fieldnotes)

Similarly Rongga *ata* is optional when used with an adjectival verb, but not with other non-adjectival verbs:

Rongga

(36) *naŋge*   *(ata)* *teʔa*
      tamarind (REL) ripe
      'ripe tamarind'   (Arka et al. 2011: 52)

Rongga

(37) *ana ata mai   ndau*
      child REL come DEM
      'the child that is coming'   (Arka 2016: 284)

However, in Keo, Baird (2002: 409) mentions that relative clauses can omit the
relativiser *ta* when the relative clause only contains one or two elements, but does
not give any examples of a relative clause consisting of a non-adjectival verb with
no relativiser.

     Many concepts encoded as adjectives in English are expressed in Central Flores
languages as verbs that can be used attributively, intransitively or transitively. Baird
(2002: 132–133) uses the example of Keo *pətu* 'hot, to heat' to illustrate this fact.
Analysing adjectives as a separate class from verbs would greatly complicate the
analysis of the numerous cases such as these. The first sentence shows an intran-
sitive predicative use, the second shows a transitive predicative use, and the third
shows an attributive use.

Keo

(38) *minu te   pətu reʔe-reʔe*
      drink DEM hot very~INTENS
      'This drink is very hot.'   (Baird 2002: 132)

Keo

(39) *rəke ha=goʔo ŋaʔo pətu ae*
      waitt SG=little 1SG   hot   water
      'Wait a moment while I heat the water.'   (Baird 2002: 133)

Keo

(40) *ŋaʔo minu kopi pətu*
      1SG   drink coffee hot
      'I'm drinking hot coffee.'   (Baird 2002: 133)

All three of these languages have verby adjectives, so I will mark all three of them
with '+' on the feature chart.

	Lio	Keo	Rongga
Verby adjectives:	+	+	+

4.8   Basic SVO constituent order

One of the features typical of Mekong-Mamberamo languages is basic SVO constituent order in transitive clauses. Although SVO constituent order is cross-linguistically extremely common, all the neighbouring languages areas (South Asian, Northeast Asian, Taiwan/Philippines, New Guinea, Australia) have other dominant constituent orders.

In the Central Flores languages, the unmarked constituent order is indeed SVO. Examples (41), (42) and (43) illustrate basic SVO sentences with full NP arguments.

Lio

   (41)   *fua   toki lako na*
            wasp bite dog DEM
            'The wasps are biting the dog.'   (Ibu_Ferdy_Frogstory)

Rongga

   (42)   *ardi poŋa ana ndau*
            Ardi hit   child DEM
            'Ardi hit the child.'   (Arka et al. 2011: xv)

Keo

   (43)   *ʔana ke   ŋae   kadʒu*
            child DEM search wood
            'That child searched for wood.'   (Baird 2002: 82)

However, SVO clauses with two NP arguments are rare. Ellipsis of core arguments is extremely common in the Central Flores languages if the referent is clear from context. Example (44) shows ellipsis of the subject of an intransitive verb in Keo:

Keo

   (44)   *bapa ena ʔemba*
            dad   LOC where
            'Where's dad?'
            ∅ *rio*
            ∅ bathe
            '(He's) bathing.'   (Baird 2002: 274)

It is equally possible to ellipse the object of a transitive verb:

Keo

(45) *na ʔemba sura ko   ŋaʔo*
      LOC where   letter GEN 1SG
      'Where is my letter?'
      ine   ŋatu ∅
      mother send ∅
      'Mother sent (it).'   (Baird 2002: 275)

Sentence (46) drawn from my Lio corpus is an example of a transitive verb with both core arguments ellipsed:

Lio

(46) *fua   paru ŋai   lako polu*
      wasp run   because dog   bark
      'The wasps run because the dog is barking.'
      ∅ iwa toki ∅
      ∅ NEG bite ∅
      '(The wasps) do not bite (the dog).'   (Ibu_Ferdy_Frogstory)

All three of these languages have basic SVO constituent order, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Basic SVO constituent order:	+	+	+

4.9   Iamitive perfects

The iamitive aspect refers to the colexicalisation of two distinct but related concepts into a single aspectual category: (1) transitions into new states which still hold at the time of reference (the perfect) and (2) events which are completed and are viewed as a finished whole (the perfective). In English, the former might be expressed with an adjective plus 'already', as in 'I'm already full', while the latter might be expressed with the past perfect, as in 'I have eaten (already).' On the other hand, in Malay, the equivalent sentences *Saya sudah kenyang* 'I'm already full' and *Saya sudah makan* 'I have (already) eaten' are formally identical.

In Lio and Rongga too, these two senses are expressed in the same way, and hence the iamitive aspect is present as an aspectual category in Lio and Rongga.

Lio

(47) *aku boʔo dowa*
1SG PERF full
'I am already full.'   (Elias 2017 fieldnotes)

Lio

(48) *aku ka dowa*
1SG eat PERF
'I have eaten.'   (Elias 2017 fieldnotes)

Rongga

(49) *somo   mbuʔe ga   siɹa hoŋa   ga*
because adult   PERF 3PL handsome PERF
'Because they are already grown-up and handsome'   (Arka 2016: 278)

Rongga

(50) *kau dadi   ga   ana ndau*
2SG give.birth PERF child DEM
'You have given birth to that child.'   (Arka 2016: 110)

However, Baird (2002) reports that Keo does distinguish between these two aspectual categories through the placement of the morpheme *neɣa, ɣa*. When placed before the predicate, it yields the 'persistent perfect' (a transition to a state that still holds at the time of speaking) but when placed after the predicate, it yields the 'perfective/completive' aspect (event viewed as a finished whole). To illustrate the difference, Baird (2002: 308) provides the pair of examples in (51) and (52):

Keo

(51) *aʔi ŋaʔo neɣa   poʔi. ŋaʔo mbana tado*
leg 1SG PER.PER break 1SG walk   cannot
'My leg is broken. I can't walk.'   (Baird 2002: 308)

Keo

(52) *aʔi ŋaʔo poʔi neɣa.   ŋaʔo ɓia   poʔi wadi*
leg 1SG break per.com 1SG not.want break again
'My leg has been broken. I don't want it broken again!'   (Baird 2002: 308)

Thus, while the two senses of the iamitive are encoded by a single morpheme in Keo *neɣa, ɣa*, there is still a formal distinction between the two senses. Therefore, I assign Keo a '+/− to indicate that it displays some of the features of iamitive perfects.

	Lio	Keo	Rongga
Iamitive perfects:	+	+/−	+

4.10   'Give' causatives

In Mekong-Mamberamo languages, causative constructions are often expressed with a morpheme identical to or derived from the verb 'to give'. This is found in eastern Malay varieties such as Papuan Malay, where causatives like Standard Indonesian *mematikan* 'to kill' are often expressed as *kasi mati* 'give die' instead.

In Lio, the most common causative serial verb construction indeed uses the verb *pati* 'to give'.

Lio

(53)   *guru   pati duke kami   ləka nia ana kəlas satu*
        teacher give kneel 1PL.EXC LOC face child class   one
        'Teacher made us kneel in front of the first grade children.'
        (Ibu_Ferdy_Scorpion_Story)

Less commonly, the verb *tau* 'to make' also serves as a causitiviser in Lio.

Lio

(54)   *ana   mo   tau   masa nia*
        child PROS make clean face
        'The child is going to wash its face.'   (Ibu_Ferdy_Bridewealth)

However, in Keo and Rongga, only the verb *tau* 'to make' is used in serial verb constructions to express causation.

Rongga

(55)   *selu   tau   mata manu   ndau*
        Selus make die   chicken DEM
        'Selus kills that chicken.'   (Arka 2016: 227)

Keo

(56)   *ʔimu tau   ɓuge ʔana ʔimu*
        3SG   make fat   child 3SG
        'She fattened her child.'   (Baird 2002: 118)

Keo and Rongga are both lacking 'give' causatives, so I will mark them with '-' on the feature chart.

	Lio	Keo	Rongga
'Give' causatives:	+	−	−

4.11 Low differentiation of adnominal attributive constructions

This feature refers to the formal similarity of three types of adnominal attributive constructions: genitival, adjectival and relative clause constructions. In English, these three types of noun phrases are distinct on the surface: genitival ('Adam's book'), adjectival ('the red book') and relative ('the book that Adam bought'). In the Mekong-Mamberamo area, languages tend to collapse these three syntactic constructions to some degree. Thus, in Minangkabau, the possessive, genitival and relative relations can all be expressed with simple juxtaposition:

Minangkabau, Western Indonesia

(57) *rumah fadʒar*
    house Fajar
    'Fajar's house'   (Gil 2015: 292)

Minangkabau, Western Indonesia

(58) *rumah ketek*
    house small
    'small house'   (Gil 2015: 292)

Minangkabau, Western Indonesia

(59) *rumah fadʒar bali*
    house Fajar buy
    'the house that Fajar bought'   (Gil 2015: 292)

As seen in Section 4.7, adjectival notions are expressed in the Central Flores languages by means of adjectival verbs, which may modify nouns with or without an intervening relativiser:

Lio

(60) *ae   (eo) kəta ina*
    water (REL) cold DEM
    'that cold water'   (Elias 2017 fieldnotes)

Rongga

(61) *naŋge (ata) teʔa*
    tamarind (REL) ripe
    'ripe tamarind'   (Arka et al. 2011: 52)

Keo

(62) *ɓaki nio   (ta) wadʒo*
    plank coconut (REL) old
    'old coconut planks'   (Baird 2002: 410)

As for relative clauses with non-adjectival verbs Lio and Rongga, the relativiser (Lio *eo*, Rongga *ata*) is not optional:

Lio

(63) *lako eo   kai gəti*
dog REL 3SG buy
'the dog that he bought'   (Elias 2017 fieldnotes)

Rongga

(64) *ana ata mai   ndau*
child REL come DEM
'that child who is coming'   (Arka 2016: 284)

On the other hand, the relativiser *ta* is optional in relative phrases in Keo, and Example (65) is equally grammatical with and without it:

Keo

(65) *puʔu kadʒu (ta)   ŋara mona nde*
trunk wood (REL) name NEG   DEM
'that tree with no name'   (Baird 2002: 410)

In Lio, Keo and Rongga, adnominal attributive possession can be expressed by simple possessum-possessor juxtaposition:

Lio

(66) *kolo lako*
head dog
'the dog's head'   (Ibu_Ferdy_Frogstory)

Keo

(67) *aʔi medʒa*
leg table
'the leg of the table'   (Baird 2002: 214)

Rongga

(68) *uma   simeon*
garden Simeon
'Simeon's garden'   (Arka 2016: 188)

In Lio (but not in Keo and Rongga) it is possible to use the relativiser *eo* to express adnominal possession:

Lio

(69) *taŋgo eo   ata   fai*
portion REL person woman
'the woman's portion'   (Ibu_Ferdy_Bridewealth)

In Rongga and Keo (but not in Lio) there is a genitive morpheme that sometimes appears between the possessum and the possessor: Keo *koʔo*, Rongga *ko*. In Keo, this particle is obligatory in some contexts, such as when the two nouns in the possessive adnominal construction are common nouns or kin terms. The genitive particle is also preferred in some contexts in Rongga, but not the same set of contexts as in Keo.

Keo

(70) *ʔana koʔo wəta*
child GEN sister
'sister's child'   (Baird 2002: 214)

Rongga

(71) *lako ko   domi*
dog GEN Domi
'Domi's dog'   (Arka 2016: 187)

In summary, all three languages can use simple juxtaposition to express adnominal possession. In addition, Lio can use the relativiser *eo*, and Keo and Rongga can use the genitive marker Keo *koʔo*, Rongga *ko* to express adnominal possession.

Overall, the picture is mixed. Lio has the lowest level of differentiation of adnominal constructions, because the relativiser *eo* can be used to form possessive, genitive and relative constructions. Keo also has a low level of differentiation of adnominal constructions, since juxtaposition is used to express possessive, genitive and relative constructions. However, in some classes of genitive constructions, the genitive particle Keo *koʔo* is obligatory, so some types of possession cannot be expressed through juxtaposition. Therefore, I assign Keo a '+/−' in the table of features. Rongga has a higher level of differentiation of adnominal constructions, since it has not only the genitive particle *ko* in many genitive constructions, it also cannot generally drop the relativiser *ata* in relative clauses. Therefore I assign Rongga a '−' in the feature chart.

	Lio	Keo	Rongga
Low differentiation of adnominal attributive constructions	+	+/−	−

4.12 Weakly developed grammatical voice

A language can be said to have a weakly developed grammatical voice system if there is no overt, morphologically marked mechanism for voice alternations such as the passive. This is generally true of the Mekong-Mamberamo languages. Voice alternations encoded only as constituent order changes still qualify as 'weakly developed grammatical voice' by the definition of Gil (2015).

The Central Flores languages do not have dedicated morphology for voice alternations and therefore show weakly developed grammatical voice. A system of voice alternation through constituent order changes is grammaticalised to the greatest degree in Rongga, where the passive is systematically expressed by promoting the Patient to subject position, and reintroducing the Agent in a prepositional phrase with *ne* 'with, by'.

Rongga
(72) *ardi poŋa ana ndau*
    Ardi hit   child DEM
    'Ardi hit that child. (Not: *That child hit Ardi.)'   (Arka 2016: 217)

Rongga
(73) *ana ndau poŋa ne ardi*
    child DEM hit   by Ardi
    'That child was hit by Ardi.'   (Arka 2016: 217)

Rongga also shows the following restriction on relative clauses: the object of an active transitive clause cannot be relativised, but must be reformulated as the subject of a passive clause first.

Rongga
(74) *ana ata ardi poŋa ndau ɓako   dʒaʔo*
    child REL Ardi hit   DEM nephew 1SG
    Failed reading: '*The child that Ardi hit is my nephew.'   (Arka 2016: 220)

Rongga
(75) *ana ata poŋa ne ardi ndau ɓako   dʒaʔo*
    child REL hit   by Ardi DEM nephew 1SG
    Acceptable: 'The child that was hit by Ardi is my nephew.'   (Arka 2016: 220)

Lio and Keo are lacking the grammaticalised voice alternation system described for Rongga. In Keo, the object in an active SVO sentence can be fronted to topicalise it, but the Patient remains the object, and the Agent remains the subject. The Agent is not demoted to become an oblique argument, and therefore this is not a true passive

construction. Lio also lacks a true passive, but uses the same object fronting strategy as Keo. The first example shows a regular active SVO sentence where 'Nus' is the subject/Agent and 'Arno' is the object/Patient:

Keo

(76) *nus boba arno*
      Nus hit   Arno
      S   V   O
      'Nus hit Arno.'   (Baird 2002: 78 – slightly modified)

When the object is fronted, this yields the OSV clause in (77), but 'Arno' remains the object/Patient despite it being in initial position. The subject/Agent 'Nus' is still required and has not been demoted.

Keo

(77) *arno   nus boba*
      Arno Nus hit
      O   S   V
      'Nus hit Arno. (Not: *Arno was hit by Nus.)'   (Baird 2002: 78)

The final example shows that it is not a true passive construction because the sentence (78) must be interpreted as an active SV(O) clause with the object elided, rather than a passive SV clause with the Patient as subject.

Keo

(78) *arno   boba*
      Arno hit
      S   V
      'Arno hit (someone). (Not: *Arno was hit.)'   (Baird 2002: 79)

Unlike in Rongga, there are no restrictions on relativising the object of an active clause in Lio and Keo:

Lio

(79) *dəlu   aku manusia eo   fua   toki*
      friend 1sg person   REL wasp bite
      'My friend is the person who the wasps bit.'   (Ibu_Ferdy_Frogstory)

Keo

(80) *ʔata   ta   ʔimu boba ke   palu*
      person REL 3sg   hit   DEM run
      'The person that he hit ran.'   (Baird 2002: 72)

All three of these languages have weakly developed grammatical voice, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Weakly developed grammatical voice:	+	+	+

4.13   Isolating word structure

Mekong-Mamberamo languages typically have a low number of morphemes per word, with grammatical morphemes expressed as independent words.

This is very much true of the Central Flores languages. They are notable precisely for their almost total absence of bound affixes, in contrast to the modest amount of morphology present in the languages of West and East Flores (see Sections 3.1, 3.2).

There is one proclitic (Lio and Rongga *sa=*, Keo *ha=*) which marks the singular number in numeral phrases and is reconstructible as Proto-Central Flores *sa= (Elias 2018). This is transparently related to the form of the numeral for one, PCF *əsa.

All three of these languages have isolating word structure, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Isolating word structure:	+	+	+

4.14   Short words

Mekong-Mamberamo languages typically have short words, which is unsurprising given their isolating morphology. In many languages, there are constraints on the maximum size of a word.

The Central Flores languages have a very restricted range of possible word shapes. They allow only open syllables and do not allow any consonant clusters, with maximally disyllabic word of form CVCV. Native, monomorphemic words do exceed two syllables (see Table 9 for an exhaustive list of possible word shapes in Central Flores). Words longer than two syllables are either loans, or formed by compounding or fossilisation of the singular proclitic (Lio, Rongga *sa=*, Keo *ha=*).

Table 9. Exhaustive list of possible word shapes in Central Flores languages

Shape	Lio	Keo	Rongga
V	*e* 'think'	*e* 'think'	*e* 'exclamation'
CV	*ka* 'eat'	*fu* 'hair'	*ɓa* 'plate'
VV	*ae* 'water'	*oa* 'request'	*ua* 'rattan'
VCV	*eko* 'tail'	*uwa* 'skin'	*aŋe* 'maybe'
CVV	*ria* 'large'	*loa* 'burn'	*lea* 'ginger'
CVCV	*pati* 'give'	*rəto* 'dip'	*təlo* 'egg'

All three of these languages have short words, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Short words:	+	+	+

4.15   Low grammatical-morpheme density

Mekong-Mamberamo languages often display a low grammatical-morpheme density. Utterances are often composed mainly of lexical items, with few other morphemes required to bind them into a grammatical utterance. This is logically distinct from isolating morphology, but often co-occurs with it.

The Central Flores languages have low grammatical morpheme density. Since very few semantic distinctions are obligatorily expressed on either verbs or nouns, rather long sentences consisting only of content words are not uncommon in the Central Flores languages.

Lio

(81) *ata   fai   koɗo tei   ana   wawi*
       person female look find child pig
       'The woman looks and sees a piglet.'   (Elias 2017 fieldnotes)

This sentence requires the use of five grammatical morphemes in English: each noun phrase must receive either a definite article or an indefinite article, each 3sg verb must take a final /-s/, and conjunction must be expressed using 'and'. None of these are required in Lio, which expresses the notion of 'looking and seeing' as a bare serial verb construction and does not obligatorily express definiteness on nouns or tense on verbs. The same holds true in the other Central Flores languages.

All three of these languages have low grammatical-morpheme density, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Low grammatical-morpheme density:	+	+	+

4.16 Optional thematic-role flagging

A feature of the Mekong-Mamberamo language area is optional thematic-role flagging, which means that the arguments of a verb are not necessarily overtly marked to indicate their relationship to the verb. Cross-linguistically, oblique arguments are more likely to require an overt marker (such as a preposition) than core arguments.

Gil (2015) cites examples from languages in the Mekong-Mamberamo area where oblique arguments do not need to be introduced by an overt marker ('bare oblique' constructions), such as in (82):

Sundanese, Western Indonesia

  (82) *dʒələma   dahar   taŋkal*
          person   eat   tree
          'The man is eating by the tree.'   (Gil 2015: 317)

Meyah, Western New Guinea

  (83) *isok   et   mega*
          man eat tree
          'The man is eating by the tree.'   (Gil 2015: 317)

While Central Flores languages do not overtly mark the core arguments of the verb, marking of oblique arguments is obligatory. Thus, the Lio sentence in (84) is ungrammatical without the use of *ləka* 'in, at' to introduce the oblique argument:

Lio

  (84) *ani   məra   ləka   puʔu   kadʒu*
          bee live   LOC trunk wood
          'The bees live in the tree.'   (Ibu_Ferdy_Frogstory)

Lio

  (85) **ani   məra   puʔu   kadʒu*
          bee live   trunk wood
          Failed reading: '*The bees live in the tree.'   (Ibu_Ferdy_Frogstory)

Oblique arguments must be overtly marked in Keo and Rongga as well.

Keo

(86) *ʔimu kere dau maʔu*
3SG   wait down beach
'He waited down at the beach.'   (Baird 2002: 89)

Rongga

(87) *kaɻi ŋgoe one raɗi*
3SG   fall   LOC stair
'He fell on the stairs.'   (Arka 2016: 144)

Since the Central Flores languages lack the bare oblique construction Gil (2015) cites in other Mekong-Mamberamo languages, I give all three languages a '–' for this feature.

	Lio	Keo	Rongga
Optional thematic-role flagging	–	–	

4.17   Optional TAM marking

In many Mekong-Mamberamo languages, the expression of tense, aspect and mood is optional. This is true of the Central Flores languages as well. A clause with no overt TAM marking is not restricted in its range of possible interpretations. Hence, the Lio sentence (88) with no TAM marking could receive a range of possible temporal interpretations:

Lio

(88) *fua toki lako na*
wasp bite dog DEM
'The wasps [are biting/bit] the dog.'   (Ibu_Ferdy_Frogstory)

TAM marking is also not obligatory in Keo or Rongga:

Keo

(89) *kaʔe   ka ʔuwi dʒawa*
older.sibling eat tuber Java
'Big brother [is eating/ate] sweet potato.'   (Baird 2002: 78)

Rongga

(90) *pondo ndau məɗo*
pot   DEM fall
'That pot [is falling/fell].'   (Arka 2016: 157)

All three of these languages have optional TAM marking, so I will mark all three of them with '+' on the feature chart.

	Lio	Keo	Rongga
Optional TAM marking:	+	+	+

4.18   Summary: Mekong-Mamberamo features in Central Flores languages

I have examined the 17 features identified by Gil (2015) as typical of the Mekong-Mamberamo linguistic area and assessed their presence or absence in the languages of Central Flores. Table 10 summarises the findings.

Table 10. Presence or absence of Mekong-Mamberamo features in Central Flores

	Lio	Keo	Rongga
The passing gesture	+	?	?
Repeated dental clicks expressing amazement	+	?	?
Conventionalised greeting with 'where?'	+	+	?
'eye day' to 'sun' lexicalisation	+	+	+
d/t place-of-articulation mismatch	+	?	?
Numeral classifiers	+	+	+
Verby adjectives	+	+	+
Basic SVO constituent order	+	+	+
Iamitive perfects	+	+/−	+
'Give' causatives	+	−	−
Low differentiation of adnominal attributive constructions	+	+/−	−
Weakly developed grammatical voice	+	+	+
Isolating word structure	+	+	+
Short words	+	+	+
Low grammatical-morpheme density	+	+	+
Optional thematic-role flagging	−	−	−
Optional TAM marking	+	+	+

The only feature which is missing from all three languages is 'optional thematic-role flagging'. In all three languages in my sample, oblique arguments are marked obligatorily, not optionally.

5.   Additional evidence from the Central Flores numeral system

In this section, I describe the multiple numeral systems found in Lio, Keo and Rongga, which shed light on certain aspects of the pre-Austronesian languages of Central Flores.

One of the defining innovations of the Central Flores languages is the restructuring of the Austronesian decimal numeral system into a mixed-base quinary-decimal system. Table 11 shows the decimal PMP numerals on the left, followed by languages of East Flores (Sika) and West Flores (Manggarai) which retained that system more or less intact. Also included are the Palu'e numerals, to show that these are of the decimal Austronesian type and not the quinary-decimal Central Flores type. These are contrasted with the reconstructed Proto-Central Flores numerals (Elias 2018) and their reflexes in Lio, Keo and Rongga in Table 12.

The Proto-Central Flores numerals (plus the obligatory general classifier *əsa) are reconstructed in Elias (2018) are as follows: *sa=[əsa] 'one', *[əsa] dua 'two',

Table 11. Inherited decimal numerals in the languages of East Flores, West Flores and Palu'e

	PMP	Sika	Manggarai	Palu'e
1	*esa	ha	tʃa	a
2	*duha	rua	sua	rua
3	*telu	təlu	təlu	təlu
4	*epat	hutu	pat	ɓa
5	*lima	lima	lima	lima
6	*enem	əna	ənəm	ʔəne
7	*pitu	pitu	pitu	ɓitu
8	*walu	walu	alo	walu
9	*siwa	hiwa	tʃiok	iwa
10	*sa-ŋa-puluq	pulu ha	tʃa mpulu	a pulu
11	–	pulu wot ha	tʃa mpulu tʃa	–
12	–	pulu wot rua	tʃa mpulu sua	–
13	–	pulu wot təlu	tʃa mpulu təlu	–
14	–	pulu wot hutu	tʃa mpulu pat	–
15	–	pulu wot lima	tʃa mpulu lima	–
16	–	pulu wot əna	tʃa mpulu ənəm	–
17	–	pulu wot pitu	tʃa mpulu pitu	–
18	–	pulu wot walu	tʃa mpulu alo	–
19	–	pulu wot hiwa	tʃa mpulu tʃiok	–
20	*duha-ŋa-puluq	pulu rua	sua mpulu	rua pulu
100	*sa-ŋa-Ratus	ŋasu ha	tʃa ratus	a tʃatu
1000	*sa-ŋa-Ribu	riwu ha	sa=səbu	a riwu

*[əsa] təlu 'three', *[əsa] wutu 'four', *[əsa] lima 'five', *[əsa] lima əsa 'six', *[əsa] lima dua 'seven', *[əsa] dua mbutu 'eight', *[əsa] təra əsa 'nine'. The higher bases are *mbulu 'ten', *ŋasu '100', *riwu '1000'.

The numerals from 6 to 9 are derived from the lower numerals by a number of different strategies. PCF *lima əsa 'six' means 'one five' and PCF *lima dua 'seven' means 'five two', using an additive strategy. PCF *dua mbutu 'eight' means 'two four' using a multiplicative strategy (note the prenasalisation on the second element compared with PCF *wutu 'four'; this may be the remnant of the PMP *-ŋa- morpheme which appears in PMP *sa-ŋa-puluq > PCF *sa=mbulu 'ten'). Finally, PCF *təra əsa 'nine' seems to be composed of an initial morpheme meaning 'to take away, to remove' followed by 'one', so it means something like 'take away one (from ten)' using a subtractive strategy.

Note that the general classifier PCF *əsa is homophonous with the morpheme used to represent the number 1 in the composed numerals 6 and 9 (and may share an etymology in PMP *əsa 'one', although they are now clearly separate in Central Flores). In order to help the reader, the instances where PCF *əsa is used as a classifier are placed between square brackets. If the Central Flores speaker were counting something that required a different classifier, such as animals, all instances of PCF *əsa 'general classifier' would be replaced with PCF *eko 'animal classifier' (Lio, Rongga *eko*, Keo *ʔeko*).

Table 12. Quinary-decimal numerals in the modern languages of Central Flores

1	Lio	Keo	Rongga
1	sa=[əsa]	ha=[ʔəsa]	sa=[əsa]
2	[əsa] rua	[ʔəsa] rua	[əsa] ɹua
3	[əsa] təlu	[ʔəsa] tədu	[əsa] təlu
4	[əsa] sutu	[ʔəsa] wutu	[əsa] wutu
5	[əsa] lima	[ʔəsa] dima	[əsa] lima
6	[əsa] lima əsa	[ʔəsa] dima ʔəsa	[əsa] lima əsa
7	[əsa] lima rua	[ʔəsa] dima rua	[əsa] lima ɹua
8	[əsa] rua mbutu	[ʔəsa] rua mbutu	[əsa] ɹua mbutu
9	[əsa] təra əsa	[ʔəsa] təra ʔəsa	[əsa] təra əsa
10	sa=mbulu	ha=mbudu	sa=mbulu
11	sa=mbulu sa=[əsa]	ha=mbudu ha=[ʔəsa]	sa=mbulu sa=[əsa]
12	sa=mbulu [əsa] rua	ha=mbudu [ʔəsa] rua	sa=mbulu [əsa] ɹua
13	sa=mbulu [əsa] təlu	ha=mbudu [ʔəsa] tədu	sa=mbulu [əsa] təlu
14	sa=mbulu [əsa] sutu	ha=mbudu [ʔəsa] wutu	sa=mbulu [əsa] wutu
15	sa=mbulu [əsa] lima	ha=mbudu [ʔəsa] dima	sa=mbulu [əsa] lima
16	sa=mbulu [əsa] lima əsa	ha=mbudu [ʔəsa] dima ʔəsa	sa=mbulu [əsa] lima əsa
17	sa=mbulu [əsa] lima rua	ha=mbudu [ʔəsa] dima rua	sa=mbulu [əsa] lima ɹua
18	sa=mbulu [əsa] rua mbutu	ha=mbudu [ʔəsa] rua mbutu	sa=mbulu [əsa] ɹua mbutu
19	sa=mbulu [əsa] təra əsa	ha=mbudu [ʔəsa] təra ʔəsa	sa=mbulu [əsa] təra əsa

Table 12. (*continued*)

1	Lio	Keo	Rongga
20	*mbulu rua*	*mbudu rua*	*mbulu ɹua*
100	*sa=ŋasu*	*ha=ŋasu*	*sa=ŋasu*
1000	*sa=riwu*	*ha=liwu*	*sa=riwu*

In addition to the quinary-decimal system, there is a quaternary (base-4) numeral system which is present in all Central Flores languages and is therefore reconstructible to Proto-Central Flores (Table 13). This system has a more restricted application, typically being used (by Lio speakers at least) when dealing with small objects which can be stacked into pyramids of 4 such as coconuts, areca nuts, or limes. Interestingly, the highest repeating base in this system (40) is reconstructed as PCF *ulu* 'head; 40'. Non-Austronesian (Papuan) languages of the region are well-known for using body part words as numerical bases (Schapper & Klamer 2014), so this may reflect semantic influence from a non-Austronesian language.

Finally, Arka (2016: 127–128) describes an intriguing additional decimal numeral system in Rongga, which brings the number of distinct numeral systems in that language to three, with three separate bases (quaternary, quinary, decimal). This system is not productive in that it does not go above 20 and the formation of numbers above 10 is not transparent (see Table 14). A notable feature of this counting system is that various numbers contain consonant clusters which violate the phonotactic rules of Rongga: particularly *ŋgwo* 'nine' and *mopla* 'twenty'. Furthermore, none of the numerals below ten except *dua* 'two' and possibly *da* 'one' have a plausible Austronesian etymology. The numerals 3–10 do not resemble the numerals in any Austronesian language, nor in any of the nearby non-Austronesian Timor-Alor-Pantar languages (Schapper & Klamer 2014). The formation of the

Table 13. Quaternary numerals in the languages of Central Flores

	PCF	Lio	Keo	Rongga
1	**sa=[əsa]*	*sa=[əsa]*	*ha=[ʔəsa]*	*sa=[əsa]*
2	**[əsa] dua*	*[əsa] rua*	*[ʔəsa] rua*	*[əsa] ɹua*
3	**[əsa] təlu*	*[əsa] təlu*	*[ʔəsa] tədu*	*[əsa] təlu*
4	**sa=liwu*	*sa=liwu*	*ha=diwu*	*sa=liwu*
5	**sa=liwu sa=[əsa]*	*sa=liwu sa=[əsa]*	*ha=diwu ha=[ʔəsa]*	*sa=liwu sa=[əsa]*
6	**sa=liwu [əsa] dua*	*sa=liwu [əsa] rua*	*ha=diwu [ʔəsa] rua*	*sa=liwu [əsa] ɹua*
7	**sa=liwu [əsa] təlu*	*sa=liwu [əsa] təlu*	*ha=diwu [ʔəsa] tədu*	*sa=liwu [əsa] təlu*
8	**liwu dua*	*liwu rua*	*diwu rua*	*liwu ɹua*
9	**liwu dua sa=[əsa]*	*liwu rua sa=[əsa]*	*diwu rua ha=[ʔəsa]*	*liwu ɹua sa=[əsa]*
10	**liwu dua [əsa] dua*	*liwu rua [əsa] rua*	*diwu rua [ʔəsa] rua*	*liwu ɹua [əsa] -ɹua*
…	…	…	…	…
40	**sa=ulu*	*sa=ulu*	*ha=ʔudu*	*sa=ulu*

Table 14. Non-productive numerals in Rongga

	Rongga
1	ɗa
2	ɗua
3	dəmu
4	dəke
5	ali
6	woe
7	sipi
8	sapa
9	ŋgwo
10	ŋguru
11	əsaŋgəsa
12	ɾuaɗua
13	təluŋgətu
14	wutuŋgutu
15	limakima
16	aŋgunae
17	nəŋgonae
18	soroila
19	watopəsa
20	mopla

numerals 11–15 is intriguing: here we see the inherited Central Flores etyma PCF *əsa 'one', *dua 'two', *təlu 'three', *wutu 'four', *lima 'five' make an appearance. It seems that the strategy for forming numerals 11–15 is by reduplicating the numerals for 1–5 with an unpredictably altered initial consonant. Thus, while the numerals *dəmu* 'three', *dəke* 'four', *ali* 'five' do not reflect PCF numerals, the numerals *təluŋgətu* 'thirteen', *wutuŋgutu* 'fourteen' and *limakima* 'fifteen' do reflect the PCF numerals. From 16 to 20, almost nothing can be ascertained about the etymology of the numerals except that *watopəsa* 'nineteen' may perhaps contain a morpheme reflecting PCF *əsa* 'one' and therefore be formed by subtraction (20–1) (Arka 2016: 127).

Throughout this number system, there is a pervasive tendency for neighbouring numerals to alliterate with each other. This is a commonly noted development cross-linguistically. In the Rongga numeral system under discussion, there seems to be frequent use of such alliteration: note the pairs *sipi* 'seven', *sapa* 'eight', or the pair *ŋgwo* 'nine', *ŋguru* 'ten'. The numeral *təluŋgətu* 'thirteen' appears to have shifted from the expected **təluŋgəlu* under the influence of the following *wutuŋgutu* 'fourteen', so that the two numerals share the same ending.

This non-productive Rongga numeral system shares a number of suggestive parallels with the 'Yan Tan Tethera' sheep-counting systems of Northern England

(Ingram 1977), which are a remarkable instance of lexical items from a substrate language enduring for many centuries.

The sheep-counting systems of Northern England were used by shepherds into the early 20th century. They are limited to numbers below 20 (like Rongga) and are derived from the extinct Brythonic Celtic languages of Northern England such as Cumbric which form a linguistic substrate in this part of England. The numerals overall resemble those found in modern Brythonic languages such as Welsh, but adjacent numerals are often altered to alliterate better (like in Rongga). In the regions of Northern England where these sheep-counting systems are used, the Celtic languages themselves have been extinct for centuries: for instance, Cumbric was extinct by the 12th century. However, the sheep-counting systems survived into the 20th century (in modified form) as a last lexical vestige of the earlier Celtic languages. Several Northern English variants of the sheep-counting system are presented in Table 15, along with modern Welsh for comparison (Ingram 1977).

In my view, the non-productive Rongga counting system laid out in Table 14 is analogous to the sheep-counting systems of Northern England: the last lexical remnant of a group of long-extinct substrate languages. Based on the lack of similarities between the forms of the numerals and any other languages of the area, I suggest that the substrate language has no living descendants or relatives in the area.

Table 15. Borrowed Celtic sheep-counting systems used in Northern England

	Welsh	Bowland	Coniston	Tong
1	*un*	*yain*	*yan*	*yan*
2	*dau*	*tain*	*taen*	*tan*
3	*tri*	*eddera*	*tedderte*	*tether*
4	*pedwar*	*peddera*	*medderte*	*mether*
5	*pump*	*pit*	*pimp*	*pick*
6	*chwech*	*tayter*	*haata*	*sesan*
7	*saith*	*layter*	*slaata*	*asel*
8	*wyth*	*overa*	*lowra*	*catel*
9	*naw*	*covera*	*dowra*	*oiner*
10	*deg*	*dix*	*dick*	*dick*
11	*un ar ddeg*	*yain-a-dix*	*yan-a-dick*	*yanadick*
12	*deuddeg*	*tain-a-dix*	*taen-a-dick*	*tanadick*
13	*tri ar ddeg*	*eddera-a-dix*	*tedder-a-dick*	*tetheradick*
14	*pedwar ar ddeg*	*peddera-a-dix*	*medder-a-dick*	*metheradick*
15	*pymtheg*	*bumfit*	*mimph*	*bumfit*
16	*un ar bymtheg*	*yain-a-bumfit*	*yan-a-mimph*	*yanabum*
17	*dau ar bymtheg*	*tain-a-bumfit*	*taen-a-mimph*	*tanabum*
18	*deunaw*	*eddera-a-bumfit*	*tedder-a-mimph*	*tetherabum*
19	*pedwar ar bymtheg*	*peddera-a-bumfit*	*medder-a-mimph*	*metherabum*
20	*ugain*	*jiggit*	*gigget*	*jigget*

6.   Historical proposal for Central Flores languages

The historical proposal that follows is informed by Thomason and Kaufman's (1988) framework of contact scenarios and their linguistic outcomes. They draw a distinction between two basic types of contact-induced change: 'borrowing' and 'substratum interference'.

Borrowing refers to "the incorporation of foreign features into a group's native language by speakers of that language: the native language is maintained but is changed by the addition of the incorporated features" (Thomas & Kaufman 1988: 37). The first changes are lexical adoptions, and with more intense contact phonological, morphological and syntactic elements may be borrowed as well. Lexical borrowing may take place without widespread societal bilingualism, while grammatical borrowing usually takes place in a situation of extensive bilingualism. An example Thomason and Kaufman (1988) give is the influence of Sanskrit on the Mainland Southeast Asian languages: while many Sanskrit words have been adopted into Thai, Khmer and other SE Asian languages, their grammatical structure has not been changed much and the bulk of Thai and Khmer speakers were never proficient in Sanskrit. Another example is the effect of Indian languages on the English of the British colonists. English adopted local words for concepts which did not exist before, but the grammar was hardly affected.

Substratum interference occurs when "a group of speakers shifting to a target language fails to learn the target language correctly" (Thomas & Kaufman 1988: 39). Unlike in borrowing, in substratum interference the group of speakers does not maintain their own native language, but adopts the native language of another group with which they are in close contact. The linguistic traces of substratum interference are very different from borrowing: "unlike borrowing, interference through imperfect learning does *not* begin with vocabulary: it begins instead with sounds and syntax, and sometimes includes morphology as well before words from the shifting group's original language appear in the T[arget] L[anguage]" (Thomason & Kaufman 1988: 39, emphasis in original). The grammatical effects of substratum interference are particularly strong when the shifting group is numerically greater than the target language speakers. Thus, the linguistic effects of substratum interference are almost opposite to those of borrowing: structural interference first, then vocabulary. Often, the target language is of higher prestige than the shifting language, and hence the vocabulary of the target language will be preferred over the speakers' native vocabulary. An example of substratum interference given by Thomason and Kaufman (1988) is the case of Dravidian influence in Sanskrit. The Dravidian languages are held to be the source of many of the distinctive typological features of Sanskrit (retroflex consonants, among others) but there is not a great amount of clearly Dravidian vocabulary in Sanskrit. This is because

Dravidian speakers shifted to Sanskrit, rather than Sanskrit speakers borrowing from Dravidian.

Another example of substratum interference is the effect of Uralic on the Slavic languages. Uralic features in Slavic are held to include phonemic palatalisation, large case systems and the use of the partitive genitive, among others. However, some have argued that since there is very little evidence of Uralic vocabulary in Slavic, this is evidence against close contact between them. In fact, this does not argue against contact: it merely argues against a situation where Slavic borrowed heavily from Uralic. It is completely consistent with a substratum interference scenario where Uralic speakers shifted to Slavic, resulting in grammatical but not lexical interference.

Another difference between borrowing and substratum interference is the amount of time required for far-reaching structural modification. For extensive grammatical borrowing to take place, usually a long period of time is required. However, substratum interference may take place in a single generation: "in fact, substratum features are more likely to enter a T[arget] L[anguage] rapidly than slowly: if the shift takes place over long centuries, then the shifting population is likely to be truly bilingual in the T[arget] L[anguage]" (Thomason & Kaufman 1988: 41). Hence, the strongest effects of substratum interference will be seen in cases where the transition happened most abruptly.

In the case of Central Flores, a substratum interference scenario is clearly preferred. The Central Flores languages show an almost total restructuring of their typological profile, but are lexically very conservative. Comparing a 100-item Swadesh vocabulary list of PMP with Lio, Keo and Rongga yielded the following result: 69% retentions from PMP in Lio, 64% retentions in Keo and 63% retentions in Rongga. A high level of lexical conservatism, combined with heavy grammatical interference, is diagnostic of a substratum interference scenario brought about by rather abrupt language shift according to Thomas and Kaufman (1988).

With this in mind, I propose the following historical scenario to account for the typological profile of the Central Flores languages: around 4000 years before the present, the Austronesians began their push into the Indonesian archipelago from the Philippines. They brought with them richly inflected VSO Austronesian languages similar to the modern languages of Taiwan and the Philippines. They began to encroach upon the Mekong-Mamberamo language area, composed of a large number of diverse lineages which had converged structurally over a long period of time due to shared historical links.

When the Austronesians began to settle Flores around 2500–1500 BCE, they encountered speakers of Mekong-Mamberamo type languages who were numerically dominant. The Austronesian settlers enjoyed a relatively high level of prestige over the pre-Austronesian inhabitants, who began to shift from their language to the

Austronesian language of the settlers. These Mekong-Mamberamo speakers shifted to Austronesian in a relatively short period of time, leaving a heavy grammatical influence but very little lexical influence. The Austronesian settlers then integrated into the local population and adopted the restructured, Mekong-Mamberamo-like Austronesian language spoken by the majority of the population: Proto-Central Flores (or an immediate ancestor thereof). This ancestral community then differentiated *in situ* into the modern Central Flores languages with no further splits, forming an archetypal linkage (see Elias 2018).

In the scenario just described, the fact that Central Flores languages retain very little lexicon but much of the grammatical structure of the pre-Austronesian languages is explained. However, I do propose that lexical influence from the pre-Austronesian languages of Flores can be seen in one domain: the non-productive Rongga numeral system (see Section 5). This counting system is full of inexplicable oddities: the lower numerals do not resemble any other known languages of the area, the strategy for forming higher numerals is unusual, and several of the numerals contain consonant clusters that violate Rongga phonotactics. All of these point to a now-extinct pre-Austronesian source language whose relatives have all vanished.

Under my historical scenario, I must also account for the fact that the languages of West Flores and East Flores conform less closely to the Mekong-Mamberamo typological profile. It is possible that the pre-Austronesian languages of Flores were themselves diverse, and only the language (or languages) of Central Flores had Mekong-Mamberamo typology to begin with. The Mekong-Mamberamo hypothesis does not claim that all languages across the entire area show all of the features, merely that many of them do. Another very likely possibility is that the relevant factors at the time of contact were different in Central Flores and the rest of Flores, such that there was less substratum interference in East and West Flores. In East and West Flores, it is possible that the pre-Austronesian population was less dense, and the ratio of Austronesian settlers to pre-Austronesian inhabitants was higher. This would be expected to lead to less structural interference. Another potential factor is the degree of bilingualism; perhaps the pre-Austronesian and Austronesian populations did not integrate with each other as much in East and West Flores, and the pre-Austronesians vanished along with their languages rather than shifting to the language of the Austronesian settlers. Another possibility is that the transition took place over longer time periods in East and West Flores, leading to less structural interference in the target language due to more complete bilingualism.

There is another logical possibility which should be mentioned here: it is possible that the level of substratum interference was comparable across all of Flores, and that all Flores languages once resembled the Central Flores languages. Then, morphological complexity was re-innovated in the languages of East and West

Flores. In this case, we would expect that the morphology seen in East and West Flores would be innovative morphology, not retentions from earlier stages of Austronesian. Overall, it is a mixed picture and the evidence is somewhat inconclusive: much of the morphology, especially in East Flores, seems to be conservative, but there is clearly innovative plural marking in Lamaholot. Similarly, the conjugational systems of Sika and Lamaholot appear at first glance to be conservative, because they are cognate with many of the widespread Central Malayo-Polynesian verbal conjugation systems in the area. However, this may be a hasty judgment, because the verbal conjugation systems are themselves transparently derived from the independent pronouns, and therefore may plausibly have grammaticalised separately in different areas.

Note that in McWhorter's account, the settlers from Sulawesi would have arrived speaking something similar to Tukang Besi, which would have become simplified as the (already Austronesian-speaking) inhabitants of Central Flores shifted to the newcomers language. On the reasonable assumption that the settlers from Sulawesi enjoyed high prestige and did not borrow massive amounts of vocabulary from the shifting speakers, we would expect the resulting Central Flores language to resemble Tukang Besi lexically, but they do not. On the other hand, if the settlers from Sulawesi did not have high prestige, that raises the question of why the pre-existing population of Central Flores would have shifted to their language at all. The distinct lack of lexical resemblances between Sulawesi and Flores is acknowledged by McWhorter himself: "lexical and grammatical data in support of this scenario are lacking" (McWhorter 2019: 195). This lack of lexical evidence in support of his theory is a serious problem and must be accounted for.

7. Conclusion

The title of this paper poses a question: are the Central Flores languages really typologically unusual? In this paper, I have shown that the Central Flores languages display many of the typological features common in the Mekong-Mamberamo linguistic area proposed by Gil (2015). I argue that the typological profile of the Central Flores languages is best explained by structural interference from a now-extinct substrate language with Mekong-Mamberamo typology. This interference occurred at the time of the original Austronesian settlement of Flores around 2500–1500 BCE.

In contrast, McWhorter (2019) claims that the isolating typology of the Central Flores languages is due to the more general process of simplification that occurs under conditions of language shift, regardless of the typology of the languages involved. He proposes that a group of settlers arrived in Central Flores from Sulawesi

in the relatively recent past (once it was already inhabited by Austronesian speakers) and the Central Flores population shifted to their language. This argument suffers from a number of flaws. The most glaring among them are that: (1) it does not account for aspects of the Central Flores languages which are not simple, such as their numeral classifier systems, (2) it is undermined by the lack of lexical evidence linking Sulawesi to Flores, and (3) it is less economical to posit two separate contact events (one between Austronesian and non-Austronesian, one between Central Flores and Sulawesi) than a single contact event.

Although there are no non-Austronesian languages spoken on Flores any more, the non-Austronesian Timor-Alor-Pantar languages are also spoken in the Lesser Sunda Islands. An obvious avenue for further research is whether or not the substrate language of Central Flores was related to the modern Timor-Alor-Pantar languages. This question needs to be investigated further, but the Timor-Alor-Pantar languages are significantly more morphologically complex than the Central Flores languages and are tentatively assigned to the Trans-New Guinea language family based on pronominal evidence (Ross 2005). If the Trans-New Guinea status of the Timor-Alor-Pantar languages is confirmed, then it becomes more likely that the Timor-Alor-Pantar languages are a more recent migration from the New Guinea mainland to the Lesser Sundas, rather than a relict group of the non-Austronesian speakers who inhabited the area before Austronesian (see Schapper 2017: 15–16 for a discussion of interactions between Proto-Timor-Alor-Pantar and Austronesian).

In a discipline such as historical linguistics, we cannot hope for experimental replication of results to validate our hypotheses. Instead, we can bolster our historical hypotheses by developing them with one set of data in mind, and showing that they accurately capture the facts in a new set of previously unseen data. In the case of Gil's Mekong-Mamberamo hypothesis, this comes as close to the gold standard as possible: the Mekong-Mamberamo typological features describe the typology of the Central Flores languages very closely, even though the proposal was not developed with the Central Flores languages in mind. Thus, this paper serves not only as a contribution to the prehistory of Central Flores, but also as a practical validation of Gil's Mekong-Mamberamo hypothesis.

Acknowledgements

References

Aoki, Eriko & Nakagawa, Satoshi. 1993. *Endenese-English Dictionary*. Ms.

Arka, I Wayan. 2016. *Bahasa Rongga: Deskripsi, Tipologi, dan Teori* (The Rongga Language: Description, Typology, and Theory). Jakarta: Atma Jaya University Press.

Arka, I Wayan, Seda, Franciscus, Gelang, Antonius, Nani, Yonahes & Ture, Ivan. 2011. *A Rongga-English Dictionary with English-Rongga Finderlist*. Jakarta: Atma Jaya University Press.

Arndt, Paul. 1933a. *Lionesisch-Deutsches Wörterbuch*. Ende: Arnoldus-Druckerei.

Arndt, Paul. 1933b. *Grammatik der Ngada-Sprache*. Bandung: A.C. Nix & Co.

Arndt, Paul. 1937. *Grammatik der Solor-Sprache*. Ende: Arnoldus-Druckerei.

Arndt, Paul. 1961. *Wörterbuch der Ngadhasprache*. Posieux: KITLV.

Baird, Lousie. 2002. A Grammar of Kéo: An Austronesian Language of East Nusantara. PhD dissertation, Australian National University.

Bellwood, Peter. 1997. *Prehistory of the Indo-Malayan Archipelago*. Honolulu HI: University of Hawai'i Press.

Blust, Robert. 2008. Is there a Bima-Sumba subgroup? *Oceanic Linguistics* 47(1): 45–113. https://doi.org/10.1353/ol.0.0006

Blust, Robert. 2013. *The Austronesian Languages*, rev. edn. Canberra: Asia-Pacific Linguistics.

Blust, Robert & Trussel, Stephen. 2019. *The Austronesian Comparative Dictionary*, online edn. Available at <http://www.trussel2.com/ACD/> (1 June 2020).

Brown, Peter, Sutnika, Thomas, Morwood, Michael J., Soejono, Raden P., Jatmiko, Saptomo, E. Wayhu & Awe Due Rokus. 2004. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431: 1055–1061. https://doi.org/10.1038/nature02999

Donohue, Mark. 2003. Bata Lu'a ca (Palu'e-Indonesian-English dictionary). Ms.

Donohue, Mark. 2005. The Palu'e passive: From pragmatic construction to grammatical device. In *The Many Faces of Austronesian Voice Systems: Some New Empirical Studies*, I. Wayan Arka & Malcolm Ross (eds), 59–85. Canberra: Pacific Linguistics.

Elias, Alexander. 2018. Lio and the Central Flores Languages. MA thesis, Leiden University.

Erb, Maribeth. 1999. *The Manggarais*. Singapore: Times Edition.

Fenk-Oczlon, Gertraud & Fenk, August. 2008. Complexity trade-offs between the subsystems of language. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94], Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 43–65. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.05fen

Forth, Gregory. 1998. *Beneath the Volcano: Religion, Cosmology and Spirit Classification among the Nage of Eastern Indonesia*. Leiden: KITLV.

Forth, Gregory. 2004. *Nage Birds: Classification and Symbolism among an Eastern Indonesian People*. London: Routledge.

Forth, Gregory. 2008. *Images of the Wildman in Southeast Asia: An Anthropological Perspective*. London: Routledge. https://doi.org/10.4324/9780203886243

Forth, Gregory. 2009. Human beings and other people: Classification of human group and categories among the Nage of Flores (Eastern Indonesia). *Bijdragen tot de Taal-, Land- en Volkenkunde* 165: 493–514. https://doi.org/10.1163/22134379-90003630

Forth, Gregory. 2016. *Why the Porcupine is Not a Bird: Explorations in the Folk Zoology of an Eastern Indonesian People*. Toronto: University of Toronto Press. https://doi.org/10.3138/9781487510053

Fricke, Hanna. 2013. *Topics in the Grammar of Hewa: A Variety of Sika in Eastern Flores, Indonesia* [Languages of the World 45]. Munich: Lincom.

Fricke, Hanna. 2019. Traces of Language Contact: The Flores-Lembata Languages in Eastern Indonesia. PhD dissertation, Leiden University.

Gil, David. 2015. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia: The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: De Gruyter Mouton.

Hull, Geoffrey. 1998. The basic lexical affinities of Timor's Austronesian languages: A preliminary investigation. *Studies in the Languages and Cultures of East Timor* 1: 97–174.

Ingram, Arthur. 1977. *Shepherding Tools and Customs*. London: Shire Publications.

Keraf, Gregorius. 1978. Morfologi Dialek Lamalera (Morphology of the Lamalera Dialect). PhD dissertation, Universitas Indonesia, Jakarta.

Levi, Ferdinandus. 1978. A Preliminary Study of Lionese. BA thesis, Sanata Dharma Teacher's Training Institute, Jakarta.

Lewis, E. Douglas. 2010. The stranger-kings of Sikka: With an integrated edition of two manuscripts on the origin and history of the Rajadom of Sikka. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde* 257. Leiden: KITLV.

McDonnell, Bradley. 2009. A preliminary description of Ende phonology. *Journal of the Southeast Asian Linguistics Society* 2: 197–288.

McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2): 125–166.

McWhorter, John. 2007. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780195309805.001.0001

McWhorter, John. 2008. Why does a language undress? Strange cases of Indonesia. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94], Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 167–190. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.12mcw

McWhorter, John. 2011. Affixless in Austronesian: Why Flores is a puzzle and what to do about it. In *Linguistic Simplicity and Complexity: Why Do Languages Undress?* [Language Contact and Bilingualism 1] John McWhorter (ed), 223–260. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781934078402.223

McWhorter, John. 2016. Is radical analycity normal? Implications of Niger-Congo and Southeast Asia for typology and diachronic theory. In *Cyclical Change Continued* [Linguistik Aktuell/Linguistics Today 227], Elly van Gelderen (ed), 49–92. Amsterdam: John Benjamins. https://doi.org/10.1075/la.227.03mcw

McWhorter, John. 2019. The radically isolating languages of Flores: A challenge to diachronic theory. *Journal of Historical Linguistics* 9(2): 177–207. https://doi.org/10.1075/jhl.16021.mcw

Nagaya, Naonori. 2011. The Lamaholot Language of Eastern Indonesia. PhD dissertation, Rice University, Houston.

Nishiyama, Kunio & Kelen, Herman. 2007. *A Grammar of Lamaholot, Eastern Indonesia: The Morphology and Syntax of the Lewoingu Dialect* [Languages of the World/Materials 467]. Munich: Lincom.

Pampus, Karl Heinz. 1999. *Koda kiwa: Dreisprachiges Wörterbuch des Lamaholot (Dialekt von Lewolema)*. Stuttgart: Steiner.

Pareira, Mandalangi & Lewis, E. Douglas. 1998. *Kamus Sara Sikka-Bahasa Indonesia* (Sika-Indonesian Dictionar). Ende: Nusa Indah.

Ross, Malcolm. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In *Papuan Pasts: Studies in the Cultural, Linguistic and Biological History of the Papuan-speaking Peoples*, Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide (eds.), 15–66. Canberra: Australian National University.

Schapper, Antoinette & Klamer, Marian. 2014. Numeral systems in the Alor-Pantar languages. In *The Alor-Pantar Languages: History and Typology*, Marian Klamer (ed.), 285–336. Berlin: Language Science Press.

Schapper, Antoinette. 2017. Introduction to The Papuan Languages of Timor, Alor and Pantar. Volume II. Berlin: De Gruyter Mouton.

Semiun, Agustinus. 2017. Clitics of Manggarai language: The case of Central Manggarai dialect in West Flores Indonesia. *International Journal of Development Research* 7(1): 11380–11387.

Thomason, Sarah & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.

Thurston, William. 1987. *Processes of Change in the Languages of Northwestern New Britain*. Canberra: Pacific Linguistics.

van Bekkum, Wilhelm. 1944. Warloka-Todo-Pongkor: Een Brok Geschiedenis van Manggarai (West-Flores). *Cultureel Indië* 6: 144–52.

Verheijen, Jilis. 1967. *Kamus Manggarai*. The Hague: KITLV.

Verheijen, Jilis. 1982. *Komodo: Het Eiland, het Volk en de Taal* (Komodo: The Island, the People and the Language). Leiden: Martinus Nijhoff.

CHAPTER 7

From Lamaholot to Alorese

Morphological loss in adult language contact

Marian Klamer
Leiden University

Alorese is a prime example of a morphologically isolating language. This paper
traces the process of morphological simplification it has undergone by addressing
the following questions: (i) What was the morphological profile of its ancestor,
pre-Alorese? (ii) When did Alorese start to lose its morphology? (iii) Which fac-
tors caused this loss? By comparing the morphological profile of current Alorese
with its sister language, Lewoingu-Lamaholot, I conclude that the morphology of
pre-Alorese was at least as complex as current Lewoingu-Lamaholot. Pre-Alorese
underwent a process of drastic and swift morphological loss after its speakers mi-
grated to Pantar island around 1300 AD. Pre-Alorese must have had a significant
proportion of adult second language speakers who acquired it imperfectly, thus
causing its morphology to be lost. Thus, this is a good example of morphological
simplification due to imperfect adult learning in a small-scale language variety.

Keywords: morphological loss, adult language contact, Alorese, Lamaholot,
morphological reconstruction

1. Introduction

Many Austronesian languages are morphologically 'agglutinative-synthetic' with
"a relative abundance of affixes" and "morpheme boundaries [that] are usually
clear" (Blust 2009: 343). Of the Austronesian languages that are morphologically
impoverished, some extreme cases are found in western and central Flores (Blust
2009: 347–48). Here I present data on an isolating Austronesian language spoken
about one hundred kilometres east of Flores, on the islands of Pantar and Alor;
see Figure 1.

 Alorese (locally referred to as *Bahasa Alor*) has some 25,000 speakers living
along the northern coast of the island of Pantar, on the south coast of the Alor penin-
sula, and on the islets in the vicinity (Grimes, Therik & Grimes 1997; Lewis, Simons
& Fennig 2017). Klamer (2011) is a short grammar of the language. Different dialects
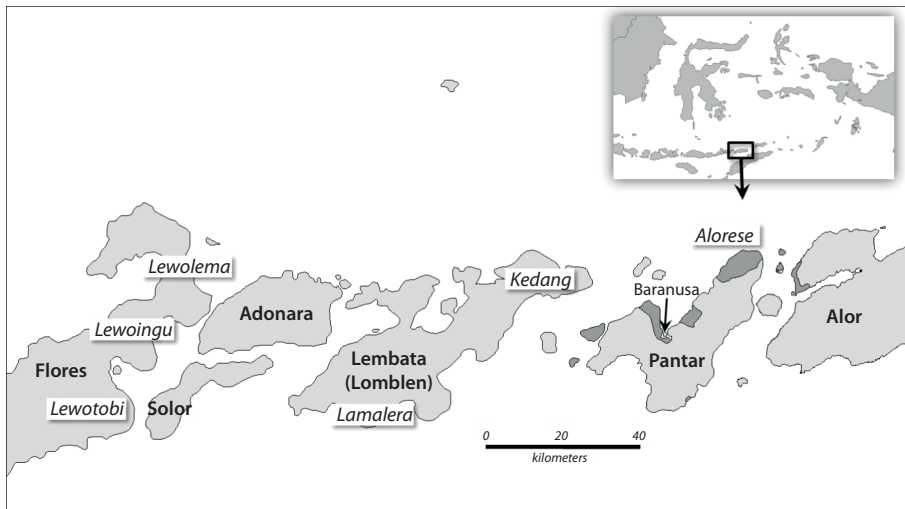
Figure 1. Alorese in its regional context. The lect names starting with *Le/La* on the western islands are Lamaholot dialects; Alorese is spoken in the dark grey areas

of Alorese are spoken, located in Munaseli, Pandai, and Baranusa on Pantar, and Alor Besar and Alor Kecil on Alor; see Figure 2. While these dialects have considerable lexical differences, their morphological and syntactic profiles are very similar. The grammatical data used in this paper is mainly from Alor Kecil; the lexical data are mostly from Baranusa (see Klamer 2011 for more information).

Alorese is a variety of the Lamaholot language/dialect chain spoken by some 325,000 people (Fricke 2019: 157–160) living on the eastern tip of Flores and neigh-bouring islands, including Solor, Lembata and Adonara; see Figure 1. Lamaholot dialects that have been described to some extent include the dialect of Lewoingu (Nishiyama & Kelen 2007), Lamalera (Keraf 1978), Lewolema (Pampus 1999), Lewotobi (Nagaya 2011), Solor (Arndt 1937; Bouman 1943; Kroon 2016), and Central Lembata (Fricke 2017a, 2017b, 2019). On the basis of lexicostatistic work in dozens of Lamaholot varieties, Keraf (1978) divides Lamaholot into three major subgroups: Western, Central and Eastern Lamaholot. Elias (2017) used Keraf's (1978) lexical data of 33 Lamaholot dialects as compiled in LexiRumah (Kaiping, Edwards & Klamer 2019) to examine regular sound changes in the dialect chain. His findings largely align with Keraf's earlier lexicostatistic work, whereby he con-firms the three subgroups Western, Central and Eastern Lamaholot. These three Lamaholot subgroups are joined at the level of Proto-Flores-Lembata, which then also includes the neighbouring languages of Sika and Kedang; see Figure 3 (Fricke 2019: 226–228). Fricke (2019) finds insufficient evidence for an innovation-defined subgroup joining these three Lamaholot groups on a more recent level.
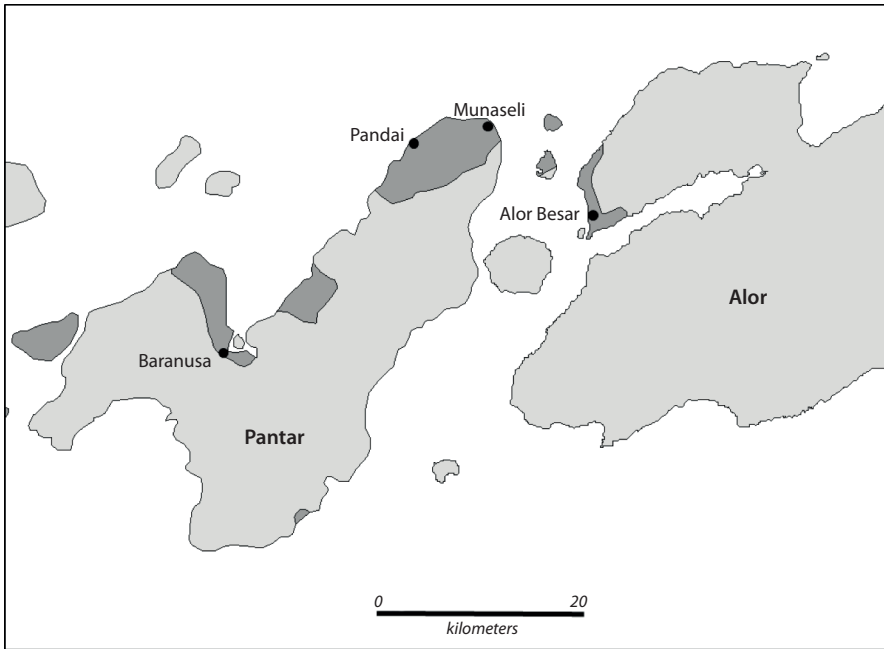
Figure 2. Locations where Alorese is spoken
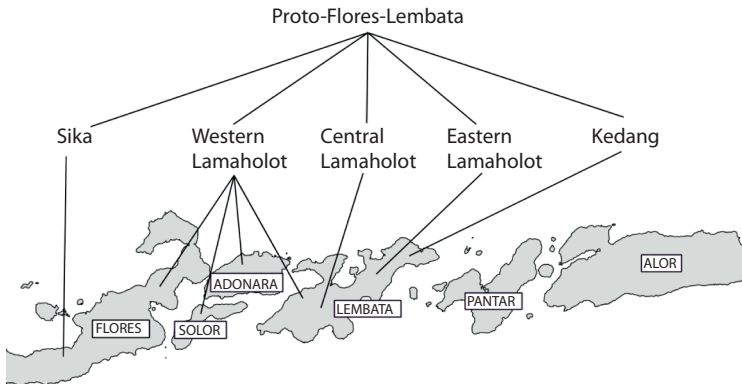
Figure 3. Geographic spread of Lamaholot varieties and their subgrouping based on regular sound changes in basic vocabulary (Elias 2017, Fricke 2019) and lexicostatistics (Keraf 1978)

Alorese appears to be most closely affiliated with the Western Lamaholot subgroup, sharing at least three innovations with it: the regular sound change *r > ʔ, a sound change in the numeral 'six' (Proto-Malayo Polynesian *enem 'six' > Proto-Flores-Lembata *ənəm > Central Lamaholot Kalikasa /ənəm/, Lerek /ənam/; Eastern Lamaholot: no data; proto-Western Lamaholot *nəmu (Adonara /namu/, Lewoingu /nəmuŋ/, Munaseli-Alorese /nəmu/, Pandai-Alorese and Baranusa-Alorese /nam:u/, Alor Besar-Alorese /namuŋ/)), and the innovation of a clause-final negator (Proto-Malayo Polynesian *salaq 'wrong' > Proto-Western Lamaholot *hala 'ɴᴇɢ' (Lamalera /hala/, Lewoingu /halaʔ/, Alorese /lahe/)) (Elias 2017; Elias p.c. February 2018; Fricke 2019: 224).

The Western Lamaholot subgroup to which Alorese belongs comprises several Lamaholot varieties. I take the Lewoingu variety described in Nishiyama and Kelen (2007) (henceforth N&K 2007) as the sister language with which I compare Alorese in this paper.[1]

This paper argues (in Section 3) that the ancestors of current Alorese speakers migrated from the region on or near Flores, where Western Lamaholot varieties are spoken, to settle on Pantar and Alor. In the process they lost contact with these Western Lamaholot varieties. The language of the ancestors of current Alorese will be referred to as 'pre-Alorese' in this paper.

In the following section (Section 2) I reconstruct the morphological profile of pre-Alorese by comparing the synchronic morphology of current Alorese and its sister Lewoingu-Lamaholot. Both languages differ greatly in terms of morphological complexity: where Lewoingu has a reasonable amount of inflectional and derivational morphology, most of which reflect proto-Malayo Polynesian forms, Alorese has virtually no morphology at all. I argue that the morphology of pre-Alorese was at least as complex as current Lewoingu, and that pre-Alorese underwent a process of morphological loss after its speakers had migrated. Then, in Section 3, I investigate when the pre-Alorese migrated away from the region where other Western Lamaholot languages are spoken. The pre-Alorese migration can be dated using evidence gleaned from accounts of oral traditions and ethnographic observations. In Section 4, I argue that the drastic and swift morphological loss observed between pre-Alorese and Alorese indicates that pre-Alorese must have gone through a stage where it had a large proportion of adult speakers who acquired it imperfectly as their second language, using Alorese as a language of trade and interethnic communication. The hypothesis that Alorese became simplified as the result of adult second language contact is based on studies of morphological simplification elsewhere in

---

1.   The morphological information in provided in Nagaya (2011) is too scanty to be used in this paper, and Kroon (2016) appeared after the current paper was written.

the world. Note that these studies typically discuss language change in relatively large speaker groups in industrial, literate societies, which is not the situation we attest for Alorese. However, recent research by Moro (2018, 2019) on second language (L2) speakers of Alorese shows that similar factors play a role in the simplification of languages spoken in small, pre-industrial societies such as Alorese. Her work indicates that the very last vestige of Alorese morphology – the subject agreement on a small number of frequent verbs – is currently also eroding. It is proposed here that similar processes of morphological simplification by adult L2 speakers caused the erosion of morphology of pre-Alorese in the past.

2.   The morphological profile of pre-Alorese

2.1   Lewoingu-Lamaholot and Alorese inflectional morphology

2.1.1   *Marking of arguments in Lewoingu-Lamaholot*
Lamaholot-Lewoingu has inflectional morphemes marking agreement of subjects on verbs, adverbs and the conjunctive element *o'on* 'and, with', while adjectives and numerals agree with the (pro)noun they modify. Lamaholot-Lewoingu pronouns and pronominal affixes are given in (1). Free pronouns are used to encode transitive (A) and intransitive (S) subjects as well as objects (P). The prefix only encodes transitive subjects (A), while the suffix encodes intransitive subjects (S). Note that this suffix also encodes nominal agreement on adjectives and numerals.

(1)   Lewoingu-Lamaholot pronouns and pronominal prefixes.

	S, A, P pronoun	A prefix	S marking suffix
1SG	*go*	*k-*	*-kən*
2SG	*mo*	*m-*	*-ko, -no*
3SG	*(ro)na*	*n-*	*-na, -nən*
1PL.EXCL	*kame*	*m-*	*-kən*
1PL.INCL	*tite*	*t-*	*-te*
2PL	*mio*	*m-*	*-ke*
3PL	*ra*	*r-*	*-ka*

(N&K 2007: 13, 31)

The A prefix obligatorily marks the subject of vowel-initial transitive verbs (N&K 2007: 98). Examples include the verbs *-a'an* 'make', *-anan* 'plait', *-ahu' wai* 'get water', *-ala/-ələ* 'pass', *-awa* 'stay', *-əwan* 'harvest, pick up, can', *-ian* 'wait', *-itə* 'sleep with', *-iu* 'hunt', *-odi* 'will', *-oi(ro)* 'know, can', *-olin* 'improve', *-urən* 'dream' (N&K 2007: 32). An illustrative paradigm is given in (2). Observe that the paradigm contains three homophonous prefixes *m-*.

(2)   Lewoingu-Lamaholot verb *-ian* 'wait' with A prefix

1SG	*k-ian*
2SG	*m-ian*
3SG	*n-ian*
1PL.EXCL	*m-ian*
1PL.INCL	*t-ian*
2PL	*m-ian*
3PL	*r-ian*

However, there are also vowel-initial verbs which do not take an agreement prefix, such as *ipu* 'sit with legs crossed', or *opən* 'tell a lie' (N&K 2007: 98), which suggests that the agreement pattern is not purely phonologically triggered, and may be eroding.

Some verbs hosting subject prefixes can function as adverbial expressions or prepositions, and appear with a 3SG default agreement *n-* prefix (N&K 2007: 103). For example, Lewoingu-Lamaholot has A-agreement on the conjunctive word *-oʾon* [oʔon] 'and, with'. In the sketch this word is variously referred to as a 'conjunction', a 'preposition' and a 'comitative' (N&K 2007: 105–108), but it is likely a verbal element 'be with' just as its cognate in Alorese is verbal. When *-oʾon* is used as a comitative predicate, the agreement prefix marks person and number of A, as in (3a). However, such contexts also allow the use of the default third person singular prefix, as in (3b). When it functions to conjoin non-nominal elements, a default 3SG prefix must be used, as in (4). For more details on the agreement marking of *-oʾon*, see N&K (2007: 10).

(3)   a.   *Go səga k-oʾon   mo*
            1SG come 1SG-with 2SG
            'I came with you'   (N&K 2007: 105)

      b.   *Go səga n-oʾon   mo*
            1SG come 3SG-with 2SG
            'I came with you'   (N&K 2007: 105)

(4)   *Mo belə n-oʾon   baʾa*
       2SG big 3SG-with heavy
       'You're big and heavy'   (N&K 2007: 103)

An intransitive argument (S) can be marked on verbs that have a 'choice of transitivity' (N&K 2007: 77). If such a verb is used intransitively, it has a suffix to encode S, and this suffix cannot appear if the verb is 'transitive'. Some examples of such optionally (in)transitive verbs are given in (5).

(5) Some Lewoingu-Lamaholot verbs taking S agreement suffixes

balik	'return'	*hode*	'burn; be angry'
bəsuk	'be born'	*horon*	'hide'
de'in	'stand'	*kirin*	'talk'
deka	'sink'	*mori*	'live'
gelu	'change'	*peku*	'turn'
həbo	'bathe'	*tannin*	'cry'
gasik	'count'	*lodo*	'go down'

An illustration is *hebo* 'bathe', used transitively in (6), and intransitively in (7). In (7a) S is marked with a pronoun and cross-referenced with an S-marking suffix. It can also be encoded with only a pronoun, as (7b) illustrates.

(6)   *Go həbo na*
    1sɢ bathe 3sɢ
    'I bathe her'   (N&K 2007: 77)

(7) a. *Go həbo-**kən***
       1sɢ bathe-1sɢ
       'I bathe'   (N&K 2007: 77)
    b. *Go həbo*
       1sɢ bathe
       'I bathe'   (N&K 2007: 77)

The S suffix cannot be used to mark a transitive agent (A), as shown in (8), and neither can the suffix denote a transitive object (P), compare (9a–b):

(8) **Go həbo-**kən** na*
    1sɢ bathe-1sɢ 3sɢ
    *'I bathe her'   (N&K 2007: 77)

(9) a. **Go həbo-**na***
       1sɢ bathe-3sɢ
       Not good for: 'I bathe her'
    b. *Go həbo **na***
       1sɢ bathe 3sɢ
       'I bathe her'   (N&K 2007: 77)

The pronominal suffix can optionally occur on adjectives in attributive function, as in *inaməlake belə / inaməlake belə-na* lit. 'man big / man big-3sɢ'. The two constructions do not differ in meaning (N&K 2007: 43).

Adjectives in a predicative or adverbial function can also take an S suffix; in these cases, the adjective has an excessive interpretation (N&K 2007: 98–99), as illustrated in (10) and (11):

(10) *Na bəlola-**na***
      3sg tall-3sg
      'He is too tall'   (N&K 2007: 98)

(11) *Mo pana lela-**ko***
      2sg walk slow-2sg
      'You walk too slowly'   (N&K 2007: 98)

In addition to adjectives in adverbial function, as *lela* 'slow' in (11), Lewoingu also has "pure adverbs" that "have no adjectival usage" (N&K 2007: 99). Such adverbs agree with the subject of the clause. Some require a prefix, for example *olo* 'earlier' in (12); others require a suffix, for example *meha'* 'alone' in (13). When applicable, the agreement on the main verb can co-occur with the agreement on adverbials such as *meha'*, as in (14).

(12) *Go səga   **k**-olo*
      1sg come 1sg-earlier
      'I came earlier / first'   (N&K 2007: 99)

(13) *Go səga   meha'-**kən***
      1sg come alone-1sg
      'I came alone'   (N&K 2007: 99)

(14) *Ra r-enun   meha'-**ka***
      3pl 3pl-drink alone-3pl
      'They drink alone'   (N&K 2007: 100)

Numerals in modifying function agree with the modified noun or pronoun in person and number using a pronominal suffix, as illustrated in (15).

(15) *Ra təlo-**ka**   səga*
      3pl three-3pl come
      'They three came'   (N&K 2007: 39)

When the modified noun is a lexical noun rather than a pronoun, the suffix on the numeral is optional and only occurs when the NP is definite, compare grammatical (16a) (definite, with suffix) with ungrammatical (16b) (indefinite, with suffix) and grammatical (16c) (indefinite, no suffix).

(16) a. *Inawae təlo-**ka**   me'enən bərin inaməlake rua-**ka** me'enən*
       woman three-3PL the   hit   man   two-3PL the
       'The three women hit the two men'   (N&K 2007: 39)

   b. **Inawae təlo-**ka**   bərin inaməlake rua-**ka***
       woman three-3PL hit   man   two-3PL
       Not good for: 'Three women hit two men'   (N&K 2007: 39)

   c. *Inawae təlo   bərin inaməlake rua*
       woman three hit   man   two
       'Three women hit two men'   (N&K 2007: 39)

Finally, Lewoingu-Lamaholot marks the addressee of imperative or hortative verbs
with a suffix, as in (17):

(17) *M-a'i -**ko**!*
       2SG-go-2SG
       'Go!'
       *Pala'e -**te**.*
       run-1PL.INCL
       'Let's run'   (N&K 2007: 75)

The agreement affixes and patterns attested in Lewoingu-Lamaholot are inher-
itances rather than innovations. The evidence for this is that similar forms and
patterns are found in its sister language Lewotobi-Lamaholot (Nagaya 2011: 103
ff.), as well as in other languages related to Lamaholot, such as Kedang (Samely
1991: 70) and Hewa, a variety of Sika (Fricke 2014: 29).

2.1.2 *Marking of possessives in Lewoingu-Lamaholot*
In Lewoingu-Lamaholot possessive constructions, pronouns and suffixes may be
used, see (18).

(18) Lewoingu-Lamaholot pronominal possessors

	pronoun (N&K 2007: 15)	suffix (N&K 2007: 13, 23, 24, 26)
1SG	*go'en*	*-kən*
2SG	*mo'en*	*-ko*
3SG	*na'en*	*-nən*
1PL.EXCL	*kame'en*	*-kən*
1PL.INCL	*tite'en*	*-te*
2PL	*mion*	*-ke*
3PL	*ra'en*	*-ka*

Illustrations with a possessor pronoun and a possessive suffix are given in (19a–b). The pronoun and suffix cannot co-occur, as shown in (19c–d). A possessor noun follows the possessed noun, in which case there is no suffix; an example is the possessor noun *guru* 'teacher' in (19e). If the possessor noun precedes the possessed noun, it must co-occur with a possessor suffix, as illustrated in (20).

(19) a. *Lango go'en*
house 1SG.POSS
'My house' (N&K 2007: 23)

b. *Lango-kən*
house-1SG.POSS
'My house' (N&K 2007: 23)

c. **Lango-kən   go'en*
house-1SG.POSS 1SG.POSS
Not good for: 'My house'

d. **Go'en   lango-kən*
1SG.POSS house-1SG.POSS
Not good for: 'My house'

e. *Lango guru*
house teacher
'A teacher's house' (N&K 2007: 24)

(20) a. *Guru   lango-nən*
teacher house-3SG.POSS
'A teacher's house' (N&K 2007: 23)

b. *Guru   lango-ka*
teacher house-3PL.POSS
'The teachers' house' (or 'faculty residence') (N&K 2007: 26)

Of the two available possessor marking strategies, the free possessor pronoun strategy (19a) is more regular and productive, while the possessor suffixing strategy (19b) is losing ground. However, the possessor suffix still exists in Lewoingu-Lamaholot, and in cases such as (20a–b), it is obligatory. The possessor patterns and forms in Lewoingu-Lamaholot were inherited from an ancestor language: similar forms and patterns are found across the Austronesian family.

2.1.3 *Marking of arguments and possessors in Alorese*

In contrast to Lewoingu-Lamaholot, Alorese has no suffixes marking subjects, possessors or any other type of agreement. It does however have prefixes marking subjects of intransitive (S) and transitive (A) verbs. The Alorese pronouns and prefix paradigms are given in (23). Note that the prefixes for 2SG, 1PL.EXCL and 2PL are homophonous, as in Lewoingu-Lamaholot.

(21) Alorese subject, object and possessor pronouns and prefixes

	Possessor pronoun	S/A prefix	S, A and P pronoun
1SG	*go*	*k-*	*go*
2SG	*mo*	*m-*	*mo*
3SG	ALIEN: *ni / ne* INAL: *no*	*n-*	*no*
1PL.EXCL	*kame*	*m-*	*kame*
1PL.INCL	*ite*	*t-*	*ite*
2PL		*m-*	*mi*
3PL		*r-*	*fe*

(Klamer 2011: 52; 60)

Alorese subject (S/A) prefixes are used on a small set of frequently used verbs; examples from my corpus (cf. Klamer 2011) are *-aka/-Vng* 'to eat', *-ei* 'to go', *-enung* 'to drink', *-oing* 'to know', and *-ong* 'to be with'. The prefixes occur only on vowel-initial verbs (see also Moro 2019). Often the subject of such inflected verbs is also expressed with an additional pronoun, which is given in brackets in (22).

(22) Two Alorese verbs with a subject agreement prefix

	-enung 'drink'	*-oing* 'know'
1SG	*(go) k-enung*	*(go) k-oing*
2SG	*(mo) m-enung*	*(mo) m-oing*
3SG	*(no) n-enung*	*(no) n-oing*
1PL.EXCL	*(kame) m-enung*	*(kame) m-oing*
1PL.INCL	*(ite) t-enung*	*(ite) t-oing*
2PL	*(mi) m-enung*	*(mi) m-oing*
3PL	*(fe) r-enung*	*(fe) r-oing*

Apart from this small set of verbs with subject inflections, Alorese has no other verbal agreement morphology. Indeed, this last 'vestige' of its morphology is currently also eroding, as will be discussed in Section 3.

Unlike in Lewoingu-Lamaholot, no productive possessor suffixes are attested in Alorese. Illustrations of Alorese (alienable) possessive constructions are given in (23).

(23) a. *Ni   uma*
         3SG.ALIEN house
         'his house'   (Alorese)

      b. *Bapa John ni   uma   being*
         father John 3SG.ALIEN house big
         'Bapa John's house is big'   (Alorese)

Alorese uses a different pronoun *no* to encode an inalienable 3rd singular pos-
sessor; see (21). In addition, both Lewoingu-Lamaholot and Alorese have a fur-
ther distinction between inalienable and alienable nouns, marked on the noun
itself by the presence of a (fossilised) final velar nasal suffix that attaches to in-
alienable body part nouns. In Alorese, the fossilised suffix is a root-final conso-
nant [-ŋ]. In Lewoingu-Lamaholot, it is [-n] on vowel-final roots, and it is [-ʔVn]
on consonant-final roots, with the V being copied from the final root vowel. In
(24), some examples are given of cognate body part nouns with fossilised pos-
sessive suffixes in both varieties, with their Proto-Malayo-Polynesian forms. In
Lewoingu-Lamaholot the nature of the possessor marker varies between a clitic
or suffix; it is optional (indicated by parentheses) in the forms in (24a), while
for the forms in (24b) it is obligatory, and in the forms in (24c) it is absent. In
Lewoingu-Lamaholot, there are also words that allow the possessor morpheme
-ng to be replaced by a modern possessor suffix from the possessor paradigm in
(18). In Alorese, however, the (originally inalienable) suffix *-ng* has become com-
pletely fossilised as a final root consonant and it has become an obligatory part of
the nominal root form; it cannot be omitted or replaced. (For more discussion, see
Klamer 2011, 2012).

(24)		Alorese (Baranusa lect)	Lewoingu-Lamaholot (N&K 2007: 174)	PMP (Blust and Trussel n.d.)	Meaning
	a.	*lima**ng***	*lima**(n)***	*qalima	'hand/arm'
		*fofa**ng***	*wəwa**(n)***	*baqbaq	'mouth'
		*rata**ng***	*rata**(n)***	unrelated *buhək	'hair'
		*fuli**ng***	*wuli**(n)***	unrelated *liqəR	'neck'
	b.	*kotu**ng***	*kotən*	unrelated *qulu	'head'
		*ale**ng***	*kola'**an***	unrelated *likud	'back'
		*lei**ng***	*lein*	unrelated *qaqay	'foot, leg'
	c.	*mata**ng***	*mata*	*mata	'eye'
		*fefele**ng***	*wewel*	unrelated *dilaq	'tongue'

In sum, pre-Alorese must have contained various inflectional paradigms of the
type that are currently still present in Lewoingu-Lamaholot: affixes encoding tran-
sitive (A) or intransitive (S) subjects, possessors, as well as agreement on adjectives
and numerals. In contrast, today's Alorese has only retained reflexes of the subject
agreement prefix on a small number of verbs, using it to encode A and S. The velar
nasal suffix/enclitic that is still recognisable as a morpheme encoding inalienable
possession in Lewoingu-Lamaholot has completely fossilised into a nominal root
consonant in Alorese.

2.2 Lewoingu-Lamaholot and Alorese derivational morphology

Lewoingu-Lamaholot has seven derivational affixes. Some of these are rather regular and productive, others are less regular, but in all derivations there is still a transparent semantic link between the base and its derivation. In this section I first present a brief summary of the various derivations, and contrast them with the derivational morphology attested in Alorese. For a fuller account and additional examples of Lewoingu-Lamaholot morphology, I refer to N&K (2007).

Prefix bə(C)-

N&K (2007: 50–51) present some forty examples of a derivation which they describe as involving the prefix *bəN-*. It is a productive derivational process. The prefix is realised as *be-* before a consonant, and as *b-, beʔ, ben*, or *ber-* before a vowel, which is why I analyse it as *bə(C)-*. The prefix *bə(C)-* derives words of various categories from nouns; and actor, action, patient and instrument nouns from verbs. Some examples are given in (25). While no current reconstruction appears to regularly account for this prefix, apparent cognates, such as Malay *ber-*, indicate that Lewoingu *bə(C)-* is derived from an Austronesian source.

(25)	N base	*rawuk*	'hair'	*bə-rawuk*	'have hairs'
		lolon	'top part'	*bə -lolon*	'high'
		wai'	'water'	*bə-wai'*	'watery'
		wola	'fat (N)'	*bə-wolan*	'fat (Adj)'
	V base	*pasak*	'shoot'	*bə-pasak*	'shooter'
		doru	'rub'	*bə-doru*	'tool for rubbing, skin scraper'
		ewik	'to slice'	*b-ewik*	'slice (N)'
		lidun	'close'	*bə-lidun*	'door'

Prefix pə-

N&K (2007: 51) provide seventeen examples of prefixing *pə-*; examples are in (26). Sometimes, the prefix co-occurs with suffix *-k*. The base of the derivation can be a noun or a verb, and the prefix derives verbs meaning 'be like the base N', and activity nouns meaning 'actor of V'. Etymologically it may be related to PMP *pa-ka-* 'treat like X' (Blust 2009: 359) and/or PMP *paR-* 'deverbal noun' (Blust 2009: 359).

(26)	N base	*tua*	'palm wine'	*pə-tuak*	'taste like palm wine'
		tana	'land'	*pə-tanak*	'feel like soil'
		kawu	'dust'	*pə-kawuk*	'grey'
		wua	'fruit'	*pə-wua*	'trees about to bear fruit'
		kua	'dregs, waste'	*pə-kua*	'feel like dregs'
	V base	*tutu'*	'speak'	*pə-tutu'*	'speaker, speaking'
		leta'	'ask'	*pə-leta'*	'asking, beggar'
		hegak	'replace'	*pə-negak*	'replacement'

Prefix kə-

N&K (2007: 52–53) present some thirty examples of the prefix *kə-*. Derived forms are nouns denoting a result of an event or a tool used in the event, but they also include derivations more vaguely related to the base word, see (27). Etymologically the prefix may be related to PMP *ka-* 'formative for abstract nouns' (Blust 2009: 359, 362).

(27) V base   *sakok*   'whisper'   *kə-sakok*   'a whisper'
      pasa   'swear'   *kə-pasa*   'oath'
      betok   'emerge'   *kə-betok*   'jump'
      Irregular *pati*   'pile'   *kə-mati*   'things piled'
      kiyuk   'close eyes'   *kə-niyuk*   'not able to open eyes'
      gate   'to hook'   *kə-nate*   'hooker'
      golo   'roll (cigarette)'   *kə-nolo*   'rolled thing (cigarette)'
      gasik   'count'   *kə-nasik*   'sum'

Infix -ən-

N&K (2007: 53–54) list some twenty examples of derivations with the infix *-ən-*. It derives nouns from verbal bases starting with a coronal consonant (i.e. /t, s, n/). The nouns denote an actor, action, state, result or tool, see (28). This prefix may be etymologically related to PAN *-in-* 'perfective, nominaliser' (Blust 2009: 372).

(28) *tali*   'add'   *t-ən-ali*   'added thing'
      teho   'wipe clean'   *t-ən-eho*   'wiping cloth'
      tubak   'stab'   *t-ən-ubak* 'stabbing tool'
      napa   'spread'   *n-ən-apa*   'things spread'
      saga   'drop from above'   *s-ən-aga*   'tool to receive a falling object'
      seok   'fry'   *s-ən-eok*   'tool for frying; fried food'

Prefix mən-

N&K (2007: 54) give seventeen examples of this derivation. The final nasal of the prefix *mən-* always replaces the initial consonant of the base. The derivation may involve an extra final consonant (N&K 2007: 54), for example *bəkə* 'angry' > *mən-əkən* 'being angry'. It derives stative verbs, or nouns (actor, action, result) or nouns referring to people with the property of the base, see (29). Etymologically it may be related to PAN *ma-* 'stative' (Blust 2009: 363–364).

(29) *ba'at*   'heavy'   *mən-a'at*   'something heavy'
      bəkə   'angry'   *mən-əkən*   'angry man, being angry'
      dira   'use a fan'   *mən-ira, nira* 'fan'
      wikak   'break in pieces'   *mən-ika'*   'piece, fraction'
      ungar   'to wound'   *mən-ungar*   'a wound'
      nange   'swim'   *mən-ange*   'swimmer'
      bohu'   'full (stomach)'   *mən-ohun*   'state of being full'

Prefix gə(C)-

N&K (2007: 49) list ten examples of derivations with this prefix. They list this prefix as *gəN-* perhaps because the final consonant in the prefix may historically have been a nasal, as it changed root-initial /p/ into /m/, and /h/ into /n/. However, this process of consonant replacement no longer regularly involves replacement with a nasal, as root initial /b/ undergoes lenition to /w/, and no consonant occurs before roots with an initial liquid (r/l). According to N&K (2007) the derived words include action, actor or result nouns; not all of the derivations involve a category change from verb to noun, see (30). Etymologically the prefix may reflect PMP *ka-* 'manner in which an action is carried out' or *ka-* 'achieved state' (Blust 2009: 362–363).

(30) | *poʾok* | 'cut' | *gəm-oʾok* | 'taking all' |
hiko	'go past'	*gən-iko*	'going past'
balik	'return'	*gə-walik*	'return (n)'
lupa	'forget'	*gə-lupa*	'forget'
redoʾ	'shake (V, N)'	*gə-redoʾ*	'shaking'

Consonant replacement

N&K (2007: 48–49) present some thirty examples of this derivation, which involves replacing the initial consonant of a root with a homorganic nasal. In most examples, the base is an activity verb, and the derivation denotes an item that is related to the activity, such as its result or location, or the instrument that is used with the activity, see (31). Etymologically this derivation may be related to PAN *ma-* 'stative' (Blust 2009: 363–364).

(31) | *pet* | 'bind' | *met* | 'belt' |
bowak	'weave cloth'	*mowak*	'patterns of cloth'
take	'cover roofs with thatch'	*nake*	'thatch'
hapen	'hang'	*napen*	'hanger'
iluʾ	'saliva'	*niluʾ*	'taste'
haman	'dance'	*naan*	'dancing place'
huro	'eat, serve, use a spoon'	*nuro*	'spoon'

In sum, in the seven types of derivational morphemes of Lewoingu-Lamaholot we can still see a semantic relation between the independently used base form and the word forms derived from them, although the relation has often become somewhat opaque. All seven affixes seem to be inherited Malayo-Polynesian/Austronesian forms. That is, they are reflexes of ancient derivational morphemes that have been around in some form or other in the ancestor language(s) of Lamaholot.

In contrast, the only word formation process attested in Alorese is reduplication. Alorese verbs and adverbs undergo full reduplication to indicate iterative or intensive activity, as shown in (32). Nominal reduplication denotes plural diversity, as in (33a–b). There are also reduplicative forms for which no root forms exist, as in (34a–b).

(32) *Akhirnya, no gena-gena dapat lahe*
      finally   3sg RDP-search find   NEG
      'Finally he searched and searched [but] did not find him'   (Klamer 2011: 92)

(33) a.   *gambe-gambe*
           RDP-grandfather
           'grandfathers'
      b.   *ina-ina*
           RDP-mother
           'mothers'   (Klamer 2011: 39)

(34) a.   *kapu-kapu* (*kapu),
           'firefly'
      b.   *uli-uli* (*uli)
           'fable'   (Klamer 2011: 39)

Full reduplications with similar functions are also attested in Lewoingu-Lamaholot, so in terms of their reduplication strategies there is no salient contrast between the two varieties.[2] The fact that reduplication occurs in both Lewoingu and in Alorese is taken here to indicate that reduplication was a part of pre-Alorese that was maintained in Alorese. But then the question can be asked, why didn't Alorese reduplication get lost like the other derivational morphology? My suggestion is that this may be due to the unique character of reduplication as a word forming process. Unlike derivational affixes, full reduplication has an iconic relation between form and meaning, and as such is also attested in improvised language behaviour, as a universal combinatory principle to derive new words (Muysken 2013: 716). This is why, for example, reduplication can spontaneously emerge in contact varieties such as creoles (Bakker and Parkvall 2005). So, if a language with derivational affixes and reduplication (like pre-Alorese) undergoes morphological loss due to language contact, we expect the affixes to be most affected. Reduplication will be affected less or not at all, as it is one of the basic strategies universally employed by imperfect language speakers to create new forms.

Unlike Lewoingu-Lamaholot, Alorese has little derivational morphology. Virtually all the morphologically complex Lewoingu-Lamaholot forms discussed above are simple forms in Alorese. In my Alorese (Baranusa lect) lexicon of approx. 600 items, I found only five words that contained a possibly fossilised affix;

---

2.   In Nishiyama and Kelen's (2007: 60) description, reduplication is described as "[…] not very productive, [but] occasionally observed." However, this quote seems to refer to reduplication of non-nominal forms; the reduplication of nouns to encode 'plural', discussed elsewhere in the sketch (p. 44) appears to be regular and productive.

they are listed in (35). The first column gives the Alorese word, the second column its meaning, the third column its cognate in Lewoingu-Lamaholot. Column four presents the PMP proto-form where known, suggesting that the current reflex in Alorese is indeed historically a morphologically complex word. The last column presents some cognate forms in other languages of the region for comparison (see Kaiping et al. 2019 for additional forms and references).

(35)   Alorese (Baranusa dialect) words with fossilised affixes
in a set of approx. 600 items

Alorese	Meaning	Lewoingu-Lamaholot	PMP (Blust and Trussell n.d.)	Example cognates in languages of the region
*pa*l*ae*	'run'	*pela?e* 'run'	PMP *laRiw	Hewa/Sika *p-lari,* Tetun Terik *ha-lai*
*ka*l*ake*	'man, husband'	*lake* 'husband'	PMP *laki 'male, masculine, man'	Hewa & Sika *la?i,* Tokodede *laki,* many Lamaholot varieties have either simple *lake* or complex *kelake*
*kaf*ae*	'wife'	*kewae* 'wife'	PMP *bahi	Many other Lamaholot varieties also have *kewae*
*ka*puhor	'navel'	*kepuhur* 'navel'	PMP *pusej	Hewa/Sika *puher,* Alorese dialects on Pantar have *puhor,* some Lamaholot varieties have *kepuser* or *kepuher*
kenamu	'fly'	*kenamuk* 'fly'	PMP *ñamuk 'mosquito'	Indonesian *nyamuk* 'mosquito', Sika *əmu* 'mosquito'

Apart from these fossilised forms, I have not attested other morphologically complex words in Alorese. In other words, except for full reduplication, Alorese did or does not have productive derivational processes. And with a few exceptions, the derivational morphology that is still attested in its sister Lewoingu-Lamaholot has been lost completely in Alorese. The only derivational prefix of which a few remnants have been retained is *kə-*.

2.3   Summary: The morphology of pre-Alorese

Table 1 and Table 2 summarise the inflectional and derivational features of today's Lewoingu-Lamaholot and Alorese.

Table 1. Summary of inflectional features in Lewoingu-Lamaholot and Alorese

	Lamaholot	Alorese
consonantal subject prefix on vowel-initial verbs	yes	few frequently used verbs
suffix to mark S on verbs	yes	no
agreement on adjectives and adverbs	yes	no
agreement on numerals	yes	no
agreement on the conjunctive element -o'on 'and, with'	yes	no
possessor suffix	yes	no
inalienable possessor suffix/enclitic	yes	fossilised

	Lamaholot	Alorese
Consonant replacement	yes	no
Prefix bə(C)-	yes	no
Prefix pə-	yes	no
Prefix kə-	yes	no
Infix -ən-	yes	no
Prefix mən-	yes	no
Prefix gə(C)-	yes	no
Reduplication	yes	yes

As argued in Section 2, the morphology attested in Lewoingu-Lamaholot is clearly of Malayo-Polynesian origin. Pre-Alorese probably had at least as much morphology as today's Lewoingu-Lamaholot: possessor suffixes, inalienable suffixes, distinct pronominal prefixes for transitive and intransitive subjects, suffixes for intransitive subjects and nominal agreement, and at least seven derivational prefixes.

In pre-Alorese, the variable order in possessor marking structures was regularised to a construction where possessors can only precede possessed nouns; and the final nasal morpheme on inalienable nouns became reinterpreted as a final consonant of the root noun (Klamer 2012). Of all the morphology in pre-Alorese, today's Alorese only retained a few frequent verbs with subject inflection, a tiny number of words containing remnants of derivational prefixes, in particular kə-, and reduplication. The rest of its morphology has been lost.

3.   When, why, and how pre-Alorese became isolating

Morphological loss that is as radical and fast as observed in the passage from pre-Alorese to Alorese suggests that the language went through a stage of imperfect second language learning. In this section I date the pre-Alorese migration from where the Western Lamaholot group was originally located (Section 3.1), investigate under what social circumstances the language could have lost its morphology (Section 3.2), and argue that adult second language speakers were the agents of the change (Section 3.3).

3.1   Dating the migration of pre-Alorese to Pantar Island

Sometime in the past, speakers of pre-Alorese moved away from the region where Western Lamaholot languages were/are spoken, on or near the eastern part of Flores Island (see Section 1). Using historical sources and oral traditions, this migration can be dated to have occurred 600–700 years ago. The evidence for this is presented in Klamer (2011, 2012) and is summarised here for convenience.

In an early Dutch geographic journal article (Anonymous 1914: 75–78)[3] the "non-indigenous" coastal populations of Pantar and Alor are distinguished from indigenous mountain populations. The only non-indigenous coastal people on the islands are the Alorese. The anonymous report also recalls a local legend that the coast of northwest Pantar (where Pandai is located, see Figure 2) was the first to be populated by these coastal people. The legend has it that a "colony of Javanese" settled there "5 to 600 years ago" [in 1914] (Anonymous: 77). However, the same source includes a footnote (p. 89) which explains that the notion *orang djawa* (lit. 'Javanese people') applies to everyone who comes from other parts of the archipelago. In other words, the so-called "Javanese" coastal settlers mentioned in the legend were people from "overseas", but not from Java.[4] The coastal communities in Pandai, Munaseli, Baranusa and Alor Besar are all Alorese speaking,[5] and do not

---

3.   This article was written by one or more unknown editors of the journal it was published in. A footnote explains that the two major sources for the article were the "Military Memories" written to report on military expeditions on the islands in 1910 and 1911, and an earlier report of a geological expedition by R. D. M. Verbeek in 1899, published as "Molukken Verslag" in *Jaarboek van het Mijnwezen in Ned. Oost-Indie, jaargang 1908*.

4.   Compare Kambera (a language of Sumba) *tau Jawa* 'stranger' (lit. 'Javanese person') and *tau Jawa bara* 'westerner' (lit. 'white Javanese person'), where *Jawa* also denotes 'stranger' (Onvlee 1984: 115).

5.   The Alorese spoken in Munaseli is referred to as *Bahasa Muna* or *Kadire Senaing* by Rodemeier (2006).

show a trace of Javanese linguistic or cultural influence. Instead, the close linguistic and cultural ties between today's Alorese and Lamaholot suggest that the colony of *orang djawa* that settled on Pandai according to the legend were in fact pre-Alorese speakers originating from (the vicinity of) east Flores, where Western Lamaholot languages are spoken.

The legend about the founding of Pandai is also the first of two legends reported in Lemoine (1969) and cited in Barnes (1973: 86, 2001: 280) and Rodemeier (2006). Today, it is still part of the oral history of the Alorese: in 2016, Francesca Moro recorded a story in Pandai in which the current king of Pandai, Rajab Suleiman Abu Bakar, tells the same legend about a Javanese king who came to Pantar and founded the village of Pandai, dating his arrival at 1,310 AD (Moro 2018). The legend recounts that two Javanese brothers, Aki Ai and his younger brother Mojopahit, sailed to Pantar, where Aki Ai treacherously abandoned Mojopahit. Mojopahit's descendants eventually colonised Pandai, Baranusa, and Alor Besar. The second legend in Lemoine (1969) tells of Javanese immigrants killing the king of Munaseli (another kingdom located further eastwards on the north coast of Pantar, see Figure 2) and destroying the Munaseli kingdom sometime between 1,300–1,400 AD. These Javanese immigrants were allied to the kingdom of Pandai. The defeated Munaseli population fled to Alor. In short, ethnographic observations report oral traditions which all agree that the Pandai and Munaseli kingdoms were in place around 1,300 AD in Pantar, and that they were established by non-indigenous groups who also colonised Baranusa and Alor Besar. From linguistic and cultural evidence, it can be inferred that these groups spoke pre-Alorese, a Western Lamaholot language.

To conclude, pre-Alorese speaking groups migrated away from the Western-Lamaholot-speaking area in the east Flores region 600–700 years ago,[6] and have been present in north Pantar at least from the 14th century. After initially settling on Pantar, they also settled in the westernmost coastal parts of Alor Island.

3.2   Alorese as a language of trade and interethnic communication

In known cases where language contact has led to loss of morphological complexity it involves adults as L2 learners who simplify non-native morphological structures, as in Afrikaans (den Besten 1989), Old English (Trudgill 2016) and in adult second language Dutch (Blom, Polišenská & Weerman 2006). While morphological loss *can* be an independent development, the relatively quick and drastic reduction

---

6.   The notion "east Flores region" is intentionally left vague. If the homeland of Lamaholot was Lembata (as argued in Fricke 2019), then it is also possible that the Western Lamaholot group started out in Lembata, with some of the Western Lamaholot languages going west, to Flores, and some, like pre-Alorese, going east.

of all morphology in Alorese suggests that the language went through a stage of imperfect language learning, by adult speakers who acquired it as a second language. This converges with what we know about the social circumstances in which Alorese was used.

The ethnicity of today's speakers of Alorese is rather mixed. Barnes (1973: 86, following Anonymous 1914: 77, 89) mentions that "the Coastal Alorese speaking coast-dwellers of Alor and Pantar […] have slowly formed from a mixture of Selayarese (Macassarese-Buginese), Solorese and Javanese and people of the former Muna (on the northern tip of Pantar)[7] and, on Pantar, also from people from Ternate." This ethnic mixing with people from overseas is to be expected, as in pre-colonial times (at least from the 12th century onwards), busy trade relations existed between the Moluccas, groups from Java, Sulawesi, possibly China, Vietnam, and northern India; and Pantar and Alor were part of these trading routes. In colonial times, Portuguese and other ships sailed the narrow but extremely deep gulf between Pantar and Alor on their way from the Moluccas and Makassar in the north, to the islands of Timor and Sumba in the south, in order to buy wax and sandalwood. When sailing this narrow gulf they had both Alor and Pantar within sight. Traders and soldiers must have frequented the islands in the 16th century when travelling between Larantuka (on Flores) and Dili (on Timor), as Alor and Pantar are located in between these two Portuguese settlements. (See Hägerdal 2010 and Klamer 2010 for more historical details and references.) In other words, after the pre-Alorese speakers settled on the coasts of Pantar and Alor some 600–700 years ago, they may very well have intermarried with members of overseas groups, as suggested by Anonymous (1914: 77). In addition, the (pre-)Alorese coastal populations also had intensive trade relations with the inland non-Austronesian populations of Alor and Pantar, exchanging e.g. fish and woven cloth for local food crops grown in the mountains (cf. Anonymous 1914: 76, 81–82). The Alorese speaking groups on the coasts of Pantar and Alor were initially relatively small and scattered (for example, Anonymous 1914: 89–90 mentions groups of 200, 300, and 600 people). As newcomer clans inhabiting coastal locations geographically remote from each other, many Alorese clans must have been outnumbered by their non-Austronesian neighbours, so it is expected that they also exchanged women with the non-Austronesian mountain clans in their immediate vicinity. Today, the Alorese on Pantar mix and intermarry with speakers of non-Austronesian languages such as Teiwa, Blagar, Kaera, Western Pantar, and Kroku; and on Alor the Alorese mix with speakers of non-Austronesian Adang and Kabola. Wellfelt (2016) presents evidence that such intensive cultural contacts also existed in the past.

---

7. The 'Muna' referred to here is an abbreviation of 'Munaseli', the mythical kingdom located in north Pantar.

Moreover, it seems that until not so long ago, exchange of slaves was also common in at least some parts of Alor and Pantar. For example, speakers of Teiwa (non-Austronesian, northwest Pantar) noted that in former days, they would sell, send, or give away people that were "useless to the clan" to the Alorese people living in Baranusa (Klamer 2010: 41 footnote 2).

It is likely that pre-Alorese was used as a language of trade and interethnic communication. Several sources mention that Alorese was used as a language of wider communication in the Alor-Pantar region till at least the mid 1970's (Stokhof 1975: 8; Grimes, Therik & Grimes 1997: 57); though note that Alorese is named 'Lamaholot' in these sources). Local speakers of Teiwa mentioned to me in 2004 that Alorese was used as a lingua franca in their area until well into the 1950s after which it was gradually replaced by Indonesian.

In sum, the pre-Alorese immigrants who had settled on the coast of Pantar around 1,300 AD had trade relations, cultural contacts, and exchanges of people through bilingual marriages and slavery; both with individuals from overseas and with neighbouring non-Austronesian clans. As a result of these contacts, there must have been significant numbers of adults who learned pre-Alorese as a second language. Their learners' omissions became part of a morphologically simplified variety that is Alorese today.

The situation of the pre-Lewoingu-Lamaholot speakers must have been very different. Unlike the (pre-)Alorese, the Lewoingu-Lamaholot speakers are not scattered in small groups along long stretches of coast, separated from each other by non-Austronesian speakers. Rather, the Lamaholot varieties cover a large area in the Flores-Lembata region, with bigger, contiguous groups of speakers living next to each other. The (pre-)Lewoingu-Lamaholot speakers did (and do) not have any long-term intensive contact with non-Austronesian speakers: in the last 600–700 years there were no longer any non-Austronesian languages spoken in the east Flores-Lembata region.[8] Finally, unlike Alorese, Lewoingu-Lamaholot has not been used a lingua franca. Indeed, in the period under consideration the entire Flores-Lembata region was monolingual Lamaholot, so that a lingua franca was not needed. The different sociolinguistic situations of (pre-)Lewoingu-Lamaholot and (pre-)Alorese thus explain why the former retained its morphology, and the latter lost it.

---

8.   Klamer (2012) presents evidence that once there were non-Austronesian languages spoken in the Flores-Lembata region on the basis of various non-Austronesian features found in all Lamaholot varieties. She argues that these features must have entered the languages at an ancient, prehistorical stage. In Fricke (2019) more detailed arguments are presented showing that the non-Austronesian features of the Lamaholot varieties were already part of their ancestor language.

3.3 Alorese was acquired by adult speakers

The loss of inflectional and derivational morphological categories in Alorese can be seen as an instance of simplification that occurred as a result of language contact, where non-native adults learned the language imperfectly (Trudgill 2010: 310–13).

Morphology is a complex, embedded part of grammatical structure with a relatively small functional load. Inflectional morphology is known to be seriously problematic for post-adolescent second language learners who have passed the 'critical threshold' (Lenneberg 1967) for language acquisition (Kusters 2003: 21, 48). And derivational morphology that is partly lexicalised, irregular and semantically opaque, represents arbitrary grammatical patterns that must be learned without a transparent relation between form and function, which is equally difficult for post-threshold adult language learners. Morphological features of a second language that are not part of a speaker's first language are more likely to be simplified or generalised (Jarvis & Odlin 2000: 552–553).

The morphological profile of the non-Austronesian languages surrounding Alorese is quite different from Alorese. The non-Austronesian languages generally lack verbal subject agreement prefixes and instead use verbal prefixes to mark objects; while possessors are marked as prefixes rather than suffixes. Overall, they have little derivational morphology on nouns or verbs (cf. the overviews in Klamer 2017 and Holton & Klamer 2018). The morphological discrepancies between Alorese and the non-Austronesian first languages of the adult speakers that acquired Alorese likely accelerated the loss of pre-Alorese morphology. In the process of learning a second language, adult learners apply principles that reduce the amount of morphology and increase the one-to-one relationship between form and meaning (Kusters 2003). In order for the simplified patterns to stabilise and carry on through the generations, the contact must have involved a community of bilinguals with a large number of L2 speakers – if there had been only a few L2 speakers, their morphologically reduced language would not have had much impact on the community language. While the simplifying L2 may initially have been used as a trade language or lingua franca, for the changes to become entrenched and passed on to the next generations as part of their L1, it must have been used as an L2 in a variety of wider communicative contexts apart from trade alone. In sum, the contact must have been long-term, intense, and multi-purpose (Kusters 2003; Trudgill 2011).

Sociolinguistic and census data collected through fieldwork by Moro (2019) reveals that today's Alorese as spoken on Alor has *as many* L2 speakers as it has L1 speakers, and that both types of speakers are interacting with each other in various cultural domains on a regular basis. In an experimental study focusing on the loss of the last remnants of Alorese morphology (the subject agreement prefixes on vowel initial verbs, see Section 2.1), Moro compares the production of subject prefixes

in a group of 6 female Alorese L1 speakers and a group of 12 female Alorese L2 speakers who all have the non-Austronesian language Adang as their L1. The results show that the L2 speakers not only make significantly more agreement errors than the L1 speakers, but also that they tend to use only a single 'default' subject agreement marker. In other words, the last vestige of Alorese morphology, subject agreement on verbs, is currently disappearing. Moro identifies the large proportion of L2 speakers in the community and the regular, multi-faceted contact between speakers of L1 Adang and L2 Alorese as the two driving forces that are crucial for this process of simplification.

Assuming that the contact situation between Alorese and Adang speakers today is not fundamentally different from the contact situation that existed between Alorese and other non-Austronesian speaker groups in the past, I suggest that Alorese was spoken in bilingual communities with large numbers of L2 speakers for at least several centuries. In a situation where half of a community consists of L2 speakers, the bilingual nature of the community is very stable, and as long as these demographics do not change, such a community can continue to exist for centuries without shifting to either of the languages. Such an ongoing long-term stable bilingual situation would have led to the erosion of Alorese morphology and the isolating nature of the language today.

The fact that Alorese simplified so much suggests that the change was caused by adults, not children. A long-term stable bilingual situation involving adult-second language learning usually leads to grammatical simplification. While such a situation can also induce morphological complexification, when it does it involves (pre-adolescent) childhood bilingualism (Kusters 2003; Trudgill 2011; Ross 2013).[9] Bilingual children are able to create new morphological forms by dissociating grammatical features from their original forms and remapping them on new forms (Sánchez 2006) while adult L2 speakers are less able to do so.

If the language contact was so intense that it resulted in morphological loss in Alorese, we might expect other levels of linguistic structure to also be affected (Thomason 2001). If (pre-) Alorese is (or was) segmentally or phonotactically more complex than the languages it is (was) in contact with, we would expect some of that complexity to be lost in adult L2 speakers.[10] However, if its segment inventory

---

9. Moro (2018) revealed that the plural word *hire* in Alorese emerged through contact with non-Austronesian languages, constituting a complexification of Alorese grammar with bilingual children as agents (2018: 194).

10. Long-term contact involving adult language contact may produce smaller inventories through imperfect learning, pidginisation, and simplification, while child bilingualism may produce large inventories through borrowing (Trudgill 2004: 314).

is (was) similar, and its phonotactics equally or less complex than the L1 of these speakers, we expect no contact-induced changes. This is indeed what we find: current Alorese and Lewoingu-Lamaholot, as well as the other Lamaholot varieties are all similar in terms of segment inventory and phonotactics; there is no reason to assume that (pre-)Alorese was phonologically more complex in the past and became phonologically simplified by non-Austronesian L2 speakers. (Nor did it become more complex under non-Austronesian influence.) Lexically, we expect to find non-Austronesian loans in the Alorese lexicon which set it apart from its Western Lamaholot relatives. One example is the fact that the Alorese decimals (e.g., *kar-to* 'ten', *kar-ua* 'twenty') contain reflexes of Proto-(Timor) Alor Pantar *qar- 'ten', unlike all the other Lamaholot varieties (Kaiping et al. 2019). A few more examples of loans from non-Austronesian neighbouring languages are given in (36). Note that the dialect of Alor Besar and the dialect of Pandai borrowed different words for 'mud', from different sources.

(36)   Some non-Austronesian loans attested in Alorese dialects

Alorese	Dialect	Meaning	Source	Language	Lewoingu-Lamaholot	PMP (Blust and Trussel n.d.)
klita(ʔ), *kalita*	Pandai, Munaseli, Baranusa, Alor Besar	'dirty'	*klitaʔ* *klitak* *kəlitah*	Teiwa Blagar-Bakalang Blagar-Kulijahi	*milaŋ*	unknown
lamiŋ	Pandai, Munaseli, Baranusa	'to wash'	*lamiŋ*	Western Pantar	*baha, puhu*	*basəq
para	Alor Besar	'mud'	*para* *parah*	Kabola Adang	*walaŋ*	*pitek
buta	Pandai	'mud'	*buta* *buta*	Blagar-Bakalang Blagar-Bama	*walaŋ*	*pitek

In conclusion, the loss of inflectional and derivational morphological categories in pre-Alorese is an instance of simplification that occurred during the past 600–700 years, as the result of a long-term, stable situation of bilingualism where a large group of adult non-native speakers learned the language imperfectly.

4.   Conclusions

The ancestor of today's Alorese, pre-Alorese, had morphology of Malayo-Polynesian origin, including possessor suffixes, inalienable suffixes, pronominal prefixes for transitive subjects, pronominal suffixes for intransitive subjects and nominal agreement, as well as at least seven derivational prefixes. Of these, Alorese retained the subject prefix on a small set of frequent verbs, using it to encode both transitive and intransitive subjects. The final nasal morpheme on inalienable nouns was reinterpreted as a root-final consonant segment. In a tiny number of words remnants of derivational prefixes, in particular *kə-*, can be found. The rest of the derivational prefixes and inflectional paradigms were completely lost.

This morphological loss happened after the pre-Alorese speaking group migrated from the Lamaholot area at least 600–700 years ago, settled in north Pantar in the early 14th century, and from there moved on to Alor. The pre-Alorese immigrants who settled on the coastal regions of Pantar and Alor were in contact with different speakers from overseas as well as with non-Austronesian speakers in the mountains. Contacts involved barter trade, and included exogamy resulting in bilingual marriages, as well as slavery. As a result of these contacts, adults who originally spoke a non-Austronesian language acquired Alorese as a second language. Their learners' omissions became part of a morphologically simplified variety that developed into today's isolating Alorese.

Acknowledgements

Funding

References

Anonymous. 1914. De eilanden Alor en Pantar, residentie Timor en Onderhoorigheden (The islands of Alor and Pantar, Residence Timor and Dependencies). *Tijdschrift van Het Koninklijk Nederlandsch Aardrijkskundig Genootschap* 31: 70–102.

Arndt, Paul S. V. D. 1937. *Grammatik der Solor-Sprache*. Ende, Flores: Arnoldus Drukkerij.

Bakker, Peter & Parkvall, Mikael. 2005. Reduplication in pidgins and creoles. In *Studies on Reduplication*, Bernard Hurch & Veronika Mattes (eds), 511–531. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110911466.511

Barnes, Robert H. 1973. Two terminologies of symmetric prescriptive alliance from Pantar and Alor in Eastern Indonesia. *Sociologus, Zeitschrift für empirische Ethnosoziologie und Ethnopsychologie* 23(1): 71–88.

Barnes, Robert H. 2001. Alliance and warfare in an Eastern Indonesian principality – Kédang in the last half of the nineteenth Century. *Bijdragen tot de Taal-, Land- en Volkenkunde* 157(2): 271–311. https://doi.org/10.1163/22134379-90003810

den Besten, Hans. 1989. From Khoekhoe foreignertalk via Hottentot Dutch to Afrikaans: The creation of a novel grammar. In *Wheels within Wheels; Papers of the Duisburg Symposium on Pidgin and Creole Languages*, 207–249. Frankfurt: Peter Lang.

Blom, Elma, Polišenská, Daniela & Weerman, Fred. 2006. Effects of age on the acquisition of agreement inflection. *Morphology* 16: 313–336. https://doi.org/10.1007/s11525-007-9110-1

Blust, Robert A. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Blust, Robert A. & Trussell, Stephen. n.d. *The Austronesian Comparative Dictionary*. <http://www.trussel2.com/ACD/> (1 June 2020).

Bouman, Maarten A. 1943. De Aloreesche dansplaats. *Bijdragen Tot de Taal-, Land- En Volkenkunde* 102: 481–500. https://doi.org/10.1163/22134379-90001235

Elias, Alexander. 2017. *The internal diversity of Lamaholot*. Ms, Leiden University.

Fricke, Hanna. 2014. *Topics in the Grammar of Hewa: A Variety of Sika in Eastern Flores, Indonesia*. Munich: Lincom.

Fricke, Hanna. 2017a. Nouns and pronouns in Central Lembata Lamaholot (Austronesian, Indonesia). *Wacana* 18(3): 746–771. https://doi.org/10.17510/wacana.v18i3.635

Fricke, Hanna. 2017b. The rise of clause-final negation in Flores-Lembata, Eastern Indonesia. *Linguistics in the Netherlands*, 47–62. https://doi.org/10.1075/avt.34.04fri

Fricke, Hanna. 2019. Traces of Language Contact. The Flores-Lembata Languages in Eastern Indonesia. PhD dissertation, Leiden University. Utrecht: LOT Publications. <https://www.lotpublications.nl/traces-of-language-contact> (1 June 2020).

Grimes, Charles E., Therik, Tom & Grimes, Barbara D. 1997. *A Guide to the People and Languages of Nusa Tenggara*. Kupang, Indonesia: Artha Wacana Press.

Hägerdal, Hans. 2010. Cannibals and pedlars. *Indonesia and the Malay World* 38(111): 217–246. https://doi.org/10.1080/13639811.2010.489356

Holton, Gary & Klamer, Marian. 2018. The Papuan languages of East Nusantara and the Bird's Head. In *The Languages and Linguistics of the New Guinea Area*, Bill Palmer (ed.), 569–640. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110295252-005

Jarvis, Scott & Odlin, Terence. 2000. Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition* 22: 535–556. https://doi.org/10.1017/S0272263100004034

Kaiping, Gereon, Edwards, Owen & Klamer, Marian (eds). 2019. *LexiRumah 3.0.0*. Leiden: Leiden University, Centre for Linguistics. <https://lexirumah.model-ling.eu/lexirumah/> (12 November 2019).

Keraf, Gregorius. 1978. Morfologi dialek Lamalera (Morphology of the Lamalera dialect). PhD dissertation, Universitas Indonesia, Jakarta.

Klamer, Marian. 2010. *A Grammar of Teiwa*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110226072

Klamer, Marian. 2011. *A Short Grammar of Alorese (Austronesian)* [Languages of the World Materials 486] Munich: Lincom.

Klamer, Marian. 2012. Papuan-Austronesian language contact: Alorese from an areal perspective. In *Melanesian Languages on the Edge of Asia: Challenges for the 21st Century* [Language Documentation & Conservation Special Publication 5], Nicholas Evans & Marian Klamer (eds), 72–108. Honolulu HI: University of Hawai'i Press. <http://scholarspace.manoa.hawaii.edu/handle/10125/4561> (1 June, 2020).

Klamer, Marian. 2017. The Alor-Pantar languages: Linguistic context, history and typology. In *The Alor-Pantar Languages: History and Typology*, 2nd edn, Marian Klamer (ed.), 1–48. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.437098

Kroon, Yosep Bisara. 2016. A Grammar of Solor-Lamaholot: A Language of Flores, Eastern Indonesia. PhD dissertation, University of Adelaide. <http://hdl.handle.net/2440/102380> (1 June 2020).

Kusters, Christiaan Wouter. 2003. Linguistic Complexity: The Influence of Social Change on Verbal Inflection. PhD dissertation, Leiden University. Utrecht: LOT Publications.

Lemoine, Annie. 1969. Histoires de Pantar. *L'Homme* 9(4): 5–32.

Lenneberg, Eric. 1967. *Biological Foundations of Language*. New York NY: Wiley. https://doi.org/10.1080/21548331.1967.11707799

Lewis, M. Paul, Simons, Gary F. & Fennig, Charles D. (eds). 2017. *Ethnologue: Languages of the World*, 20th edn. Dallas, Texas: SIL International. <http://www.ethnologue.com> (1 June 2020).

Moro, Francesca. 2018. The plural word *hire* in Alorese: Contact-induced change from neighboring Alor- Pantar languages. *Oceanic Linguistics* 57(1): 177–198. https://doi.org/10.1353/ol.2018.0006

Moro, Francesca. 2019. Loss of morphology in Alorese (Austronesian): Simplification in adult language contact. *Journal of Language Contact* 12: 1–26. https://doi.org/10.1163/19552629-01202005

Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(4): 709–730. https://doi.org/10.1017/S1366728912000727

Nagaya, Naonori. 2011. The Lamaholot Language of Eastern Indonesia. PhD dissertation, Rice University. <http://hdl.handle.net/1911/70366> (1 June 2020).

Nishiyama, Kunio & Kelen, Herman. 2007. *A Grammar of Lamaholot, Eastern Indonesia: The Morphology and Syntax of the Lewoingu Dialect* [Languages of the World/Materials 467]. Munich: Lincom.

Onvlee, Louis. 1984. *Kamberaas (Oost-Sumbaas)-Nederlands Woordenboek*. Dordrecht: Foris.

Pampus, Karl-Heinz. 1999. *Koda Kiwa: Dreisprachiges Worterbuch des Lamaholot (Dialekt von Lewolema)* [Vol. 4. Abhandlungen Fur Die Kunde Des Morgenlandes 52] Stuttgart: Steiner.

Rodemeier, Susanne. 2006. *Tutu Kadire in Pandai-Munaseli – Erzählen und Erinnern auf der vergessenen Insel Pantar (Ostindonesien)* [Passauer Beträge zur Südostasienkunde 12]. Berlin: LIT Verlag.

Ross, Malcolm. 2013. Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact* 6(1): 5–47. https://doi.org/10.1163/19552629-006001002

Samely, Ursula. 1991. *Kedang (Eastern Indonesia): Some Aspects of its Grammar* [Forum Phoneticum 46]. Hamburg: Helmut Buske.

Sánchez, Liliana. 2006. Bilingual grammars and creoles: Similarities between functional convergence and morphological elaboration. In *L2 Acquisition and Creole Genesis* [Language Acquisition and Language Disorders 42], Claire Lefebvre, Lydia White & Christine Jourdan (eds), 277–294. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.42.17san

Stokhof, Wim A. L. 1975. *Preliminary Notes on the Alor and Pantar Languages (East Indonesia)*. Canberra: Pacific Linguistics.

Trudgill, Peter. 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8: 305–320. https://doi.org/10.1515/lity.2004.8.3.305

Trudgill, Peter. 2010. Contact and sociolinguistic typology. In *The Handbook of Language Contact*, Raymond Hickey (ed.), 299–319. Malden MA: Wiley-Blackwell. https://doi.org/10.1002/9781444318159.ch15

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

Trudgill, Peter. 2016. Contact-related processes of change in the early history of English. In *The Cambridge Handbook of English Historical Linguistics*, Merja Kytö & Paiva Pahta (eds), 318–334. Cambridge: CUP. https://doi.org/10.1017/CBO9781139600231.020

Wellfelt, Emilie. 2016. Historyscapes in Alor. Approaching Indigenous Histories in Eastern Indonesia. PhD dissertation, Linnaeus University, Växjö, Sweden.

Double agent, double cross?

Or how a suffix changes nature in an isolating language: *dór* in Tetun Dili

Catharina Williams-van Klinken and John Hajek

Dili Institute of Technology / University of Melbourne

In East Timor, there have been centuries of contact between the strongly isolating Austronesian language Tetun Dili and the morphologically-rich Romance language Portuguese. In all this time, only one derivational morpheme has been borrowed into Tetun Dili for use with native lexicon. This is *-dor*, a transparent agentive suffix which neatly fits the word order and stress patterns of existing Tetun Dili agentive compounds. Tetun Dili has borrowed numerous nouns with this suffix. However when in combination with native roots, it has shifted in terms of its semantics, word class of the root and derivation, and even word status, bringing it more in line with pre-existing native agentive morphemes. In other words, Tetun's strongly isolating nature has won, at least for now.

Keywords: language contact, isolating language, Tetun Dili, Portuguese, Austronesian contact, morphological borrowing

1. Introduction[1]

Tetun Dili is the primary lingua franca used throughout the newly independent nation of East Timor, spoken by some 940,000 people, mostly as a second language. It is also one of East Timor's official languages, alongside Portuguese.

It has, over a period of several hundred years, developed from the Austronesian language Tetun Terik, which has limited morphology, including subject marking on some verbs, an agentive circumfix *mak n*, a few other derivational prefixes and suffixes, and partial and full reduplication, in addition to compounding (van Klinken 1999: 58–98). In becoming a lingua franca, Tetun Dili has become largely

---

1. An earlier version of this paper was presented at the Eleventh International Conference on Austronesian Linguistics in Aussois, France, 22–26 June 2009.

isolating in nature. It has lost subject marking and almost all derivational possibilities, retaining only compounding, causative *ha*, a limited use of detransitivising *nak*, and minimal reduplication (Williams-van Klinken, Hajek & Nordlinger 2002: 17–24).

Tetun Dili is unusual for Austronesian languages in that it has been greatly influenced by Portuguese, the language of colonial rule for several centuries until the Indonesian invasion in 1975. This impact is primarily seen in vocabulary, with the proportion of Portuguese loans varying from 10–20% in everyday conversations in the capital Dili, and up to 30–40% in newspaper articles. Amongst open-class words, the percentage of Portuguese loans in the media can rise as high as 75%.

Tetun Dili (henceforth Tetun) thus provides a case study of an isolating Austronesian language in very close contact with a morphologically-rich Romance language, Portuguese. In this study we look at how these two influences play out in the case of the Portuguese suffix *dor*, which is the only Portuguese affix to have been borrowed into Tetun as a productive derivational morpheme used with native and nativised lexicon. We show how there are now two related but different manifestations of *dor* in Tetun. The first is clearly a suffix which is attached only to Portuguese transitive verb roots, with the same features as in Portuguese. The second reveals a clear semantic and grammatical shift, and shows signs of a move away from simple suffixal status.

2.   Historical background

As mentioned above, Tetun Dili developed from the vernacular Austronesian language Tetun Terik. Tetun Terik was the basis of a lingua franca that has been used throughout much of the eastern part of the island for at least the last five hundred years (*documento Sarzedas,* quoted in Nordholt 1971: 161).

Portuguese involvement with Timor started five hundred years ago, with their first shipment of sandalwood in 1515. However, intensive contact between Portuguese and Tetun only started in 1769, when Portugal moved its capital from Lifao in the west to its current location of Dili, located within the region in which Tetun was a lingua franca. By 1845, the tiny capital of Dili was reported to be bilingual in Tetun and Portuguese (*Annaes Maritimos e Coloniaes*, cited in Thomaz 2002: 104).

In addition to standard Portuguese, there was until the 1960s also a creole Portuguese spoken mainly by mixed-race residents in the Dili suburb of Bidau (Baxter 1990; Hull 2002a). This creole is not likely to have been influential in the adoption of the suffix *dor* into Tetun Dili, since the creole was not widespread, and this suffix is not mentioned in Baxter's (1996) description of it.

The development of modern Tetun Dili, as a heavily Portuguese-influenced variety of Tetun, dates to after World War II (Thomaz 1981). This came about largely through the rise of a new educated elite in Dili, and a decline in influence of the traditional elite, who had closer ties to Tetun Terik.

During the period of Indonesian rule (1975 to 1999) the influence of Portuguese reduced, as Indonesian was the official language, and the use of Portuguese in public was actively discouraged by the regime.

Since 1999, when East Timor voted to become independent, Tetun and Portuguese have become the official languages. Throughout this period, Tetun has been used in a much wider range of spheres than before, including the media, parliament and education, with most new terms required for these fields being taken from Portuguese.

Not only has Tetun gained new functions since independence, but it has also gained many speakers, with the proportion of the population who speak Tetun Dili at home rising markedly. In 1999 only about 7% of the population spoke Tetun Dili as a home language.[2] According to the 2015 census, less than twenty years later, 360,000 people, or 31% of the population, speak Tetun Dili as their main language at home. A further 580,000, or 57% of the population, speak it as a second or third language.

3. Portuguese nominal loans with *dór*

Tetun has borrowed numerous Portuguese nouns ending in the suffix *dor*.[3] These are mostly terms found in formal registers, particularly in bureaucratic and technical fields. There are 180 examples in the lexical database on which this study is based, with potential borrowings being almost limitless. In many cases, the Portuguese root is borrowed as well. In all cases noted so far, the root is, at least in its original Portuguese form, a transitive verb.

The loans fall into two semantic categories. The first are nouns referring to humans, identifying their profession or role, as illustrated in Table 1. The root verb specifies their typical activity. (Examples are given in Tetun spelling.)

---

2. This percentage was calculated by Williams-van Klinken, Williams and Brites da Silva (2016) based on the number of speakers listed in Grimes (2005) and population figures given in National Statistics Directorate (2006: 28).

3. In Portuguese orthography the suffix *dor* does not have an accent mark. However, in official Tetun orthography an accent is used to indicate stress on final or antepenultimate syllables. Hence Portuguese writes *administrador*, but Tetun *administradór* 'administrator.MASC.SG', whereas feminine *administradora* is spelled identically in both languages.

Table 1. Portuguese loans specifying human roles

Derivation	Root
administradór 'administrator'	Vt. *administra* 'administer' (rarely used in Tetun)
governadór 'governor'	Vt. *governa* 'govern' (rare in Tetun; it prefers *ukun*)
fasilitadór 'facilitator'	Vt. *fasilita* 'facilitate'
moderadór 'chairperson'	Vt. *modera* 'chair'
peskizadór 'researcher'	N. *peskiza* 'research' (Vt or N in Portuguese, N in Tetun)

The other category is that of instruments, with the root verb specifing what the instrument is typically used for. Table 2 presents some examples.

Table 2. Portuguese loans specifying instrument

Derivation	Root
gravadór '(tape…) recorder'	Vt. *grava* 'record'
karegadór '(battery) charger'	Vt. *karega* 'charge'
regadór 'sprayer'	Vt. *rega* 'spray'
radiadór 'radiator'	(Portuguese Vt. *radiar* 'radiate' not known in Tetun)
agrafadór 'stapler'	Vt. *agrafa* 'staple'

Some of these derivations are also used as adjectives in Portuguese, for instance *fundador(a)* 'N. founder; Adj. founding'; *colonizador(a)* 'N. coloniser, Adj. colonising'. However, such adjectival use is extremely rare in Tetun, apparently being restricted to speakers who are exceptionally influenced by Portuguese.

In Portuguese, nouns and adjectives are marked for both gender and number, whereas in Tetun native lexicon, they are marked for neither. The conflict between these two systems with respect to Portuguese loans in Tetun is resolved in different ways depending on the specific word, the formality of the situation and the degree of Portuguese influence on the speaker (Hajek & Williams-van Klinken 2019). In formal situations, where Portuguese influence is stronger, some speakers use both feminine and plural endings in loans, while others feel uncomfortable with them. For instance, television reporters regularly address their viewers as *telespetador-es sira* 'far-watch-AGENT-PLUR DEF.PLUR' = '(TV) viewers', which most young people consider rather unacceptable, since it combines a Portuguese plural *es* and a Tetun plural morpheme *sira*. Similarly, in meetings one will hear *modera-dor-a* 'moderate-AGENT-FEM' to refer to female moderators; those less influenced by formal meeting terminology are likely to use *moderadór* neutrally to refer to either men or women. Many people reject feminine endings on Portuguese loans unless they are used to hearing them. So, for instance, ten years ago many people rejected feminine *embaixadora* 'ambassador-FEM' or *kantador-a* 'singer-FEM', since at the time these forms were rarely heard. Now they are becoming more widely used and accepted.

4. *Door* with single Tetun roots

The Portuguese suffix *dor* has also been productively borrowed into Tetun. In Tetun, however, it behaves in strikingly different ways from the original Portuguese source form. This includes the semantics of the derived term, the class of the root, the adjectival status of the derivation (discussed in Section 8), and its word status (discussed in Section 9, but foreshadowed here by its spelling as a separate word *door*).

Semantically, *door* derives only human nouns. In most cases, the root describes a habitual activity, which is usually negatively valued, as can be seen from the examples in Table 3.

Table 3. Derivations specifying negatively valued characteristics

Derivation	Root
husu door 'person who keeps asking for a loan or favour'	Vt. *husu* 'ask'
mama door 'habitual betel nut chewer'	Vt. *mama* 'chew betel nut'
haluha door 'forgetful'	Vt. *haluha* 'forget'
lao door 'gadabout'	Vi. *lao* 'walk'
moe door 'very shy'	Adj. *moe* 'shy'
dukur door 'sleepyhead'	Adj. *dukur* 'sleepy'
nervozu door 'quick-tempered, irritable'	Adj. *nervozu* 'irritated' (Portuguese *nervoso* Adj. 'nervous; touchy; worked up; irritated')

There are, however, also a few derivations which indicate a person's position or skill, and at least one (*servisu door* 'diligent') which is about habitual behaviour that is positive. These are illustrated in Table 4.

Table 4. Derivations specifying position, skill or positively-valued characteristics

Derivation	Root
siik door 'fortune-teller'	Vt. *siik* 'guess'
tiha door 'net fisherman'	Vt. *tiha* 'fish with a net'
servisu door 'diligent, hard-working'	Vi. *servisu* 'work' (Portuguese *serviço* N. 'service, work')
tiru door 'sharpshooter'	Vt. tiru 'shoot' (Portuguese *tiro* N. 'shot, shooting')

As shown by the examples given above, the root in Tetun derivations is often a transitive verb (e.g. *husu* 'ask'), but unlike Portuguese, it can also be an intransitive verb (e.g. *lao* 'walk') or an adjective (e.g. *moe* 'shy').

Some Tetun derivations with *door* use Portuguese roots, but produce new forms which are not possible in Portuguese since in Portuguese the roots in question are not transitive verbs. This is the case for the above examples *nervozu door*

'quick-tempered, irritable', *servisu door* 'diligent' and *tiru door* 'sharpshooter' (for which Portuguese uses *atirador* derived from the verb *atirar* 'shoot').

   Some derivations based on nativised Portuguese roots are used in two contrasting ways, the one based on Tetun semantics, and the other on Portuguese semantics. Some such contrasts are shown in Table 5.

Table 5. Contrasts between Tetun and Portuguese derivations

Tetun derivation	Portuguese loan	Root
empresta door 'someone who always borrows and doesn't return'	*emprestadór* 'borrower, lender' (term used in translations but not widely accepted)	*empresta* 'borrow'
joga door 'habitual gambler, card-player'	*jogadór* 'football player'	*joga* 'play, gamble'
konsumi door 'person who loves to eat a particular thing (e.g. cassava)'	*konsumidór* 'consumer (e.g. of electricity)'	*konsumi* 'consume'

Unlike true Portuguese loans ending in *dor*, Tetun derivations using *door* cannot take a Portuguese plural or feminine suffix, even when the form involves a loaned Portuguese root. It is thus possible to get a contrast between a Portuguese loan noun which can take a feminine suffix and indicates someone's position, and a Tetun derivation which uses the same root, but which can never take a feminine suffix and which indicates behaviour (in the case below, a once-off role).[4] In this way a subtle but important semantic difference is clearly maintained.

(1) *Joana nee* <u>*kordena-dor-a*</u>   *língua.*
     Joana this coordinate-AGENT-FEM language
     'Joana is the language coordinator.' [official position – Portuguese-influenced interpretation]

(2) *Joana nee* <u>*kordena-dór*</u>   *ba língua   nian.*
     Joana this coordinate-AGENT for language POS
     'Joana is coordinating language issues.' [what she is doing, for instance, in the current meeting – Tetun-influenced interpretation]

---

4. Thanks to Tetun teacher Hendriana da Costa Marçal who managed to analyse and point this out when one of the authors attempted to 'correct' her *kordenadór* to feminine *kordenadora* in a text she was writing.

5. Three-unit sequences

Although *door* usually combines with a single verbal or adjectival root, it can also combine with a sequence of a transitive verb and its object, again usually for negatively valued behaviour. In this case *door* either precedes or follows the object.

In some derivations, such as those illustrated in Table 6, speakers prefer the object to be final. This is always the case in those (rare) instances in which the object consists of more than one word (such as the final example *konta door ema nia vida* below). Limited evidence suggests this is also the strongly preferred order if the object is a single word consisting of more than two syllables (such as *sasaan* 'goods').

Table 6. Derivations with transitive verb, *door*, and object

Derivation	Literal translation
baku door feen 'wife basher'	'bash AGENT wife'
haluha door sasaan 'person who keeps forgetting their possessions'	'forget AGENT goods'
empresta door sasaan 'person who keeps borrowing things and doesn't return them'	'borrow AGENT goods'
konta door ema nia vida 'gossip'	'person AGENT person POS life'

This order follows that of Portuguese, which uses a genitive *de* between the agentive noun and the object (e.g. *cobrador de impostos* 'collector of taxes'). Although this genitive marker is not carried over into Tetun, many of the expressions are, such as *kobradór impostu* 'collect-AGENT tax' = 'tax collector', and *administradór munisípiu Dili* 'administer-AGENT municipality Dili' = 'Dili municipal administrator'.

In contrast, final *door* is preferred when the verb-object pair cannot be separated, so forcing *door* into final position. This order is required when reciprocal *malu* 'each other' is used; this is consistent with the fact that *malu* cannot be separated from the verb in any construction. It also seems that idiomatic expressions, such as *lori lia* 'carry word' = 'spread malicious rumours' or *futu manu* 'tie bird' = 'do cock-fighting' cannot readily be separated by *door* (e.g. only 2 out of 5 assistants accepted the alternative *futu door manu*). This order, illustrated in Table 7, is uniquely Tetun, being not in any way possible in Portuguese with *dor*.

Table 7. Derivations with transitive verb, object and *door*

Derivation	Literal translation
baku malu door 'person who habitually gets involved in fights'	'bash RECIPROCAL AGENT'
istori malu door 'quarreller'	'quarrel RECIPROCAL AGENT'
tolok ema door 'person with a dirty mouth'	'swear.at person AGENT'
lori lia door 'rumour-monger, tattle-tale'	'carry word AGENT'
hemu tua door 'drunkard, wino'	'drink wine AGENT'
futu manu door 'cock-fighter, gambler on cock-fights'	'tie bird AGENT'

For yet other expressions, there is variation, either because speakers readily accept both orders, or because they disagree as to the appropriate order. This is the case for many (but not all) expressions in which the object is the generic *ema* 'person'. Table 8 presents examples of this variable constituent order.

Table 8. Derivations with transitive verb, and *door* either preceding or following the object

Derivation	Literal translation
konta door ema = *konta ema door* 'gossip'	'recount person' + 'AGENT'
haan door ema = *haan ema door* 'person who uses black magic'	'eat person' + 'AGENT'
baku door ema = *baku ema door* 'person who often bashes people'	'bash person' + 'AGENT'

Some of these verbs appearing in three-unit sequences can also readily have *door* without an object (e.g. *haluha door* 'forget AGENT' = 'forgetful person', alongside *haluha door sasaan* previously cited above). Others, though, require an object. For instance *hafuhu door* 'spy.on AGENT' requires an object to specify what the person is spying on, such as *hafuhu door festa* 'spy.on AGENT party' = 'person who often spies on parties' or *hafuhu door ema* 'spy.on AGENT person' = 'peeping Tom'.

The objects in these constructions are non-referential, and nearly always consist of a single word. They are usually either very generic, such as *ema* 'person', or chosen from only a small set of possible candidates for that verb, such as *tua* 'wine' or *kafé* 'coffee' for the verb *hemu* 'drink' (e.g. *hemu tua door* 'drink wine AGENT' = 'drunkard, wino' or *hemu kafé door* 'drink coffee AGENT' = 'coffee addict'). As with two-word derivations, they usually involve negatively valued behaviour. So, while *door* can be used for wine or coffee drinkers (both of which substances are considered bad if over-indulged in), attempts to use *hemu xá door* 'drink tea AGENT' to mean 'old soak, heavy tea drinker' elicit amusement, with comments such as 'But drinking tea isn't bad for you!' or 'Tea isn't addictive!'

With regard to constituent structure, three-word sequences with medial *door* clearly have *door* more closely tied to the verb than to the following object (i.e.

the structure is [Vt *door*] O). In this Tetun follows, as previously noted, the source Portuguese construction, where the *dor*-derivation and the object are separated by the genitive preposition *de*. Another indicator that the object is separate is that writers are all agreed that it should be written as a separate word, even though they do not agree on whether *door* should be written attached to the verb. Finally, objects in final position sometimes consist of more than one word, although this is rare (e.g. *konta door ema nia vida* 'person AGENT person POS life' = 'gossip').

Where *door* occurs finally, the constituent structure is less clear. Semantically, *door* belongs with the verb or the verb-object combination, not with the object alone (hence, [Vt O] *door*). In terms of spelling, most of our assistants considered that these should be written as three separate words (e.g. *hemu tua door*), as occurs with the one example under the *dor* entry in Morris' (1984) dictionary. The remaining assistants made a closer written link between the *door* and the object than between the verb and *door* (e.g. *hemu tua-door, hemu-tuador*), contrary to what one would expect from the semantic structure. This spelling, however, presumably reflects people's recognition that *dor* is normally written attached to what precedes it (based on it being a suffix in Portuguese), rather than reflecting their conception of constituent structure. These expressions are in any case so rarely written that no consensus has yet been achieved. In contrast to our assistants, Hull and Guterres Correia (2005: 7) spell their one example as a single word (*bukaliadór* 'seek-word-AGENT' = 'one who stirs up trouble'), consistent with their analysis of *door* as a suffix, while Thomaz (1981: 117) spells his one example with hyphens (*futu-manu-dor* 'tie-bird- AGENT' = 'cockfighter, gambler on cockfights').

6. Use and creativity

While Portuguese loans with the suffix *dor* occur predominantly in more formal and technical situations, Tetun derivations with *door* are mainly found in informal contexts. Perhaps their usually pejorative connotations contribute to this. In fact, in a corpus of 400,000 words of written texts, there were only 20 examples of Tetun *door* expressions, with these being found in less formal writing, such as critical postings to an email list, and a common-language Bible translation.

These Tetun derivations are, however, quite productive, as well as rather unstable. The lexical database on which this study was conducted included 120 examples with Tetun roots, with speakers readily able to come up with more. One student, when asked for three-word examples, came up with almost 50 expressions, most of which were subsequently rejected by other students (and so not included in the example count above). In fact, many *door* expressions which are used by some people are rejected by others. In some cases, they are even rejected by the speakers

themselves. For instance, young people who were explaining *hafuhu door* 'spy.on AGENT' to one of the authors, listed as examples *hafuhu door festa* 'spy.on AGENT party', *hafuhu door ema hariis* 'spy.on AGENT person bathe', and *hafuhu door ema troka* 'spy.on AGENT person change.clothes', but then denied that the latter two would use *door*. This variability is to some degree a characteristic of the language; most of its speakers are not native speakers of Tetun Dili, and there are many other terms too on which there is no consensus. The lack of consensus seems particularly extreme for three-word *door* expressions, perhaps also because they are not reinforced by being seen in writing, and do not have parallels in Portuguese or in Indonesian.

Derivations using *door* can communicate effectively as one-off creations, even when speakers claim the resultant expression is 'not Tetun'. For instance, during the 2006 military-political crisis in Timor, a senior opposition figure speaking on the radio listed some terrible crimes allegedly committed by his opponents, using the presumably innovative expression *fahe door kilat sira nee* 'distribute AGENT gun DEF.PLUR this' = 'these people who hand out guns (to unauthorised recipients)'.

The corpus even includes one innovative *door* expression based on an Indonesian root, a very well-known nominal acronym *KKN*, meaning 'collusion, corruption and nepotism'. In this opinion blog, a group of leaders was described as:

(3) *Nauk-ten,   bandidu, otonomista   Kkn   dor   mak*
     steal-dung bandit   automomy-supporter corruption AGENT FOCUS
     sira   ne'e.
     DEF.PL this
     'Thieves, bandits, through-and-through corrupt supporters of autonomy [as opposed to independence of East Timor], that's what they are.'[5]

There are several characteristics which clearly indicate that *door* derivations are lexemes rather than syntactic phrases. Firstly, as noted above, there is a degree of lexicalisation, in that speakers have opinions about which sequences are 'Tetun', though with large inter-speaker variation and room for creativity. Secondly, in contrast to phrases, the verb cannot be individually modified; for instance, one cannot say **moe loos door* 'shame very AGENT'. Thirdly, in three-word sequences, the object of the verb is necessarily non-referential, as is typical for derivations but not for phrases. Finally, it is argued below that these expressions are adjectival. Since all phrases in Tetun are endocentric (e.g. all adjective phrases are headed by an adjective), there is no precedent for having an adjectival phrase which does not even include an adjective as one of its constituents.[6]

---

5. Note that examples from written texts are in the original spelling.

6. In contrast, there is one class of compounds which has a different word class to any of the constituent words, namely adverbial compounds from coordinate verbs (e.g. adverbial *tuun-sae* 'up and down, all over the place' from the two verbs *tuun* 'descend' and *sae* 'ascend')

7. Alternative strategies for agentive terms

Tetun Terik (the Austronesian source language for Tetun Dili) has relatively little productive morphology, but does have a productive agentive circumfix *mak n* (with variants *mak, ma k,* and *ma*), which derives actors (van Klinken 1999: 70–76). Examples include *ma-kawen* 'AGENT-marry' = '(one) who married', *mak-leo-n* 'AGENT-protect' = 'protector, guardian', and *mak-sosa-n* 'AGENT-sell' = 'seller'. Like Tetun Dili *door* derivations, derivations with *mak n* can take an object, for instance *ma-ho kabau malae* 'AGENT-have buffalo non-native' = 'who have horses'. Also like *door*, the resulting derivation is not a noun; rather it is analysed as a special sub-class of verbs with some adjective-like properties. Unlike *door*, however, these do not describe habitual behaviour, and do not have negative connotations.

Tetun Dili has lost this circumfix, except in a few fixed terms used mostly in the conservative liturgical register, notably *mak-soi-n* 'AGENT-save' = 'Saviour', and *mak-sala-k* 'AGENT-sin' = 'sinner'. This loss follows most of the surrounding languages, which are even more isolating than Tetun Terik, and which use periphrastic constructions for agentive terms (Hull 2001: 107). These include Mambae, the language with the largest number of speakers and the original language of the area where the capital Dili is now located. Such influence is to be expected since the vast majority of Tetun Dili speakers have, throughout its history, been native speakers of other Timorese languages.

In place of this Tetun Terik circumfix, Tetun Dili uses three morphemes which all follow the verbal root to derive actors. By far the most productive of these morphemes is *door*. The other two are native Tetun nouns, but they are used to form compounds not found in Tetun Terik. Unlike *door* and *mak- -k*, neither of them is used in undisputed three-morpheme constructions.

The first of these native forms is *teen*, meaning 'dung, excreta'. Not surprisingly, this is used for negatively valued characteristics (with about 20 examples in the database), being more strongly pejorative than *door*. Examples include *bosok-teen* 'lie-dung' = 'habitual liar' and *baruk-teen* 'lazy-dung' = 'lazybones'. Hull (2001: 107) suggests that the *teen* derivations could have derived from a proto form **teras* meaning 'hard' (the cognates of which are still used in some other Timorese languages such as Waima'a to form agentive terms). He suggests that this developed into **tees* in Tetun, which was associated by folk etymology with *teen* 'excrement', with the result that it merged formally and semantically with *teen*. This suggestion seems the most plausible to date, since other local languages do not use 'dung' to create pejorative terms. Almost all *teen* derivations are adjective-like in that they are normally used predicatively or attributively, and can readily be intensified (e.g. *O beik-teen liu!* '2s.INFORMAL stupid-dung very' = 'You're such an idiot!'). However, unlike many other adjectives, they cannot be negated using the nonemphatic negator *la*.

The other native root used in actor compounds is *nain*, a noun meaning 'master, lord, owner'. This was traditionally used with verbs or adjectives to derive a handful of agentive nouns. All are positive, such as *kaben nain* 'marry master' = 'married person', and *matenek nain* 'clever master' = 'expert', with the exception of the church term *sala nain* 'wrong master' = 'sinner'. In recent years *nain* has started to be used in formal contexts such as the media and conferences for new derivations showing role. These derivations, which are not at this stage well accepted, include *lee nain* 'read master' = 'reader' (replacing the Portuguese loan *leitór*), *hakerek nain* 'write master' = 'writer' (replacing the rarely used Portuguese loan *eskritór* 'writer'), and *rona nain* 'listener' (replacing Portuguese *ouvinte*). These expressions are nominal. Although they rarely refer to particular individuals (with the exception of *matenek nain* 'expert'), they frequently occur as the head of definite plural NPs (e.g. *ukun nain sira* 'rule master DEF.PLUR' = 'the rulers'). The use of a term meaning 'master' to derive agentive nouns follows other languages of the region (Hull 2001: 107), e.g. Waima'a *bale-buu* 'steal-master' = 'thief'. Note that *nain* is also used with nominal roots in the more direct sense of 'owner, master'; e.g. *loja nain* 'shop master' = 'shop keeper', *rai nain* 'land master' = 'spirit of the land; indigenous person', just as it is in Tetun Terik and other languages of the region, e.g. Waima'a *busa-buu* 'field-master' = 'farmer'. This construction has in recent years been extended to include mastery of characteristics, in expressions such as *dame nain* 'peace master' = 'Lord of Peace' (e.g. describing God) and *justisa nain* 'justice master' = 'judge'.

Since *teen* and *door* are both pejorative, and *nain* rather unproductive, the lexical gap left in Tetun Dili by the loss of Tetun Terik *mak k* has not yet been adequately filled. Portuguese loans make up some of the short-fall, particularly in higher registers, e.g. *vendedór* 'seller', while in formal contexts some people are extending *nain* to role terms such as *sosa nain* 'sell master' = 'seller'. There remain, however, many actor concepts for which Tetun Terik derives terms with *mak*, but for which Tetun Dili can only use phrases, such as 'picker' (Tetun Terik *ma-hili-k* 'AGENT-pick', Tetun Dili *ema nebee hili* 'person REL pick'), and 'cutter' (e.g. Tetun Terik *ma-kotu husar* 'AGENT-sever umbilical.cord' = 'person who cuts the umbilical cord', Tetun Dili *ema nebee kotu husar* 'person REL cut umbilical.cord').

8. Word class

Although the Portuguese suffix *dor* always derives a noun in the first instance, in Tetun, *door* derives adjectives, thus following both the pejorative agentive *teen* (but not *nain*) and Tetun Terik's agentive circumfix *mak k*.[7] The evidence for this is as follows.

Firstly, these expressions are usually predicative or attributive, describing a person rather than referring to one.

(4) *O hatene nia hirus door,   keta provoka nia hirus nee!*
    2s know 3s   anger AGENT don't provoke 3s   anger this
    If you know (your spouse) is quick to anger, don't stir up their anger!

(5) *Ameu, … Hau la   gosta besik ba ema   fuma door.*
    (name)   1s   not like   close to person smoke AGENT
    'Ameu, (just sit further away). I don't like being close to smokers.'

Secondly, it is possible to intensify these expressions. This is however rare, perhaps because *door* expressions are already strong.

(6) *Heis, o   para kesar   door!*
    EXCL 2s so   tell.on AGENT
    Heh, you're such a telltale!

Thirdly, there is the test of negation, which is again rare, with only two negated examples with the negator *la* in our textual database. When speakers were asked for their judgments, there was significant variation. However most accepted negation using unemphatic *la* 'not' or the potentially more emphatic *la … ida* lit. 'not … one', used with verbs and adjectives, rather than insisting on *laos* 'not', which is the only way to negate nouns.

(7) *Hasee ema   diak, kalma, mos bele foo perdua   ba ema   bainhira*
    greet   person good calm   also can give forgiveness to person when
    ema   halo sala   ba nia, la   nervozu door.
    person do   wrong to 3s   not irritated AGENT
    (Describing someone who is *oin mamar* (face soft)): '(S/he) greets people well, is calm, can also forgive people when they do wrong to him/her, is not irritable.'

---

7. The authors in their previous descriptions of Tetun *door* derivations (Hajek & Williams-van Klinken (2003: 58), Williams-van Klinken, Hajek & Nordlinger 2002: 20) assumed, with insufficient justification, that these were nouns. Hull and Guterres Correia (2005: 7) list *dor* as a nominal suffix, while Hull and Eccles (2001: 7) say it derives agentive nouns or adjectives. Hull's dictionary (2002b) lists some derivations as nouns (e.g. "*laodór* n. good walker"), some as adjectives (e.g. "*hamnasadór* adj. laughing; smiling; jovial, cheerful", and some as both (e.g. "*halimardór* adj. playful; n. joker").

Finally, *door* expressions fail the critical tests for nounhood in Tetun Dili. They cannot be enumerated, and cannot head a possessor noun phrase. Nor, as previously stated, can they be used to refer directly to people. For instance, when one of the authors suggested using **Baa bolu kanta door sira mai* 'go call sing AGENT come' to call singers onto the stage, the assistants protested that this would be interpreted as making fun of the singers. In other words, it would be a pejorative description of them rather than a term of reference, for which Portuguese *kantór* 'singer' is used.

A summary comparison with *nain* and *teen* is presented in Table 9 below.

Table 9. Grammatical and semantic characterisation of agentive derivations in Tetun

	door	*teen**	*nain*
Source of morpheme	Portuguese suffix *dor*	Noun 'dung'	Noun 'master, lord, owner'
Usual syntactic function	Attributive or predicative	Attributive or predicative	Some (e.g. *kaben nain* 'married person'): Attributive or predicative Others (e.g. *matenek nain* 'expert'): head NP
Can be intensified?	Yes (but rare)	Yes (easily)	No
Can be negated with verbal/adjectival negator *la*?	Yes (but rare)	No (speakers dubious)	No
Can be negated emphatically with *la … ida*	Yes	Yes	No
Can be used to refer?	No	No	Yes, but only to groups, not to individuals (except *matenek nain* 'expert')
Can be enumerated?	No	No	No
Can head possessor NP?	No	No	Yes
Has pejorative sense?	Normal (not very strong)	Always (strong)	Never

* The only exception is *naok-teen* 'thief', which is clearly a noun, being used to refer, countable, and able to head a possessor NP. It is not clear why this should be the case for this word only.

Table 9 demonstrates that forms with *door* and *teen* are strikingly similar in most respects, and are grammatically much more adjectival in nature than forms with *nain*, which are clearly nominal in behaviour. The greater dispreference for negation of *teen* forms with nonemphatic *la* appears to reflect the greater pejorative and emphatic sense of such forms in comparison to those with *door*.

9.   Word status

Writers on Tetun, as well as writers in Tetun, disagree as to whether *door* with Tetun roots is a suffix, a separate word, or somewhere in between. For instance, Thomaz (1981: 66f) says that it is a separate word though writing it with a hyphen, but experts involved with the National Institute of Linguistics consider it a suffix (Hull & Eccles 2001: 7, Hull & Guterres Correia 2005).

Amongst Tetun speakers and writers, confusion is to be expected. On the one hand, the existence of many Portuguese loans in which *dor* is clearly a suffix must push speakers in favour of analysing *door* on Tetun roots too as a suffix, to be spelled attached to the root and with a single 'o' as in Portuguese. On the other hand, the fact that grammatical behaviour, word-position and stress-bearing pattern parallel those of the native Tetun roots *nain* and especially *teen* favour analysis as a separate root, potentially spelled with double 'o' to show that the vowel is long. These competing analyses are not helped by the fact that Tetun Dili is a second language for most of its speakers, and is in any case written anarchically in practice, with the result that speakers are often unsure in their judgments about what is correct Tetun (see also below).

Phonological behaviour is not a particularly useful diagnostic for *door*, while spelling practice is conflicting (see immediately below). Nevertheless, its productivity and its flexibility of position in three-word expressions indicate that *door* has moved along the cline from suffixal (as in Portuguese) in the direction of a separate word, much like *nain* and *teen*.

Phonology

Phonologically, there is no apparent difference between *door* expressions and true Portuguese loans with suffixal *dor*: both appear to be single prosodic units, with destressing of the preceding root; e.g. [joga'do:r] *jogadór= joga door*. However, similar destressing of the left-most element also occurs with the other two agentive morphemes in Tetun, *teen* and *nain*. All three agentive morphemes thus behave quite unlike any native Tetun affixes, since Tetun Terik has only unstressed prefixes and consonantal suffixes. There are therefore no other unambiguously stressed suffixes with which a comparison can be made.

Spelling

Spelling practices present evidence that Tetun *door* is perceived by many Tetun writers not as an affix, but rather as a separate root. As already noted, Tetun spelling is, in practice, still unstandardised, varying greatly from one writer to the next. Nevertheless, certain patterns are present.

With respect to Portuguese loans using *dor*, all writers are agreed in writing them as single words, with a single 'o' in *dor*. The only variation is in whether to follow official National Institute of Linguistics spelling (Instituto Nacional de Linguística 2002) in placing an accent over the 'o' (e.g. *administradór* 'administrator'), or whether, like most writers, they omit the accent (*administrador*) as in Portuguese.

In contrast, spelling of Tetun derivations varies enormously. Most people who assisted with this research, as well as most examples in our written corpus, write two-morpheme Tetun *door* expressions as two words (e.g. *siik door* or *siik dor*), while some use a hyphen (e.g. *siik-door* or *siik-dor*). In both cases, some people use a double vowel in *door* to show its length, while some others use a single vowel as per Portuguese, with many writers in any case not consistently marking length in Tetun. In addition, there are also people who write these derivations as one word with a single vowel in *dor* (e.g. *siikdor*), as per Portuguese loans and in line with the officially sanctioned National Institute of Linguistics spelling (except that it places an accent on the 'o', e.g. *siikdór*).

Some of those who assisted with this research distinguished so clearly between Tetun and Portuguese derivations that they would spell 'player' as two-word *joga door* when they interpreted it according to Tetun semantics as 'gambler', and as single-word *jogadór* when they interpreted it in the Portuguese sense of 'football player'.

It should, however, be pointed out that variation in spelling word boundaries is also found in writing compounds with *teen* and *nain*. As shown in Table 10 below, for *teen* compounds, the National Institute of Linguistics and the present authors use a hyphen (e.g. *baruk-teen*); other texts not influenced by these two sources are evenly split between writing *teen* as a separate word (e.g. *baruk teen*), or writing it directly attached to the root (usually with a single vowel, e.g. *barukten*). In contrast, *nain* compounds are usually written as two separate words, except by authors who use the National Institute of Linguistics (INL in Portuguese) spelling system, in which all compounds are written with hyphens.

It might be argued that some people prefer to write *door* as a separate word because of a reluctance to attach an obviously Portuguese morpheme *dor* to a Tetun root. This seems unlikely, however, since there are a few derivations which mix the Tetun causative prefix *ha* with a clearly Portuguese root, and which are relatively widely

Table 10. Spelling patterns in Tetun

	door	*teen*	*nain*
Two words	most common	36% (of these 70% *teen*, 30% *ten*)	75%
Hyphen	some	33% (almost all based on Dili Institute of Technology or INL spelling)	25% (all based on INL spelling)
Single word	some	31% (of these 90% *ten*)	rare

accepted and invariably written as a single word, namely *hapara* 'make-stop' = 'Vt. stop', with 81 textual examples, and *haforsa* 'make-strong' = 'strengthen', with 14 examples in our corpus (and thousands more on the internet). If people were reluctant to mix languages within the one word, it would be easy to replace these derivations with the periphrastic phrases *halo para* 'make stop' or *halo forsa* 'make strong', both of which also occur in texts.

A lexeme *door*?

There is one example in the corpus, from a slanderous flame on an email list, which uses *door* apart from a root. The reduplication of *door* in this example (presented below in the original spelling) indicates not only that the author has more than one unacceptable behaviour in mind, but also that they are of various kinds.

(8) *Uluk   ne'e pasti   mau-hu   ida, …, lanu-ten   ida,*
   formerly this surely older.brother-blow one   drunk-dung one
   'In the past he must have been an informer, …, a drunkard,
   futu manu dor,   joga feto   dor,   deve-dor
   tie   bird   AGENT play woman AGENT owe-AGENT
   a cock-fighter, one who goes to prostitutes, always in debt,
   no   dor-dor   seluk Tan   karik.
   and AGENT-AGENT other as.well perhaps
   and probably various other bad things as well.'

Aikhenvald (2007: 51) notes that in rare cases, derivational affixes can be used independently. In this case, they are usually used in the plural, such as in the English book title "Isms and Ologies", and in the *dor-dor* example here.

10. Borrowability of *door*

Several factors facilitate the borrowing of Portuguese *dor*, even though suffixes are not readily adopted from other languages for use with native lexicon (Thomason & Kaufman 1988: 74).

One is the intense long-term exposure of Tetun to Portuguese, both during the long period of Portuguese rule, and as co-official language since Timor achieved full independence in 2002. It thus fits Thomason and Kaufman's (1988: 74) observation that derivational affixes may be borrowed and added to native roots in situations of relatively intense language contact.

Another contributing factor is that *dor* occurs, as already noted, in a large number of Portuguese loans.

In addition, *dor* is a very 'transparent' morpheme. In particular, it is clearly delineated phonologically, in that it always has the form *dor*, even though this can in Portuguese be followed by feminine *a* or plural *es/ as*. The suffix is also regular semantically, in that it always derives either an actor or instrument noun. As such it is easier to borrow than less transparent affixes (Winford 2003).

Given the loss of Tetun Terik agentive *mak k*, there was a lexical gap for agentive terms. Portuguese *dor* (and *door*) has the marked advantage of being a less pejorative way of producing such terms than the native alternative *teen* 'dung, shit'. On the other hand, it also has a wider semantic range than native *nain* 'owner, master, lord'. The frequently pejorative sense of *door* with native roots is balanced by the neutral nature of *-dor* typical of Portuguese loans (e.g. *vendedór* 'seller', *administradór* 'administrator').

That Portuguese *dor* is easy to borrow is shown by the fact that it has spread further to other vernacular languages in East Timor, including both Austronesian languages such as Waima'a (e.g. *maudór* 'drunkard' (Hull 2001: 108)), and non-Austronesian ones such as Makasae (e.g. *logo dor* 'liar' from *logo* 'lie' (Francisca Cecilia X. dos Santos p.c. 2018)). Outside of Timor, too, this agentive suffix has been productively borrowed from Spanish into the Austronesian language Chamorro and the South American language Quechua (Chamoreau 2012).

11. The double life of Portuguese *dor* in Tetun

That we now have two manifestations of the Portuguese suffix *dor* in Tetun is not in doubt. Despite their many shared features, suffixal *dor* and more word-like *door* are also very different in behaviour and categorisation. While truly suffixal *dor* is found only on loans, nativised *door* can be attached to native roots, but also, as we have seen, to Portuguese roots. In Table 11 we summarise the grammatical differences between the two forms.

Table 11. A grammatical comparison of *dor* and *door* in Tetun

	Loan: e.g. *kordenadór*	Tetun derivation: e.g. *kordena door*
Word class	Noun	Adjective
Adjectival properties:		
Can be intensified?	No	Yes
Can be negated with verbal/adjectival negator *la*?	No	Yes
Can be negated emphatically with *la … ida*?	No	Yes
Nominal properties:		
Can be used to refer?	Yes	No
Can be enumerated?	Yes	No
Can head possessor NP?	Yes	No
Portuguese grammatical properties:		
Takes Portuguese gender marking?	Yes in acrolect	No
Takes Portuguese plural marking?	Yes in acrolect	No
Type of Root	Vt	Vt, Vi, Adj, (one example of N)
Possible Vt *dor/door* Object order?	Yes	Yes
Possible Vt Object *dor/door* order?	No	Yes
Elements separable in spelling?	No	Yes

A comparison of Table 11 with Table 9 demonstrates the marked shift of *door* away from its use as a nominalising suffix to an adjectivalising morpheme that strongly matches the full lexical element *teen* in its grammatical properties and behaviour in derived agentive forms (as seen previously in Table 6).

12. Conclusion

In conclusion, how do speakers of an isolating language respond to affixation possibilities when borrowing heavily from a morphologically-rich language? In the case of Tetun Dili, they have, despite centuries of intense contact, only adopted for use with native lexicon a single transparent suffix which neatly fits the word order and stress patterns of existing Tetun agentive compounds, while filling a gap in the lexicon. They have reanalysed the suffix at least partially in the direction of a bound root if not lexical element within a compound.[8]

---

8.   See Janda (1995) for other cross-linguistic examples of borrowed suffixes shifting along the grammatical cline to clitics and bound roots.

Not only this, but they are able to successfully distinguish between Portuguese *dor* and nativised Tetun *door*, with the latter taking on Tetun semantics, word class, and even word status in a manner consistent with pre-existing native agentive morphemes *teen* and *nain*.

In other words, Tetun's strongly isolating nature has won. Or rather, it has won so far. With the renewed intensive contact with Portuguese of the last decade, it is very possible that *-dor* will be "re-borrowed", or, at least, again take on more of its Portuguese features – both in loans and in native innovations. Time will tell.

References

Aikhenvald, Alexandra Y. 2007. Typological distinctions in word-formation. In *Language Typology and Syntactic Description, Vol. III: Grammatical Categories and the Lexicon*, Timothy Shopen (ed.), 1–65. Cambridge: CUP. https://doi.org/10.1017/CBO9780511618437.001

Baxter, Alan N. 1990. Notes on the Creole Portuguese of Bidau. *Journal of Pidgin and Creole Languages* 5(1): 1–38. https://doi.org/10.1075/jpcl.5.1.02bax

Baxter, Alan N. 1996. Portuguese and Creole Portuguese in the Pacific and Western Pacific rim. In *Atlas of Language of Intercultural Communication in the Pacific, Asia, and the Americas*, Stephen A. Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds), 299–338. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110819724.2.299

Chamoreau, Claudine. 2012. Spanish diminutive markers *-ito/-ita* in Mesoamerican languages: A challenge for acceptance of gender distinction. In *Morphologies in Contact* [Studia Typologica 10], Martine Vanhove, Thomas Stolz, Aina Urdze & Hitomi Otsuka (eds), 71–91. Berlin: De Gruyter.

Grimes, Barbara F. (ed.). 2005. *Ethnologue: Languages of the World*, 18th edn. Dallas TX: Summer Institute of Linguistics.

Hajek, John & Williams-van Klinken, Catharina. 2003. Um sufixo românico numa língua austronésia: *-dor* em Tetun (A Romance suffix in an Austronesian language: *-dor* in Tetun). *Revue de Linguistique Romane* 67: 55–65.

Hajek, John, & Williams-van Klinken, Catharina. 2019. Language Contact and Gender in Tetun Dili: What Happens When Austronesian Meets Romance? *Oceanic Linguistics*, 58 (1), 59–91.

Hull, Geoffrey. 2001. A morphological overview of the Timoric Sprachbund. *Estudos de Línguas e Culturas de Timor-Leste / Studies in the Languages and Cultures of East Timor* 4: 98–205.

Hull, Geoffrey. 2002a. *The languages of East Timor: Some basic facts*. Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.

Hull, Geoffrey. 2002b. *Standard Tetum-English Dictionary*, 3rd edn. Winston Hills: Sebastião Aparício da Silva Project in association with Instituto Nacional de Linguística.

Hull, Geoffrey & Eccles, Lance. 2001. *Tetum Reference Grammar*. Winston Hills: Sebastião Aparício da Silva Project, in conjunction with Dili: Instituto Nacional de Linguística, Universidade Nacional de Timor Lorosa'e.

Hull, Geoffrey Stephen & Guterres Correia, Adérito José. 2005. *Kursu gramátika Tetun ba profesór, tradutór, jornalista no estudante universidade sira* (Tetun Grammar Course for Teachers, Translators, Journalists and University Students). Dili: Instituto Nacional de Linguística.

Instituto Nacional de Linguística. 2002. *Matadalan ortográfiku ba Tetun-Prasa* (Spelling Guide for Tetun Prasa). Dili: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.

Janda, Richard D. 1995. From agreement affix to subject "clitic"– and bound root: *-mos > -nos* vs. (-)*nos*(-) and *nos-otros* in New Mexican and other regional Spanish dialects. In *CLS 31: Papers From the 31st Regional Meeting of the Chicago Linguistic Society, Vol. II: The Parasession on Clitics*, Audra Dainora, Rachel Hemphill, Barbara Luka, Barbara Need & Sheri Pargman (eds), 118–139. Chicago IL: Chicago Linguistic Society.

Morris, Cliff. 1984. *Tetun – English Dictionary*. Canberra: Pacific Linguistics.

National Statistics Directorate. 2006. *Timor-Leste Census of Population of Housing 2004: Atlas*. Dili.

Schulte Nordholt, Herman G. 1971. *The Political System of the Atoni of Timor*. The Hague: Martinus Nijhoff.

Thomason, Sarah Grey & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.

Thomaz, Luis Filipe F. R. 1981. The formation of Tetun-Praça, vehicular language of East Timor. In *Papers on Indonesian Languages and Literatures*, Nigel Phillips & Anwar Khaidir (eds), 54–83. Paris: Cahier d'Archipel 13.

Thomaz, Luís Filipe F. R. 2002. *Babel Loro Sa'e: O Problema Linguístico de Timor-Leste* (Babel of the East: The Linguistic Problem of East Timor). Lisboa: Instituto Camões.

van Klinken, Catharina. 1999. *A Grammar of the Fehan Dialect of Tetun, an Austronesian Language of West Timor*. Canberra: Pacific Linguistics.

Williams-van Klinken, Catharina, Hajek, John & Nordlinger, Rachel. 2002. *Tetun Dili: A Grammar of an East Timorese Language*. Canberra: Pacific Linguistics.

Williams-van Klinken, Catharina, Robert Williams & Helio Brites da Silva. 2016. Tetun Dili laos Dili nian ona: Tetun nudar lian inan tuir dadus sensus nian (Tetun Dili no longer belongs to Dili: Tetun as mother tongue according to census data). In *Timor-Leste: The Local, the Regional and the Global*, Sarah Smith, Antero B. da Silva, Nuno Canas Mendes, Alarico da Costa Ximenes, Clinton Fernandes & Michael Leach (eds), 65–71. Dili, Timor Leste: Timor Leste Studies Association.

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Malden MA: Blackwell.

The origins of isolating word structure in eastern Timor

Antoinette Schapper

Vrije Universiteit Amsterdam / Lacito-CNRS

This paper addresses the issue of isolating word structure and its origins in the Austronesian and Papuan languages of eastern Timor. McWhorter (2007) claims that both groups of languages evidence extensive loss of grammatical complexity as a result of "interrupted transmission" due to significant non-native acquisition. I refute McWhorter's assertion that the eastern Timor languages are not "normal" through a detailed exposition of their morphological complexities. Whilst recognising that they are isolating leaning, I argue that there is nothing "unnatural" about the grammars of these languages and that phonological changes within the Timorese Sprachbund provide sufficient explanation of their morphological profiles.

Keywords: Timor languages, phonological erosion, irregularity, lexicalisation, isolating word structure, convergence

1. Introduction

McWhorter (2007) attempts an exciting piece of research, putting forth the radical argument that all cases where a language appears to have been simplified to an extent not explainable by means of regular linguistic change are due to the intervention of non-native learners. Claims about the apparent lack of complexity in some languages have to be taken with several grains of salt, since McWhorter's (2007) criteria for assessing complexity are of highly debatable heuristic value (see, e.g., papers in Sampson, Gil, and Trudgill 2009 for an assessment of different claims, among many others). More significant for this paper, however, is the radical claim that particular languages evidence extensive loss of grammatical complexity such that they must be regarded as having had their normal development "interrupted" by widespread non-native acquisition. I shall focus on McWhorter's claims about the languages of eastern Timor that are at odds with actual language data. I argue

that the claims reveal a lack of concern for explaining patterns in language over asserting a model.

This paper addresses the issue of isolating word structure and its origins in the Austronesian and Papuan languages of eastern Timor. The island of Timor is located at the eastern end of the Minor Sundic island chain (Map 1). It is home to around a dozen languages of the Austronesian family (speckled on the map) and four Papuan languages of the Timor-Alor-Pantar family (dark grey on the map). Both groups of languages of the eastern half of Timor are largely isolating, typically having little inflectional morphology beyond a set of verbal agreement affixes and a set of possessor affixes. I look at claims made in McWhorter (2007, repeated in 2008) that the isolating structure of the languages of eastern Timor evidence their reduced complexity as the result of significant non-native acquisition in the past. I argue that the evidence for McWhorter's scenario of 'interrupted transmission' causing morphological 'stripping' and overall reduced complexity is weak, and does not stand up to scrutiny on proper examination of data from the languages of eastern Timor. I show that these languages evidence the kinds of complexity that McWhorter says to be typical of older languages. In fact, I propose that morphological loss can be accounted for by patterns of phonological changes shared across eastern Timor languages.

This paper is structured as follows. Section 2 outlines McWhorter's view of complexity, while Section 3 presents his arguments for viewing eastern Timor languages as having wide non-native acquisition in their histories. I take a deep dive into data from the Austronesian and Papuan languages of eastern Timor in Sections 4 and 5 respectively, arguing that there is more morphological complexity to be observed in the languages than McWhorter would have us believe. Section 6

Map 1. The languages of eastern Timor and surrounds

then compares the eastern Timor languages to their nearest neighbours and relatives, with a view to understanding how reduced the languages in fact are. I argue that whilst eastern Timor languages show morphological reduction in comparison to their nearest relatives, they cannot be characterised as drastically morphological stripped and their differences are readily explainable by means of normal processes of language change. In Section 6, I discuss alternative explanations for the isolating structure, both in terms of a Timorese Sprachbund, and a larger, more ancient convergence pattern.

2. McWhorter's complexity

McWhorter argues in a series of publications (2001, 2005, 2007, 2008) that isolating word structure in language is 'unnatural'. McWhorter (2001) writes, "[i]n the uninterrupted transmission of a human language, radical loss of complexity throughout the grammar is neither normal, occasional, nor rare, but impossible…". He continues, '[o]lder languages at all times retain a degree of accreted complexity distinguishing them from languages that were born as pidgins'. That is, he claims that isolating languages arise exclusively as the result of 'interrupted language transmission', either from creolisation (2001, 2005), or from 'Non-hybrid Conventionalized Second-Language varieties' (2007).

McWhorter (2007) elaborates further on his earlier claims that all cases where a language appears to have been simplified to a degree not explainable by means of regular linguistic change are due to the intervention of non-native learners. He defines a matrix of complexity by which the simplicity of a language can be assessed. Three factors contribute to a language's complexity:

i.   *Overspecification*: This refers to the differing degrees to which languages overtly and obligatorily mark semantic distinctions (McWhorter 2007: 21–29). For McWhorter, overspecification means that languages evidence features such as noun class marking including numeral classifiers, possessive classes such as inalienable versus inalienable distinctions, definiteness marking, TAME marking, markers of valency change, multiple degrees of demonstrative gradation, numerous negators expressing different negative semantics, abundant pragmatic particles, etc.

ii. *Structural elaboration*: This refers to the number of rules in morphosyntax and elements in (morpho)phonology that derive surface structures (McWhorter 2007: 29–33). Linguistic features that McWhorter views as structurally elaborate are, for instance, complex morphophonemics, large phonemic inventories or tones with multiple contrasting levels, grammatical gender systems, and declension and conjugation classes.

iii. *Irregularity*: This refers to the lexical specification of grammatical features and of paradigm cells that are not the expression of generalised rules, but must be learned by rote. For McWhorter, irregularity in language can be seen in the presence of features like assignment of grammatical gender/noun classes, irregular plural formation, suppletion in inflectional paradigms, etc.

There are, of course, many other ways in which a language could be seen to display complexity, or a lack thereof. McWhorter (2007: 268) gives the example of Pirahã, an indigenous language of South America, observing that the absence of numerals, colour terms, and clausal embedding in the language could be regarded as evidence of its simplicity. Yet, for McWhorter, these features are irrelevant to complexity as he defines it and are merely the result of an "unelaborated cultural perspective among its speakers". McWhorter sees that despite a small phonemic inventory, Pirahã has ample complexity with its two-tone contrast, inflections for aspect and evidentiality, and nominalisation morphology.

McWhorter maintains that extreme lack of complexity – in the form of over-specification, structural elaboration, and irregularity – is not attributable to chance in the world's languages. In his view, non-native acquisition tends to shave away features such as these as they are less necessary to communication. According to McWhorter, languages which display high levels of the above features are older languages; while languages which do not have these features have had 'interrupted language transmission' as a factor in their history. In the strongest version of his hypothesis McWhorter argues that simple grammars are *impossible* without extensive non-native acquisition.

3. McWhorter's explanation of isolating word structure in Timor languages

Turning to Austronesian languages, McWhorter (2007: 242–251, 2008) makes the observation that extreme isolating structure appears to be cross-linguistically rare within the Austronesian family. He contrasts the morphologically rich profile of many Austronesian languages to that presented by some of the languages of Flores and of Timor. He explains that these languages show "unusual morphological sim-plification" (2007: 247), having "shed all or most of their inflections" (2007: 248). He likens them to creoles, stating that it is "extremely unusual for an older language to hover this closely above [the level of complexity of] creoles" (2007: 251), asserting that their isolating structure must be "traced to heavy non-native acquisition at some point in the past" (2007: 248).

On Timor, McWhorter (2007: 242–251, 2008: 175–181) draws a contrast be-tween the many isolating languages of the eastern half of Timor to those spoken

in the western half. He writes that Uab Meto (also known as Dawan(ese), Atoni or Timorese) has two sets of subject prefixes with unpredictable distributions, metathesis and irregular verbs, 'complex' features which he says to be not present in eastern Timor. Similarly, McWhorter also notes that Rotinese[1] is of greater complexity than the Austronesian languages of eastern Timor, with verbal subject prefixes, eight numeral classifiers and possessive enclitics conditioned by constituent class. He concludes that western Timor languages show the 'normal complexity' (2008: 178) of eastern Austronesian languages that he establishes from inspection of surrounding languages such as Kambera (Sumba), Sika (Flores) or Tukang Besi (south-east Sulawesi).

McWhorter observes that it is not only the Austronesian languages of eastern Timor but also the Papuan ones that are 'morphologically stripped' and 'simplified' (2007: 248). McWhorter further asserts that the Timor languages contrast with the typically synthetic grammars of Papuan languages in general (McWhorter 2007: 248–249) or their nearest Papuan relatives on New Guinea (McWhorter 2008: 178–179). Although no data is presented in support of either of these claims, McWhorter concludes that the shared simplification of Papuan and Austronesian languages on Timor points to a common event of 'interrupted transmission' in their histories. He rejects the idea that the Papuan languages of Timor could have developed in an isolating direction because of contact with their morphologically reduced Austronesian neighbours, declaring that this would mean ascribing a highly unusual degree of structural loss to the contact in question. He admits that areal pressure can result in morphological reduction, but maintains that a mere contact account is not sufficient to explain "why they lost so much morphology overall that they stand as strangely analytic, or analytic-leaning, languages" (McWhorter 2007: 248).

Following Hull (1998, 2001), McWhorter (2007, 2008) hypothesises that the historical event that led to the morphological stripping of Timorese languages was an invasion from Central Maluku approximately eight hundred years ago. This theory is based on a Timorese myth involving incoming Ambonese (Hull 1998: 161–164) and the existence of six Timorese placenames similar to placenames in Ambon (Hull 1998: 162). McWhorter concludes that although the details of the Ambonese invasion are unknown, "treating this migration as the cause of the strangely low level of complexity in Timor languages' grammar is more scientific than ascribing the anomaly to chance" (2008: 181). Whilst it is not the intention of this article to dissect the claims of Hull which McWhorter bases his reasoning on, it is worth

---

1.   Note that there is no single language of Rote, but a cluster of languages, potentially not very closely related to one another (Edwards 2018a, 2018b). "Rotinese" is typically used by McWhorter and in the general literature for the Termanu language as described in the pioneering work of Jonker (1915). I will follow this practice here for the sake of simplicity.

noting that no linguists working in the area have taken them up, let alone even bothered to refute them in writing. Oral traditions such as origin myths are well known in the region to reflect political expediencies rather than historical realities (Wellfelt 2016), while similar placenames are often found over areas where related languages are found due to similarities in the strategies used to name places.[2] What is more, Hull (1998) makes clear that the "Ambonese signature" is strongest in the languages of Roti and Uab Meto, spoken in western Timor and precisely the groups whose languages McWhorter claims to have normal levels of complexity. In short, the historical scenario used by McWhorter to explain isolating structure in eastern Timor is without merit and I won't engage with it further in this paper.

In what follows, I restrict myself to addressing McWhorter's claim that the languages of Timor are 'morphologically stripped' and, in general, lack the complexity of older languages. My discussion of complexity will focus on morphological complexity, though the reader should bear in mind that morphological complexity is only one kind of complexity. McWhorter (2007) himself emphasises this, but also observes that morphological complexity is still the most decisive feature as it constitutes the first wave of grammatical simplification in the wake of which various complexities of other kinds can remain.

4.   Austronesian languages in Timor[3]

The Austronesian languages of Timor are thought to be divided into two subgroups, with some differences between authors (compare Hull 1998 and Edwards 2018a). The Central Timor subgroup is small, containing just Tokodede, Kemak, Mambae and Welaun (Edwards 2019: 42–49). The remaining languages of both east and west Timor, with the exception of Helong, appear to all belong together in a single, large "Timor-Wetar-Babar" subgroup that stretches well beyond Timor to the Babar islands in southern Maluku (see the following for various subgrouping arguments for

---

2.   For example, Lutur is a placename found in the Aru, Kei and Tanimbar islands, but the placename itself is not the result of contact between groups on these islands. *Lutur* is a noun meaning 'fort, stone wall' in the indigenous languages of southern Maluku (Schapper 2019) and places with prominent features such as forts were often named after those landmarks.

3.   The following sources were used for the eastern Timor Austronesian languages discussed here: Dadu'a, Penn (2006); Galoli (aka Galolen), Hull (2003); Habun, Hull (2001); Rahesuk (aka Hresuk), Boarccaech (2013); Idate, Alcantara (2015); Kemak Atsabe, own fieldnotes, Schapper (2009); Kemak Marobo, Chuck Grimes p.c.; Lakalei, Hull (2001); Mambae Ainaro, own fieldnotes; Mambae Same, Grimes et al. (2014); Naueti, Veloso (2016); Tetun Fehan (West Timor), van Klinken (1999); Tetun Terik (East Timor), Hull (2001); Waima'a, Bowden et. al. (nd.), Hull (2002); Welaun (aka Wekais), da Silva (2012), Edwards (2019).

the region: van Engelenhoven 1987, 2009a, 2010; Mills 1991; Taber 1993; Hull 1998; Edwards 2018a: 86–88). Within this group on Timor, the so-called Kawaimina (*Ka*irui, *Wai*ma'a, *Mi*diki and *Na*ueti) languages all are very closely related and seem to have arisen out of a differentiated dialect chain. The low-level subgroupings of other Austronesian languages in eastern Timor have not as yet been established, while those of the western Timor languages have been explored extensively in recent times (e.g., Edwards 2018b).

In what follows, I present a wide-range of data from the Austronesian languages spoken in the eastern half of Timor (including the Central Timor languages), illustrating the diverse range of morphological structures that the languages present.

4.1 Verbal agreement prefixes

As is common in eastern Indonesia, the majority of Austronesian languages in eastern Timor have verbal prefixes agreeing with their subject. Only the Kawaimina languages, Tokodede and Kemak are exceptions, having no known verbal agreement prefixes. Examples of these prefixes are given in Table (1). We see that languages differ in the number of persons that are marked. For most of the languages, these prefixes are only found on vowel- and sometimes h-initial roots (often replacing initial *h*, Hull 2001: 153–154).[4] In Habun, however, the prefixes occur both on vowel and consonant initial roots. In Mambae Ainaro *n-* appears erratically on a small number of vowel-initial verbs.

Table 1. Subject agreement prefixes in the AN languages of eastern Timor

	Galoli	Habun	Welaun	Tetun Fehan	Lakalei	Mambae Ainaro
1SG	*ʔ-*	*k-*	*k-*	*k-*	–	–
2SG	*m-*	*m-*	*m-*	*m-*	*m-*	–
3SG	*n-*	*n-*	*n-*	*n-*	*n-*	*n-*
1PL.INCL	*t-*	*t-*	–	–	–	–
1PL.EXCL	*r-*	*h-/--*	–	–	–	–
2PL	*r-*	*h-/--*	–	–	–	–
3PL	*r-*	*r-*	*n-*	*n-/r-*	–	–

Whilst the Austronesian languages above have just a single set of prefixes, there are others that have multiple sets. For example, Idate has one set that occurs on vowel initial verbs and another on consonant initial verbs (Table 2).

---

4. The appearance of agreement prefixes may also be affected by discourse pragmatics in multi-verb clauses (see, e.g., the description of Tetun Fehan subject agreement in van Klinken 1999: 174–176).

Table 2. Idate agreement prefixes paradigms

	Set 1		Set 2	
		'buy'		'open'
1SG	–	*ala*	*u-*	*uloʔe*
2SG	*m-*	*mala*	*o-*	*oloʔe*
3SG	*n-*	*nala*	*na-*	*naloʔe*
1PL.INCL	–	*ala*	*ta-*	*taloʔe*
1PL.EXCL	–	*ala*	–	*loʔe*
2PL	–	*ala*	–	*loʔe*
3PL	*r-*	*rala*	*ra-*	*raloʔe*

Rahesuk also has two sets the forms of which are phonologically conditioned
(Table 3). Set 1 goes on verbs with an initial sonorant consonant. The full set appears
on initial verbs with initial liquids, but is reduced in different ways on verbs with
initial nasals: on *n*-initial verbs the 3rd person singular prefix *n*- is lost; on m-initial
stems the 3rd person singular is infixed as <n>, while the *m*- prefix for 2nd persons
and 1st person plural exclusive is lost; finally, on *ŋ*-initial stems only the *k*- prefix for
1st person plural inclusive and third person plural is retained. Set 2 prefixes appear
on vowel initial verbs and are used as infixes on *h*-initial verbs. Verbs beginning
with other consonants do not take agreement markers.

Table 3. Rahesuk agreement prefixes paradigms

	Set 1			
	l-initial	'go'	**n-initial**	'sow'
1SG	–	*laʔa*	–	*naho*
2SG	*m-*	*mlaʔa*	*m-*	*mnaho*
3SG	*n-*	*nlaʔa*	–	*naho*
1PL.INCL	*k-*	*klaʔa*	*k-*	*knaho*
1PL.EXCL	*m-*	*mlaʔa*	*m-*	*mnaho*
2PL	*m-*	*mlaʔa*	*m-*	*mnaho*
3PL	*k-*	*klaʔa*	*k-*	*knaho*
	m-initial	'come'	**ŋ-initial**	'swim'
1SG	–	*ma*	–	*ŋaɲi*
2SG	–	*ma*	–	*ŋaɲi*
3SG	*<n>*	*mna*	–	*ŋaɲi*
1PL.INCL	*k-*	*kma*	*k-*	*kŋaɲi*
1PL.EXCL	–	*ma*	–	*ŋaɲi*
2PL	–	*ma*	–	*ŋaɲi*
3PL	*k-*	*kma*	*k-*	*kŋaɲi*

Table 3. (*continued*)

	V-initial	'drink'	h-initial	'(re)turn'
		Set 2		
1SG	–	*enum*	–	*hali*
2SG	*m-*	*menum*	*<m>*	*hmali*
3SG	*n-*	*nenum*	*<n>*	*hnali*
1PL.INCL	*r-*	*renum*	*<r>*	*hrali*
1PL.EXCL	*m-*	*menum*	*<m>*	*hmali*
2PL	*m-*	*menum*	*<m>*	*hmali*
3PL	*r-*	*renum*	*<r>*	*hrali*

Dadu'a is similarly complex in that there are two agreement sets but verbs are lexically assigned to them (forms and examples in Table 4). In addition, prefixation is associated with a raft of morphophonological changes to roots. For instance, verbs with initial /b/ and /p/ show the following changes when set 1 prefixes are attached: *k + b > f* and *h + p/b > f*. In addition, Dadu'a also has a set of verbs with irregular prefixal paradigms, some of which are illustrated in Table 5.

Table 4. Dadu'a agreement prefixes paradigms

		Set 1			Set 2	
		'live'	'close'		'blow'	'make'
1SG	–	*mia*	*pa?a*	–	*afuu*	*oi*
2SG	–	*mia*	*pa?a*	*m-*	*mafuu*	*moi*
3SG	–	*mia*	*pa?a*	*n-*	*nafuu*	*noi*
1PL.INCL	*k-*	*kmia*	**kba?a**	*t-*	*tafuu*	*toi*
1PL.EXCL	*h-*	*hmia*	*fa?a*	*r-*	*rafuu*	*roi*
2PL	*h-*	*hmia*	*fa?a*	*r-*	*rafuu*	*roi*
3PL	*h-*	*hmia*	*fa?a*	*r-*	*rafuu*	*roi*

Table 5. Dadu'a irregular prefixing verb paradigms (irregular forms bolded)

	'die'	'go'	'cut'	'enter'	'injure'	'reach'
1SG	*mate*	*laa*	*looh*	*tama*	**namani**	**raik**
2SG	*mate*	*laa*	**hlooh**	*tama*	**namani**	**raik**
3SG	**nate**	*laa*	*looh*	*tama*	**namani**	**raik**
1PL.INCL	**kmate**	**ka?a**	**klooh**	*tama*	*tamani*	*rai*
1PL.EXCL	**hmate**	**hlaa**	**hlooh**	**tahma**	*ramani*	*hraik*
2PL	**hmate**	**hlaa**	**hlooh**	**tahma**	*ramani*	*hraik*
3PL	**hmate**	**hlaa**	**hlooh**	**tahma**	*ramani*	*hraik*

Far from being stripped of verbal inflectional morphology, we have seen that most Austronesian languages of eastern Timor have subject agreement prefixes that appear on at least a subset of the verbal lexicon. For several languages we find multiple sets of verbal agreement prefixes whose choice of host may be lexical or phonological. Morphophonemic processes also can be observed to frequently play a role in determining the surface forms of prefixes and roots.

4.2 Derivational prefixes and associated complexification

Austronesian languages are well-known for their derivational morphology (see, e.g., Blust 2014). Like other eastern Austronesian languages, however, the languages of Timor have none of the voice morphology that characterises languages in western Indonesia and the Philippines. Nonetheless, a range of Austronesian derivational prefixes is still found in eastern Timor languages. Depending on the language, these prefixes may be fossilised, productive or have, in some cases, even fused with inflectional morphology. They have not been simply 'shed', but are accreted as part of the system. What is more, accreted prefixes can also be observed to have caused considerable complexification in other domains of linguistic structure.

A prime example of this is the large consonant inventories of Waima'a and Naueti that have come into being through the fossilisation of prefixes in the languages. Considerably above the average Timorese language consonant inventory of 12–15 consonants, Naueti has 27 native consonant phonemes (Table 6) and Waima'a 30 (Table 7). Beyond their large size, these consonant phoneme inventories are cross-linguistically unusual in that they include phonological rarities such as ejectives, pre-glottalised consonants, post-glottalised consonants, contrastively aspirated plosives, and voiceless sonorants. The historical source of these additional consonant phoneme series in prefixes is apparent from their being limited to word-initial position and being present in relatively few items compared to their "regular" (non-glottalised, non-aspirated, non-devoiced) counterparts.

Waima'a and Naueti aspirated stops and voiceless sonorants have their origins in the absorption of a prefix through the following steps: PMP *pa-[5] > *ha- (cf. Tetun *ha-*) > *h- > *^{h}C > C^h / _plosive, Ç / _sonorant, as exemplified in (1) and (2). Once

---

[5]. Note that the exact source prefix cannot be regarded as certain at this stage. I associate the origin of the Kawaimina aspirated plosives and voiceless sonorants here with the PMP *pa- which was a causative prefix. However, the lack of causative semantics on many of the relevant forms and cognate causative morphemes on many of the relevant forms in neighbouring languages suggests that another morpheme may also be involved. A reviewer suggests that a fossilised agreement prefix *h- such as that found in Habun an Dadu'a would be a possibility.

Table 6. Naueti consonant phonemes

		Labio-velar	Bilabial	Alveolar	Velar	Glottal
Plosive	voiceless			t	k	ʔ
	voiced		b	d	g	
	aspirated		pʰ	tʰ	kʰ	
Fricative	voiceless			s		h
Nasal	voiceless		m̥	n̥		
	voiced		m	n		
	preglottalised		ʔm	ʔn		
Lateral	voiceless			l̥		
	voiced			l		
	preglottalised			ʔl		
Rhotic	voiceless			r̥		
	voiced			r		
	preglottalised			ʔr		
Approximant	voiceless	w̥				
	voiced	w		j		
	preglottalised	ʔw				

Table 7. Waima'a consonant phonemes

		Labial	(Post)-Alveolar	Velar	Glottal
Plosive	voiceless		t	k	ʔ
	voiced	b	d	g	
	aspirated	pʰ	tʰ	kʰ	
	ejective	p'	t'	k'	
Fricative	voiceless		s		h
	glottalised		sʔ		
Nasal	voiceless	m̥	n̥		
	voiced	m	n		
	glottalised	mʔ	nʔ		
Lateral	voiceless	l̥			
	voiced	l			
	glottalised	lʔ			
Rhotic	voiced	r			
	glottalised	rʔ			
Approximant	voiceless	w̥			
	voiced	w			
	glottalised	wʔ			

having entered the phonemic inventories of the languages, these phonological features appear to have taken on a life of their own, assimilating additional lexemes that were never marked by *pa-.[6] What is more, they have been extended to create new phonemes, notably, /pʰ/ in Waima'a and Naueti appears to have developed analogously to /kʰ/ and /tʰ/ for the assimilation of Tetun (and other) loans with initial f.

Origin of Waima'a-Naueti aspirated plosives

(1) PMP *pa- + *kaən 'eat' > Waima'a, Naueti *kʰaa* (cf. PMP *k > Waima'a-Naueti k, e.g., PMP *kahiw 'tree, wood' > Waima'a, Naueti *kai* 'tree, wood')
PMP *pa- + *takut 'fear' > Waima'a, Naueti *tʰaku* 'fear' (cf. Tetun *hamtaʔuk* 'be in fear (of something)', *taʔuk* 'be fear') (cf. PMP *t > Waima'a-Naueti *t*, e.g., PMP *tasik 'sea' > Waima'a, Naueti *tasi* 'sea')

Origin of Waima'a-Naueti voiceless sonorants

(2) PMP *pa- + *bahuq 'odour, stench' > Waima'a *w̥au*, Naueti *w̥ou* 'stink' (cf. PMP *b > Waima'a-Naueti *w*, e.g., PMP *buaq 'fruit' > Waima'a *wuo*, Naueti *wua* 'fruit')
PMP *pa- + *ma-hataq 'raw, uncooked' > Waima'a, Naueti *m̥ata* 'raw, uncooked, unripe, green' (cf. PMP *m > Waima'a-Naueti *m*, e.g., PMP *manuk 'chicken' > Waima'a, Naueti *manu* 'bird, chicken')
PMP *pa- + *laRiw 'run, run away' > Waima'a *l̥ai* 'quick', Naueti *l̥ai~l̥ai* 'very fast, immediate' (cf. Kemak Atsabe *plai*, Dadu'a *hlai*, Tetun Terik *halai* 'run' < *pa-laRiw) (cf. PMP *l > Waima'a-Naueti *l*, e.g., PMP *qalima 'hand, five' > Waima'a, Naueti *lima* 'arm, hand')

A similar process of prefix absorption can be seen to have given rise to Naueti Waima'a glottalised consonants through the following steps: PMP *ka-[7] > *k- > *ʔ- > ʔC (Naueti, where C is [+sonorant]) > Cʔ (Waima'a, where C is [-plosive]). I present some examples of the emergence of glottalised sonorants in Waima'a-Naueti in (3). Thus far, there are no PMP etyma reflected in Waima'a that have either /sʔ/ or a member of the ejective phoneme series (a known phonetic progression of post-glottalisation). In addition to the absence of corresponding glottalised phonemes in Naueti, this indicates again that these additional phonemes were created by extension, particularly for the assimilation of borrowings.

---

6. Veloso (2016: 4) notes that in general across Kawaimina languages there is a significant lack of correspondence in aspirated stops and voiceless sonorants.

7. In some cases, this prefix may have been originally been PMP *paka- that subsequently reduced to *ka- > *k- (cf. Tetun *hak-*). As with PMP *pa- above, it is by no means certain that PMP *ka was the (only) source for the Kawaimina ejectives and preglottalised stops. Other prefixes or phonological phenomena may have played a role in giving rise to these unusual segments.

Origin of Waima'a-Naueti glottalised consonants

(3)   PMP *ka- + *nahik 'climb' > Naueti *ʔnai*, Waima'a *nʔai* 'climb' (cf. PMP *n >
      Waima'a-Naueti n, e.g., PMP *nunuk 'banyan tree' > Waima'a *(kai-)nunu*,
      Naueti *nunu*)

      PMP *ka- + *muRmuR 'gargle, rinse the mouth' > Naueti *ʔmumu* 'hold between
      teeth', Waima'a *mʔumu* 'rinse' (cf. Tetun *hak-mumu* 'to wash or rinse the mouth,
      to gargle') (cf. PMP *m > Waima'a-Naueti *m*, e.g., PMP *matay 'die, dead' >
      Waima'a, Naueti *mata*)

      PMP *ka- + *waRi 'sun' > Naueti *ʔwai* 'dry in sun', Waima'a *wʔai* 'dry in sun'
      (cf. PMP *w > Waima'a-Naueti *w*, e.g., PMP *wahiR 'water' > Waima'a, Naueti
      wai)

Whilst the same level of phonological complexification that we observe in Waima'a
and Naueti is not found elsewhere, PMP verbal morphology is not simply shed in
other Austronesian languages of Timor. Moreover, we do find instances of fossilised
morphology giving rise to additional phonemes in other Timorese languages. For
instance, *m-* (< PMP *ma-) is found fossilised on many stative monovalent verbs in
many languages of eastern Timor, as illustrated with Kemak Atsabe in (4). PMP *p
is usually reflected as /p/ and PMP *b as /h/ in Kemak Atsabe. However, *p > b in
Kemak Atsabe under prefixation of *ma- as follows: *ma- + *p > *mp > *mb > b
(Blust 2008: 96–97). In other situations where consonant clusters are created due to
the presence of *m-*, an epenthetic vowel is inserted between /m/ and the first con-
sonant of the root. This unstressed vowel may be realised as schwa or harmonised
to the first vowel of the root.

Kemak Atsabe stative prefix

(4)   *mdu*   'sit'   [mə'du ~ mu'du]   < PMP *ma- + *tudan
      mnahu 'fall'   [mə'nahu ~ ma'nahu]   < PMP *ma- + *nabuq
      mnaru 'long'   [mə'naru ~ ma'naru]   < PMP *ma- + *anaduq
      banasa 'hot'   [ba'nasa]   < PMP *ma- + *panas

The high frequency with which some prefixes are found suggests that prefixes may
have been productive until quite recently. An example of this is the *k-* (< *ka-, itself
perhaps a reduction of PMP *paka-) prefix which is found on a large number of
Tokodede verbs, but not on cognates in nearby languages, as in (5). As in the Kemak
examples above, an epenthetic vowel is inserted between /k/ and the first consonant
of the root to break up the resulting consonant cluster.

Tokodede *k-* prefix

(5)   *kbaas* 'slap'   [kə'baas ~   < *k-baas   cf. Welaun *basa*, Kemak *basa*,
                       ka'baas]   Mambae *baas*, Tetun *basa* 'slap'

      kdede 'knock'   [kə'dede ~   < *k-dede   cf. Tetun *dere* 'hit repeatedly'
                        ke'dede]

      kmus 'kiss, suck'   [kə'mus ~   < *k-mus   cf. Kemak Atsabe *muusu* 'suck'
                            ku'mus]

      kdula 'turn'   [kə'dula ~   < *k-dula   cf. Tetun *dulas*, *hak-dulas*,
                       ku'dula]   Dadu'a *dulah* 'twist, wind'

      kdole 'crawl'   [kə'dole ~   < *k-dole   cf. Tetun *dolar*, Dadu'a *dolah*
                        ko'dole]   'crawl'

      kbut 'close eyes'   [kə'but ~   < *k-but   cf. Welaun *buta* 'sleep'
                            ku'but]

In fact, other eastern Timor Austronesian languages have clearly productive verbal prefixes marking valency changes. Most widespread are causative prefixes (e.g., Welaun *a-* and Tetun *ha-*, reflecting PMP *pa-). Idate has a fuller range of valency changing suffixes: *si-* (and its allomorph *di-* occurring before /l/) marking an anti-causative derivation of a transitive verb (6); *a-* marking a causative derivation of a intransitive verb, (7), and; *ma-* marking stativity (as opposed to dynamic) on intransitive verbs (8).

Idate valency-changing morphology

(6)   Transitive ~ anticausative alternation
      a.   *au* **u-loʔe**   *lala   mata-k*
           1SG 1SG-open   path eye-1SG
           'I open the door (lit. path eye).'
      b.   *lala   mata-k*   **di-loʔe**
           path eye-1SG   ANTIC-open
           'The door (lit. path eye) opens itself.' or 'The door is open.'

Intransitive ~ causative alternation

(7)   a.   **turu,**   *asu!*
           descend dog
           'Get down, dog.'
      b.   *ami*   **a-turu**   *bandera*
           1PL.EXCL CAUS-descend flag
           'We lower the flag.'

Stative ~ causative alternation

(8)   a.   *au* **ma-nahu** *hori kareta*
           1SG STAT-fall from car
           'I fell out of the car.'

b. *au* **a-nahu** *livru*
1SG CAUS-fell book
'I dropped the book.'

Of course, derivational morphology such as this does not have the same status as inflectional morphology in McWhorter's model (2008: 18–20). But in many eastern Timor Austronesian languages, causatives are not synchronically separable from inflections. Table 8 presents languages in which we find subject agreement prefixes have fused with causative *pa-. In Waima'a *ra-* (a fusion of a 3PL prefix *da- with *pa-) has generalised to all persons; the absence of cognates in its sister language, Naueti, may indicate the Waima'a causative is a borrowing from Galoli.

Table 8. Subject agreement markers fused with causative

	Eastern Tetun[†]	Dadu'a	Waima'a
1SG	*ka-*	*a-*	
2SG	*ma-*	*ma-*	
3SG	*na-*	*na-*	
1PL.INCL		*ta-*	*ra-*
1PL.EXCL	*ra-*		
2PL		*ra-*	
3PL			

† These come from Hull (2001: 150). Note that van Klinken (1999: 172) does not regard the Tetun Fehan verbal inflections fused with the causative *ha-* prefix, but rather a result of regular morphophonological rule where initial *h* is replaced by an inflection, as described in the previous section. Such a rule would, presumably, explain the origin of the fused forms.

Galoli itself has two series of valency changing morphemes that are fused with subject inflections, as presented in Table 9. The anticausative paradigm is characterised by several morphophonemic rules that determine its surface form (on b-initial verbs, its form is *Ca-*, on g-initial verbs *Cam-*, on s-initial verbs *Can-*, on l-initial verbs *Car-*, and elsewhere *Cak-*).

Table 9. Galoli fused subject agreement and valency changing prefixes

	Causative	Anticausative
1SG	*ʔa-*	*ʔak-*
2SG	*ma-*	*mak-*
3SG	*na-*	*nak-*
1PL.INCL	*ta-*	*tak-*
1PL.EXCL		
2PL	*ra-*	*rak-*
3PL		

4.3 Possessive morphology and possessive classes

Possession in the Austronesian languages of eastern Timor is a domain that exhibits a range of complex structures.

Numerous Austronesian languages of eastern Timor have a paradigm of person-number suffixes occurring on nouns that encode possessors (Table 10). These suffixes represent continuations of conservative Austronesian inflectional morphology, despite some obvious paradigm levelling. Even in languages like Kemak which is lacking verbal prefixes, we find a full paradigm of possessor suffixes. Likewise, Waima'a and Naueti retain the possessive suffix *-n* in the third person, although they have no verbal inflections whatsoever.

Table 10. Possessor suffixes

	Galoli	Kemak Atsabe	Idate	Lakalei	Tetun Fehan	Waima'a	Naueti
1SG	*-k*	*-gV*	*-k*	*-k*		–	–
2SG	*-m/--*	*-mV*			*-n*	–	–
3SG	*-n*	*-V*	*-n*			*-n*	*-na*
1PL.INCL						–	–
1PL.EXCL				*-n*		–	–
2PL	*-r*	*-rV*	*-r*		*-n/-r*	–	–
3PL						*-n*	*-na*

While one suffixal set is typical, Welaun has two sets of phonologically conditioned possessive suffixes: Set 1 used on nouns with a final vowel, and Set 2 used on nouns with a final consonant (Table 11). Dadu'a is unusual in Timor in that it has a paradigm of possessor prefixes (Table 12).

Across eastern Indonesia and Oceania, possessive suffixes are associated with inalienable possession, while unbound possessive markers, typically preceding the possessum, are widely used for alienable possession (Donohue & Schapper 2008).

Table 11. Welaun possessor suffixes

	Set 1	Set 2
1SG	*-k*	*-aak*
2SG	*-n*	*-aan*
3SG	*-n*	*-aan*
1PL.INCL	*-t*	*-aat*
1PL.EXCL	*-t*	*-aat*
2PL	?	?
3PL	*-n*	*-aan*

Table 12. Dadu'a possessor prefixes

1SG	*a-*
2SG	*o-*
3SG	*ni-*
1PL.INCL	*ita-*
1PL.EXCL	*ami- ~ am-*
2PL	*mi-*
3PL	*sia- ~ si-*

In eastern Timor, such a system of possessive classification is found in Kemak. We see in (9a) that the possessor of the inalienable body part noun *gara-* 'head' is encoded by a possessive suffix, while in (9b) a free possessive pronoun encodes the possessor of the alienable noun *uma* 'house'.

Kemak Atsabe possessive classes

(9) Inalienable
   a. *gara-**ga***
       head-1SG
       'my head'
   Alienable
   b. ***au***   *uma*
       1SG.POSS house
       'my house'

In most Austronesian languages of eastern Timor, however, the morphosyntactic distinction between alienable and inalienable possession such as found in Kemak has broken down. Possessive suffixes and free possessive markers have instead entered into different paradigmatic relationships with one another. In Waima'a and Naueti, the split is in person: 3rd person possessors of all kinds are encoded with a suffix, e.g., Naueti *-na* (10a and b), while a free possessive marker is used for other persons, such as the Naueti 1st person singular possessive form *au* (10c and d).

Naueti possessive coding

(10) 3rd person possessors

   a. *uma-**na***
       house-3
       'his/her/their house'
   b. *lima-**na***
       arm-3
       'his/her/their arm'

1st person possessors

c. **au**   *uma*
   1SG.POSS house
   'my house'

d. **au**   *lima*
   1SG.POSS arm
   'my arm'

In Idate, the split between possessive marking strategies is morphophonological: nouns ending in a vowel take possessor suffixes (11a and b), while nouns ending in a consonant take free possessive markers (11c).

Idate possessive coding

(11)   Vowel-final nouns

   a.   *namo-**k***
        garden-1SG
        'my garden'

   b.   *ibo-**k***
        mouth-1SG
        'my mouth'

   Consonant-final nouns

   c.   *betuk*   **auk**
        bamboo 1SG.POSS
        'my bamboo'

In these Austronesian languages, a remnant of the system of possessive classification is that typical inalienable nouns like body part nouns occur obligatorily with a possessor, while alienable nouns do not require the expression of a possessor to be well-formed. This feature is relatively rare world-wide and represents one manifestation of alienable/inalienable possessive systems that McWhorter deems complex (Bickel & Nichols 2013). Timorese languages add the further complexity of allowing free possessive markers to be either pre-posed or post-posed to the possessed noun, with which fine shades of closeness in possessive relations can be signalled (see van Klinken 1999: 145–152; Schapper 2009).

Possessive morphology is also widely found in the Austronesian languages of eastern Timor in non-possessive contexts, with possessor morphology being used in attributive constructions (cf. Ross 1998 on similar constructions in Oceanic). In Kemak, for instance, nominal (and less often verbal) attributes can be marked as if they were possessed with the 3rd person affix *-V*, while the referent noun of the NP behaves like a possessor, occurring before the possessed attribute. Attributive marking with the possessor suffix is not limited to inalienably possessed nouns, but is found on a wide range of items, such as *tasi* 'sea' and *mate* 'dead' in (12).

Kemak Atsabe possessive attributive marking

(12) a. *nipe tasii* 'sea snake'   < *nipe* 'snake'   + *tasi-V* 'sea-3SG'
     b. *nua matee* 'old coconut' < *nua* 'coconut' + *mate-V* 'dead-3SG'

The parallel between possessive and attributive marking is illustrated on the basis of Naueti *-na* in (13) and Idate *-n* (14). Similar constructions have been described for Waima'a (Bowden et al. nd) and Welaun (Edwards 2019: 39–40).

Naueti *-na*

(13) Possessive *-na*

    a. *asukai bui-**na***
       man   cat-3
       'man's cat'

    Attributive *-na*

    b. *asukai riku-**na***
       man   rich-3
       'rich man'

Idate *-n*

(14) Possessive *-n*

    a. *ni iwa-**n***
       3SG mouth-3SG
       'his, her mouth'

    Attributive *-n*

    b. *ruut hutu-**n**   isa*
       grass bind-3SG one
       'a bundle of grass'
       (lit. one bound grass)

Again in the domain of possession, the languages of eastern Timor manifest a range of features from inflectional morphology to (in)alienability contrasts that are not consistent with McWhorter's picture of stunningly 'stripped' languages.

4.4   Synchronic metathesis

Synchronic metathesis refers to a process whereby the expected linear ordering of sounds in a word is reversed in certain morphosyntactic environments, thus, xy becomes yx. Metathesis is a striking feature in the Austronesian languages of south-west Maluku, being reported in many of the languages in Timor and southern Maluku (Schapper 2015: 135–138). McWhorter claims that this typologically unusual morphophonological feature is limited to western Timor (found in, e.g., Helong, Bowden

2010; Uab Meto varieties, Edwards 2016). In fact, metathesis is found among the languages that McWhorter points to as being the most simplified in Timor.

In Mambae, one of the languages McWhorter names as being most simplified in Timor (2007: 247), metathesis and the related process of apocope is productively used to mark dependency relationships in phrases. In Mambae Ainaro, for example, many nouns have two forms: a vowel-final form which appears phrase finally, and a consonant-final form which appears phrase non-finally, as in the NN compounds in (15). The different forms that are realised in Mambae Ainaro by metathesis are the result of interactions of vowels with one another. Final high vowels such as in (15a) are maintained when they metathesise into a position next to a low vowel (15b). The mid-vowel /e/ (15c) is assimilated when it metathesises into a position next to /i/ (15d).

Mambae Ainaro metathesis
  (15) Vowel-final form
      a.   *kud*   **tali**
           horse rope
           'bridle' (lit. 'horse rope')

      Consonant-final form
      b.   **tail**   *mata*
           rope eye
           'trap' (lit. 'rope eye')

      Vowel-final form
      c.   *an*   **hine**
           child female
           'daughter'

      Consonant-final form
      d.   **hiin**   *ana*
           female child
           'girl'

For the Same dialect of Mambae, Chuck Grimes (p.c.) estimates metathesis to affect some 30% of the nouns and verbs of the language, as well as some members of closed classes such as pronouns and numerals. A selection of the items from Grimes et al. (2014) is presented in Table 12. Here we see that metathesis results in several morphophonemic changes in the surface form of the metathesised items. Most obvious is the assimilation of final /a/ to the quality of the previous vowel on metathesis. Additionally, on metathesis, final /i/ optionally lowers to /e/ when the preceding vowel is /a/.

Kemak Marobo also has synchronic metathesis, though it appears to be less extensive than that found in Mambae dialects (Chuck Grimes p.c.). (16) presents some preliminary examples of the metathesis. The morphosyntactic rules governing

Table 12. Metathesising items in Mambae Same

	Vowel-final	Consonant-final		Vowel-final	Consonant-final
1PL.EXCL	*ami*	*aim ~ aem*	'mouth'	*kuku*	*kuuk*
'when'	*arfila*	*arfiil*	'tongue'	*lama*	*laam*
'child'	*ana*	*aan*	'hand, arm'	*lima*	*liim*
'slap'	*basa*	*baas*	'front'	*muna*	*muun*
'approach'	*fedesi*	*fedeis*	'money'	*osa*	*oos*
'crow' (v)	*foni*	*foin*	'road'	*sala*	*saal*
'gather'	*futu*	*fuut*	'return'	*sila*	*siil*
'stone'	*hatu*	*haut*	'grandparent'	*tata*	*taat*
'female'	*hina*	*hiin*	'year'	*tona*	*toon*
'one'	*ida*	*iid*	'worm'	*ula*	*uul*
1PL.INCL	*ita*	*iit*	'house'	*uma*	*uum*

this phenomenon in Kemak remain to be determined by future work. What is clear from these examples is that several morphophonemic rules likely impact on the surface form of the metathesised items, just as in Mambae.

Kemak Marobo metathesis

(16) Vowel-final form

  a. **manu**   *hui*
      chicken wild
      'bird' (lit. 'wild chicken')

  Consonant-final form

  b. **man**   *telo-n ~ maun telo-n*
      chicken egg-3SG
      'chicken egg'

  Vowel-final form

  c. **ama**   *naʔi*
      father royal
      'father's elder brother'

  Consonant-final form

  d. **aam**   *cuan*
      father old
      'grandfather'

In sum, synchronic metathesis is attested to be present in what are among the most affix-poor languages in eastern Timor.[8]

---

8. Owen Edwards (p.c.) states that there is also reason to believe that synchronic metathesis is present in Tokodede, a close relative of Mambae and Kemak.

4.5   Numeral agreement

Numeral classifiers are found in the Austronesian languages of eastern Timor, how-ever, in relatively small numbers compared to many other Austronesian languages. Of the languages which have been described, Tetun Fehan has around a dozen different numeral classifiers (van Klinken 1999: 140ff), and at least four numeral classifiers have been identified for Naueti (Veloso 2016: 46). Despite this, there is a feature of numerals in Kawaimina languages that is similar to numeral classifica-tion, but is more complex in that it is inflectional.

Simplex numerals between 'two' and 'nine' in both Waima′a and Naueti must agree in animacy with the referent of the quantified noun. Agreement is indexed by prefixes on the numeral distinguishing HUMAN versus NONHUMAN. The agreement behaviour for numerals is illustrated for the two languages in (17) and (18). In both languages the quantity interrogative 'how much, many?' also takes the numeral agreement prefixes to agree with the animacy of the referent whose quantity is questioned.

Waima′a numeral agreement

(17) HUMAN agreement

   a. *anu-ata **wuo**-hitu*
       woman HUM-seven
       'seven women'

   NONHUMAN agreement

   b. *kumu **kai**-hitu*
       pigeon NHUM-seven
       'seven pigeons'

Naueti numeral agreement

(18) HUMAN agreement

   a. *kii   **wua**-lima*
       person HUM-five
       'five people'

   NONHUMAN agreement

   b. *uma   **kai**-lima*
       house NHUM-five
       'five houses'

The Waima′a and Naueti agreement prefixes represent grammaticalisations of nu-meral classifiers that were independent lexical items: the HUMAN prefix originates in PMP *buaq 'fruit', while the NONHUMAN prefix is from PMP *kahiw 'tree, wood'. In both languages, the numeral agreement prefixes are still transparently related to

the lexemes from which they grammaticalised (cf. Waima'a *wuo*, Naueti *wua* 'fruit', and Waima'a, Naueti *kai* 'tree, wood').

A comparative study of numerals in Timor and surrounds suggests that this grammaticalisation did not occur in the immediate ancestor of Waima'a and Naueti (i.e., proto-Kawaimina), but at an earlier stage within the subgroup. Right across languages of the eastern half of the Timor-Babar subgroup, numerals between two and nine occur with fossilised agreement prefixes, reflecting the HUMAN agreement prefix from PMP **buaq* (e.g., numerals 2–5 in Kisar *wo-ro?o*, *wo-kelu, wo-?akka, wo-lima*, and in Dadu'a *wa-rua*, *wa-telu, wa-ak, wa-lima*).[9] This indicates that, despite exhibiting little person-number inflection on verbs and nouns, Kawaimina languages have not been entirely stripped of inflectional morphology, but preserve agreement prefixes that have lost productivity in nearby related languages.

5.   Papuan languages of Timor[10]

There are four Papuan languages spoken in two parts of Timor: Bunaq is located in central Timor, while Makasae, Makalero and Fataluku occupy a contiguous region at the island's eastern tip. On Kisar Island, just off the north-eastern end of Timor, is a fifth language, Oirata, a close relative of Fataluku. Bunaq and the four Eastern Timor languages form two primary subgroups of the Timor-Alor-Pantar (TAP) language family. Within the Eastern Timor subgroup, Fataluku and Oirata subgroup together, as do Makasae and Makalero. The remaining members of the family are spoken on and between the Alor and Pantar islands, forming a third primary subgroup of the family.

The nearest relatives of TAP languages, Mbaham, Iha and Kalamang, are spoken on and around the Bomberai peninsula at the western tip of New Guinea. Together, TAP languages and West Bomberai languages have been seen to form the western extreme of the hypothesised Trans-New Guinea family (Usher & Schapper ms).

In the following sections, I illustrate the structures of the Papuan languages of Timor, highlighting areas in which they display indexes of McWhorter's complexity, namely, overspecification, structural elaboration and irregularity.

---

9.   It is not clear that the NONHUMAN agreement prefix from PMP **kahiw* is found outside of Kawaimina languages. The numerals 2–5 in Galoli *i-rua, i-telu, i-haat, i-lima*, for example, still appear to be reflexes of the HUMAN prefix.

10.   The following are the sources for the Papuan languages discussed here: Bunaq, Schapper (2010a), own fieldnotes; Fataluku, van Engelenhoven (2009b), van Engelenhoven & Huber (2020), Heston (2015); Makalero, Huber (2011); Makasae, Huber (2008), Correia (2011); Oirata, de Josselin de Jong (1937).

5.1   Person agreement prefixes

A single paradigm of agreement prefixes, occurring on both verbs and nouns, is found in Bunaq (Table 13). The paradigm consists of three person prefixes and two valency-reducing prefixes. On prefixation to consonant initial roots, consonantal prefixes appear with an epenthetic vowel. Vowels in prefixes harmonise with the first vowel of the root, similar to what is found with prefixes in the neighbouring Austronesian languages, Kemak and Tokodede (see Section 4.2). Where the root is vowel initial, the prefixal vowel is deleted. This means that for the first person inclusive and second person, there is no surface manifestation of the prefix (marked as Ø-). On verbs, a prefix typically coindexes an animate P, though there is a small number of verbs with a prefix for animate S. On nouns, a prefix indicates an inalienable possessor.

Table 13. Bunaq agreement prefixes with examples

		bol 'value'	*wit* 'fetch'	*il* 'water'	*obon* 'hang'
1EXCL	n-	no-bol	ni-wit	n-il	n-obon
1INCL/2	V-	o-bol	i-wit	Ø-il	Ø-obon
3AN	g-	go-bol	gi-wit	g-il	g-obon
REFL	d-	do-bol	di-wit	d-il	d-obon
RECP	t-	to-bol	ti-wit	t-il	t-obon

The Bunaq agreement system has a host of lexical complexities. There are seven different conjugational classes of verbs, chiefly based on inflectional behaviour of the verb in the 3rd person. The class with the largest membership is the zero conjugation; this has no agreement prefix coindexing a 3rd person inanimate P but uses the prefix *g-* for a 3rd person animate P. The remaining conjugations are distinguished by the initial consonants which are present on the verb with a 3rd person inanimate P. As illustrated in Table 14, these initial consonants are replaced by the *g-* prefix when the P is animate. Not all verbs that have these initial consonants show this

Table 14. Conjugation classes of Bunaq verbs

	INANIMATE	ANIMATE	
zero conjugation	*iwal*	*giwal*	'pick'
	teke?	*geteke?*	'look at'
h-conjugation	*hukat*	*gukat*	'lift'
s-conjugation	*sumi*	*gumi*	'hide'
t-conjugation	*tinik*	*ginik*	'cook'
d-conjugation	*doenik*	*goenik*	'remember'
l-conjugation	*logo*	*gogo*	'move sth'

replacement. Conjugation classes are therefore not the result of predictable morphophonemic processes, but rather their membership is a lexical property of verbs. Most of the initial consonant conjugation classes contains many dozens of members, involving therefore a considerable amount of learning on the part of speakers.

Bunaq also has various morphophonemic rules associated with agreement prefixes. Notable amongst these is the metathesis of roots of the shape $CV_1{}'V_2C$ under prefixation where V_1 is high and V_2 non-high to the shape $-V_1{}'CV_2C$. Table 15 presents two examples, one noun *luel* and one verb *sieʔ*.

Table 15. Examples of Bunaq metathesis under prefixation

		luel 'peel'	*sieʔ* 'tear'
1EXCL	*n-*	*n-ulel*	*n-iseʔ*
1INCL/2	*V-*	*Ø-ulel*	*Ø-iseʔ*
3AN	*g-*	*g-ulel*	*g-iseʔ*
REFL	*d-*	*d-ulel*	*d-iseʔ*
RECP	*t-*	*t-ulel*	*t-iseʔ*

Numerous irregular root mutations are found on Bunaq verbs under prefixation. The changes in roots typically involve the duplication of a segment or a segment's deletion, as illustrated in Table 16. The changes are not predictable based on the shape of the root.

Table 16. Examples of Bunaq irregular verb root mutation under prefixation

	Unprefixed form	**3rd person prefixed form**
'split'	*bagal*	*gagabal*
'gather'	*binun*	*gibibun*
'wash'	*ili*	*gigili*
'tell'	*pilaʔ*	*gipiala*
'steal'	*bini*	*gibi*
'beat'	*tuʔu*	*gutuʔ*
'stretch'	*mene*	*gemen*
'clear'	*naman*	*gaman*
'cover'	*bolok*	*gobok*
'break'	*pili*	*gipi*

In Makalero and Makasae, agreement similar to that found in Bunaq is limited to the 3rd person. Cognates of the Bunaq 3rd person inflection *g-* are found on numerous vowel-initial items. Makalero has a set of verbs and preverbs that inflect for *k-* marking a 3rd person P argument (for more details on the morphosyntactic conditions of *k-*, see Huber 2011: 349ff). Makasae has a fossilised *g-* which is a

reflex of the same prefix. Table 17 presents the set of cognates between Makalero and Makasae where reflexes of *g- are found.[11]

From Table 17 we see that the Makasae cognates of the Makalero inflecting verbs have, for the most part, fossilised the prefix and thus almost always have initial /g/. There are few exceptions to this fossilisation pattern, such as Makasae *amuʔu* 'smell, stink'. In some cases, semantics seems to play a role in prefix fossilisation. With *gapu* 'with' and *apu* 'carry, cradle', Makasae retains both prefixed and unprefixed forms, albeit with differing semantics (similar to Makalero). Makasae *umu* 'die' has no fossilised prefix, likely because inflected forms had a transitive meaning 'kill' (cf. Makalero *k-umu-* 'to kill someone by an action', Bunaq *g-ume* 'kill'). Such a transitivising function of *g- is also apparent on the Makasae verb *gira* 'to water' < *g-* + *ira* 'water' (n).

In Makasae we can also see a tendency for the *g-* prefix to be retained in contexts where it is 'trapped' between morphemes. This is apparent from the Makasae verb *ena* 'see, look, watch', which when marked with the intensifying prefix *nehe-* exhibits the prefix *nehe-gena* 'see from afar'. "Trapped" *g-* is also found fossilised on nouns in Makasae with inalienable semantics in other compounds. For example, Makasae *gaʔawai* 'place' occurs in compounds such as *basara-gaʔawai* 'marketplace' and *tana-gaʔawai* 'fingerprint' (lit. hand place), or Makasae *gauhaa* 'master, owner' which occurs in compounds such as *oma-gauhaa* 'host' (lit. house owner) and *keta-gauhaa* 'farmer' (lit. ricefield owner).[12] Similarly, an inflected form of *ira* 'water' is also found in the Makasae compound *awa-gira* 'penis water' for 'sperm' (cf. Bunaq *g-il* 3-water for 'juice (of a fruit), bodily fluid'). Makalero does not appear to retain reflexes of *g- on nouns, instead using a newly grammaticalised form of the third person pronoun *ki* (cf. Makasae *gi* '3') in its place.

The evidence from Makasae and Makalero indicates that in their immediate common ancestor the third person agreement *g- was still used productively on vowel-initial (and some h-initial) nouns and verbs to mark inalienable possessors and P arguments respectively. By contrast, in their closest relatives, Fataluku and Oirata, *g- has not been preserved. However, there is evidence that a cognate agreement prefix was present in the common ancestor of these languages.

---

11. These are not the only items with a reflex of *g- on them in Makalero and Makasae. There are several other inflecting verbs in Makalero, but they do not have known cognates in Makasae. They are: *ako, k-ako* 'steal', *asu, k-asu* 'for', *ati-, k-ati-* 'downwards', *horu, ko-horu* 'with', *uan-, k-uan-* 'bigger, more', *uri, k-uri-* 'obscured, hidden', *uta-, k-uta-* 'hidden from view', *aʔa-, k-aʔa-* 'onto'. There are also several *g*-initial items in Makasae which look like they may have a fossilised *g-*. They are: *goba ~ guba* 'with', *gamun(u)* 'hold in hand', *gume* 'pick fruit', *guhur(u)* 'blow'.

12. These nouns are frequently inalienably possessed in Timor languages, for instance, Bunaq *g-omo* 'owner, master' and *g-oloʔ* 'place, spot'.

Table 17. Comparison of appearance of reflexes of *g- prefix on cognate roots in Makalero and Makasae

Makalero			Makasae			
vowel form		k-form		vowel form		g-form
ali-	'all over'	kali-		–		gali	'back, around'
afa-	'away from'	kafa		–		gafa	'from, leave behind'†
afi-	'sideways'	kafi-		–		gafi	'beside, next to, across'
afu-	'carry, with'	kafu	'carry a child'	apu	'carry, cradle'	gapu	'with, for, bring, take'
amuʔ	'smell'	kamuʔ		amuʔu	'smell, stink'	–	
ata-	'in contact'	kata-		–		gata	'next to, beside, near'
e-	'firm'	ke-		–		ge-	'firm'
ena	'see'	kena		ena	'see, look, watch'	(nehe)gena	'see from afar'
ene	'hit, strike'	kene		–		gene	'hit, afflict'
eta-	'apart'	keta-		–		geta	'apart'
haʔawein	'place' (n)	–		–		gaʔawai	'place' (n)
ia-	'under'	kia-		–		gia gini	'under, inside'
ini-	'do, make'	kini		–			'do, make, give'
ira	'water'	–		ira	'water'	gira	'to water', 'juice, internal liquid'
isa	'bake, roast'	kisa		–		gisa	'roast'
isi-	'be at'	kisi-	'originate, belong'	isi	'at, in, to, from, since'	–	
ou-	'towards'	kou-		–		gau	'for, towards, at'
ouar	'master, owner'	–		–		gauhaa	'master, owner'
ua-	'top'	kua-		–		gua goe	'on (top of), over'
ue-	'around'	kue-					'around'
umu-	'die, dead'	kumu-	'to kill by an action'	umu	'dead, die'	–	
uta	'kill'	kuta		–		guta	'kill'
utu	'cover, block, wear'	kutu		–		gutu	'put on, wear, bar'
utuʔ	'mind, look after'	kutuʔ		–		gutu	'to look after, keep eye on'

† The only instance of this item in Correia (2011: 74) is glossed as postposition the following: *i bese-bese tagara bo mikorlet i hau gafa riaʔa* 2PL quickly walk so minibus 2PL PERF POSP run 'You must walk quickly, otherwise you will miss the minibus'. The meaning of *gafa* here appears to be malefactive. Huber (p.c.) has one example of *gapa* from the Makasae Ossu dialect where *p* and *f* have not fully merged: *ini wata=e naʔu gapa laʔa* 1PL.INCL coconut=DEF just from go 'We left, leaving the coconuts behind'.

Fataluku has a prosthetic vowel that coreferences an argument on consonant initial verbs. The vowel is found on a restricted set of verbs and is a lexical property of those verbs. Typically, the prosthetic vowel is a copy of the first vowel of the root. However, in a few cases, the prosthetic vowel is unpredictable. Example verbs, both with and without prosthetic vowel, are set out in Table 18.

Table 18. Fataluku verbs with vowel prosthesis (irregular prosthetic vowels bolded)

/a/	'send'	*har*	*ahar*	/o/	'embrace'	*kolev*	*okolev*
	'flare'	*kan*	*akan*		'recognise'	*nof*	*onof*
	'multiply'	*ruka*	*aruka*		'cut up'	*fot*	*ofot*
	'warp'	*ha*	*aha*		'be.inside.PL'	*fo*	*ofo*
/e/	'take'	*me*	*eme*		'be.inside.SG'	*to*	*oto*
	'read'	*ler*	*eler*	/u/	'clothe'	*lavere*	***u**lavere*
	'count'	*keh*	*ekeh*		'sweep'	*lur*	*ulur*
	'wipe'	*fer*	*efer*		'catch in hand'	*nam*	***u**nam*
	'measure'	*te*	*ete*		'grab'	*fal*	***u**fal*
/i/	'cook anew'	*tih*	*itih*		'feed'	*fan*	***u**fan*
	'wait'	*hir*	*ihir*		'plant'	*tu*	*utu*
	'bind'	*sil*	*isil*		'spoon up'	*huleve*	*uhuleve*

The morphosyntactic properties of the Fataluku prosthetic vowel are very similar to the properties displayed by agreement prefixes in its relatives. The prosthetic vowel in Fataluku coindexes P on transitive verbs and S on intransitive verbs.[13] This parallels the split-S distribution of agreement prefixes in the related languages of Alor and Pantar as well as Bunaq. Marking a verb with a prosthetic vowel prevents another prefix from being added to the verb and never triggers initial consonant mutation (see Section 5.4 where this process is discussed as the result of the historical loss of a prefix *n-). These features indicate that the prosthetic vowel is not simply the result of a morphophonemic process, but rather fills the prefix slot on the verb. In Bunaq and Makalero an agreement prefix also cannot co-occur with locative *n-*. These features indicate that what has been described as a prosthetic vowel in Fataluku is in fact an agreement prefix *V-*.

A prefix of this form would be the expected reflex of the Proto-Timor-Alor-Pantar (PTAP) *ga- '3' (reflected in Proto-Alor-Pantar *ga-; Proto-Maka (the common ancestor of Makasae and Makalero) *g- > Makasae fossil *g-*, Makalero *k-*; Bunaq *g-*). PTAP *g > *ʔ medially in Proto-Eastern Timor (PET), but was retained as *g initially. These phones were continued with the same values in Proto-Maka, but in

---

13. Often the prosthetic vowel is anaphoric, referring back to an earlier established argument that is elided in the clause with the prosthetic vowel, but it is by no means restricted to contexts where a referent is not expressed by independent nominal constituents.

Proto-Frata (the common ancestor of Fataluku and Oirata), the sound change progressed with PTAP *g > *ʔ in all positions. However, glottal stop is not contrastive in initial position in Fataluku, thus leaving only an initial vowel as a possible reflex of PTAP *ga- in Fataluku. The remaining vowel reflex of *ga- became unspecified V- either in Fataluku or at an earlier stage.[14] A parallel case supporting the change of *g > *ʔ > Ø initially in Fataluku is the PET pronoun *gi '3.POSS', reflected as Makasae *gi*, Makalero *ki* and Fataluku *i*.

The irregular prefixal vowels that we find on some Fataluku verbs appear to be the result of reanalysis of root vowels as prefixes. For example, the irregular prefixal vowel *u*- on Fataluku *nam-e* 'catch in hand-VBLZ' goes back to PTAP *amun 'grab', a form which is reflected as Makasae *g-amun* 'hold in hand' and Bunaq *amuʔ* 'seize'. The PTAP form metathesised to *unam- in Proto-Frata. The initial *u vowel was reanalysed as an agreement prefix leaving the modern-day root as *nam-* in Fataluku and Oirata. The same reanalysis can also be seen in Fataluku *laver-e* 'clothe-VBLZ', a verb derived from the Fataluku noun *ulavari* 'waist'.[15] The 'prefixed' form of the verb, *ulaver-e*, reflects the original initial vowel segment of the root still preserved on the nominal root.

In sum, like the Austronesian languages, the Papuan languages of Timor are not without verbal inflectional morphology. Bunaq has the most extensive paradigm of agreement prefixes. Prefixation in Bunaq is also associated with lexicalised conjugation classes, morphophonemic processes like metathesis, and a host of irregular root changes. While agreement prefixes are reduced in the Eastern Timor Papuan languages, they have by no means been stripped from the languages. Makalero and Makasae preserve reflexes of PTAP *ga- on vowel initial verbs and nouns. In Fataluku a reflex of the same agreement prefix V- appears on a defined set of consonant initial verbs. Reanalysis of root initial vowels as instantiations of the V- prefix have also led the prefixal vowel to have an unpredictable form on numerous verbs.

5.2   Animacy and agreement

The Papuan languages of Timor manifest animacy distinctions on a range of agreement targets.

---

14. Recall that Makalero and Makasae only have retained agreement prefixes on vowel initial verbs and hence the vowel is lost.

15. Cf. Oirata *ulawara* 'waist, loins'. The verb *laver-e* 'clothe-VBLZ' is likely to have originally meant 'put on a loincloth' or 'wrap cloth around waist'. The initial *ula-* of this item reflects PTAP *[w]ula 'tail' and is found in numerous complex nominals in Fataluku and Oirata, e.g., Fataluku *ulafuka* 'tail', *vehula* 'youngest child'.

Some lower numerals in Eastern Timor Papuan languages have different agreement forms for human versus non-human referents. In Makalero, this semantic distinction applies to 'two', 'three' and the quantifier 'many' (Table 19). In Makasae, 'two', 'three' and 'four' have distinct forms for human and non-human referents (Table 20). Human numerals generally combine with a human classifier, Makalero *amu* 'body' and Makasae *anu* 'person'. The formation of these distinct numerals does not follow any know inflectional pattern in the languages, but does appear to involve fossilised prefixes on the numeral bases (e.g., *lol- for non-human, *mV- for human).

Table 19. Makalero numeral agreement

	NON-HUMAN		HUMAN
'one'		*u(n)*	
'two'	*loloi*		*meih*
'three'	*lolitu*		*itu*
'four'		*fat*	
'five'		*lima*	
'six'		*douh*	
'seven'		*fitu*	
'eight'		*afo*	
'nine'		*siwa*	
'ten'		*ru(ru)*	
'many'	*roual*		*rial*

Table 20. Makasae numeral agreement

	NON-HUMAN		HUMAN
'one'		*u*	
'two'	*lolaʔe*		*mahe*
'three'	*lolitu*		*mitu*
'four'	*loloha*		*pae*
'five'		*lima*	
'six'		*daho*	
'seven'		*pitu*	
'eight'		*apo*	
'nine'		*siwa*	
'ten'		*ruru (u)*	
'many'		*baun*	

Fataluku also exhibits animacy based agreement on the numerals for 'two' and 'three' (Table 21). These numerals can be marked with *-afu* when reference is to non-humans and *-tere* when reference is to humans (see Fataluku plural markers

Table 21. Fataluku numeral agreement

	UNMARKED	NON-HUMAN	HUMAN
'one'			*ukani*
'two'	*etse*	*etsafu*	*etsatere*
'three'	*utuʔe*	*utuʔafu*	*utuʔatere*
'four'		*fate*	
'five'		*lime*	
'six'		*neme*	
'seven'		*fitu*	
'eight'		*kafa*	
'nine'		*siva*	
'ten'		*taʔane*	

discussed in Section 5.4). Unlike Makalero and Makasae, Fataluku also has unmarked numerals for these values which can be used for any referent.[16]

Bunaq has a gender system based on a two-way distinction of INANIMATE versus ANIMATE. Gender is a covert property of Bunaq nouns that is reflected on two agreement targets, determiners and 3rd person prefixes on verbs. Determiners in Bunaq must agree in animacy with the head noun of the NP, and each determiner has both an INANIMATE and ANIMATE agreement form (Table 22). On verbs, Bunaq displays differential marking of Ps based on animacy. While INANIMATE Ps are unmarked by a verbal prefix, animate Ps are prefixed on the verb with *g-* '3AN-' (see Table 14 in Section 5.1 for the verbal conjugations which complicate this basic agreement system).

Table 22. Animacy agreement on Bunaq determiners

	INANIMATE	ANIMATE
Definite article	*ba*	*bi*
Proximal demonstrative	*bare*	*bari*
Non-proximal demonstrative	*baʔa*	*baʔi*
Specifier demonstrative	*doe*	*doi*
Contrastive demonstrative	*homo*	*himo*
Counter-expectational demonstrative	*bere*	*beri*

---

16. The Fataluku system looks like a reduction of a previously more productive system that has broken down. De Josselin de Jong (1937: 195–197) describes the numeral system of the closely related Oirata language. In this work we see that numerals above one have two forms, one with *-een* and one without. He does not identify any animacy difference between the two forms, though it is notable that numerals marked with *-een* only appear with human nouns in the examples provided. He also notes that the form *-apu* is also found with numerals in Oirata, where it means 'all', likely related to Fataluku *-afu*.

The Bunaq animacy agreement system is described here as a gender system because animacy is a grammatical rather than semantic property of nouns in Bunaq. That is, although the agreement system has a strong semantic basis, it is not sufficient to know the meaning of a noun in order to determine what agreement form it will take. Whilst all nouns denoting animates take ANIMATE agreement, not all nouns denoting inanimates take INANIMATE agreement. For example, *zap* 'dog' in (19a) is determined by the ANIMATE form of the definite article and takes the 3rd person ANIMATE agreement prefix *g-* on the verb *teke?* 'look at'. The INANIMATE noun *zo* 'mango' in (19b) is determined by the ANIMATE form of the definite article and does not agree on the verb. By contrast, the ANIMATE noun *pa?ol* 'maize' in (19c) has a plant referent but takes the ANIMATE agreement forms that we saw with *zap* 'dog' in (19a).

Bunaq gender agreement

(19)   Animate referent with ANIMATE agreement
   a.   *neto **zap bi**   **ge**-teke?*
       1SG dog DEF.AN 3AN-look.at
       'I'm looking at the dog.'

   Inanimate referent with INANIMATE agreement
   b.   *neto **zo**   **ba**   teke?*
       1SG mango DEF.INAN look.at
       'I'm looking at the mango.'

   Inanimate referent with ANIMATE agreement
   c.   *neto **pa?ol bi**   **ge**-teke?*
       1SG maize DEF.AN 3AN-look.at
       'I'm looking at the maize.'

In sum, animacy plays a role in agreement across the Papuan languages of Timor. Contrastive numerals for human versus nonhuman referents appear irregularly in the Eastern Timor languages. The grammatical gender system of Bunaq with its two-way agreement contrast qualifies in McWhorter's terms as moderately complex.

5.3   Locative and applicative prefixes

Unlike in other TAP languages, verb serialisation is quite limited in Eastern Timor Papuan languages. While other TAP languages have serialisation, similar functions in the Eastern Timor languages are fulfilled by what has variously been described as verbal prefixation, verb compounding and incorporation into a preverbal slot.

    Verbal prefixes derived from verbs typically have a dedicated truncated form in Eastern Timor Papuan languages. We can see this variation between full verb

and verb prefix particularly well in Makasae, a language where the move from verb serialisation to verb prefixation is not entirely complete. Comparison in Makasae can be expressed in two ways: (i) the verb *litaka* 'pass, surpass' introduces the standard of comparison in serialisation with a stative property verb such as *rau* 'good' (20a), or; (ii) the verbal prefix *lita-*, obviously related to the verb *litaka*, introduces the standard as the applied object of the stative property verb (20b). By contrast, Makalero has almost no verb serialisation and its comparatives are formed exclusively by prefixation (Schapper & de Vries 2018).

Makasae

(20)   Serialised exceed comparative

    a.   *fi   welafu ehani **rau** fi   boba laneʔe gigeʔe **litaka***
       1PL.INCL life   now   good 1PL.INCL father PL   POSS   EXCEED
       'Our lives nowadays are better than our parents' lives.' (lit. 'Our lives now are good exceeding those of our parents')

    Incorporated/prefixed exceed comparative

    b.   *fi   welafu ehani fi   boba laneʔe gigeʔe **lita-rau***
       1PL.INCL life   now 1PL.INCL father PL   POSS   EXCEED-good
       'Our lives nowadays are better than our parents' lives.'

                                        (Correia 2011: 318)

The move away from serialisation to prefixation has meant that the Eastern Timor Papuan languages have developed large inventories of verb prefixes, unparalleled in their relatives. Table 23 sets out a small number of the locative prefixes that are found on verbs in Fataluku. See Section 5.4 for the related issue involving the prefixation of locative verbs marked by *-ne* and the initial consonant mutations on verbs described in the following.

Table 23. Examples of Fataluku (Eastern) prefixes and related *-ne* marked verbs

	Prefixal form	Verbal form		Prefixal form	Verbal form
'on'	*a-*	*ane*	'at'	*metse-*	*metsene*
'reach'	*atsa-*	*atsane*	'on'	*mitsa-*	*mitsane*
'cover'	*atsu-*	*atsune*	'near'	*mini-*	*minine*
'aside'	*afa-*	*afane*	'in front'	*mira-*	*mirane*
'amidst'	*apa-*	*apane*	'inside'	*mutsu-*	*mutsune*
'far'	*tso-*	*tsone*	'around'	*poro-*	*porone*
'front'	*fanu-*	*fanune*	'up, upon'	*puhu-*	*puhune*
'up'	*hia-*	*hiane*	'between'	*ulu-*	*ulune*
'with'	*horu-*	*horune*	'backwards'	*uta-*	*utane*
'under	*iti-*	*itine*	'away'	*ura-*	*urane*

Whilst many of the verbal prefixes in Eastern Timor Papuan languages are productive and have semantics transparently related to their source verb, semantic bleaching of verbal prefixes can also be observed. In these situations, lexicalised relationships between verb and verb prefix tend to emerge. For example, the Makasae verbal prefixes *mi-*, *ne-* and *ge-* are considered by Correia (2011) to be derived from the verbs *mini* 'follow', *nehe* 'very, be excessive' and *gehele* 'firm, tight'. As can be seen in Table 24, the combination of verbs with these verb prefixes frequently yields unexpected semantics.[17]

Table 24. Examples of Makasae semantically bleached verbal prefixes

mi- **prefix**			
fusa	'peep'	*mifusa*	'spy'
gamunu	'hold'	*migamu*	'feel, grope'
laʔa	'go, walk'	*milaʔa*	'follow'
loʔi	'wipe, clean'	*miloʔi*	'anoint'
maʔene	'know'	*mimaʔene*	'recognise'
saga	'look for'	*misaga*	'search'
suri	'let go, set free'	*misuri*	'follow in numbers'
tamunu	'mention'	*mitamu*	'name after'
ne- **prefix**			
akasa	'try, attempt'	*neakasa*	'try hard'
daʔiri	'flatter, praise'	*nedaʔiri*	'flatter, praise greatly'
gini	'do, make'	*negini*	'treat really badly'
guta	'kill, slaughter'	*neguta*	'beat badly'
lolo	'tell, say'	*nelolo*	'scold excessively'
ge- **prefix**			
base	'hit, strike'	*gebase*	'hammer'
diʔara	'sit'	*gediʔara*	'establish oneself'
booro	'tie up'	*gebooro*	'tighten'
gesi	'close'	*gegesi*	'lock, bolt'
koʔolo	'hug'	*gekoʔolo*	'embrace, include'
sifa	'hold, drive, catch'	*gesifa*	'arrest'
siʔili	'bind, tie'	*gesiʔili*	'fasten, secure'

---

17. In some cases, it would perhaps be more accurate to say that it is not always clear from Correia's (2011) glosses what the semantic difference between prefixed and unprefixed forms are.

5.4   Initial verb root mutations

The Eastern Timor Papuan languages collectively display irregular initial consonant mutations on verbs that are triggered by what, as we saw in the previous section, has been variously described as prefixation or incorporation into a preverbal slot. For example, Oirata has two patterns of initial consonant mutations: *t > -r* (e.g., *tipare* 'run, flee' > *ura-**ripare*** 'go back, run back'), and *p > -h*, (e.g., *pai* 'make, do' > *ura-**hai*** 'open, uncover', lit. do back). These mutations are not found on all verbs with the appropriate initial segments; rather they are irregular, non-predictable changes that occur to a verb on prefixation of particular morphemes.

The patterns of initial consonant mutation and the set of verbs subject to them varies from language to language. In Makasae, only around a dozen verbs mutate following two patterns *s > -d* and *t > -d* (Table 25).[18] In Makalero, the number is larger with around 40 verbs showing consonant mutation following two patterns *h > -s* and *t > -d* (Table 26).

Table 25. Makasae verbs with initial consonant mutations

	Free	Bound			Free	Bound	
s > -d	*sege*	*-dege*	'difficult'	*t > -d*	*tamu*	*-damu*	'name'
	seʔele	*-deʔele*	'jump'		*taru(nu)*	*-daru*	'put, place, bury'
	seriki	*-deriki*	'tie down'		*taʔe*	*-daʔe*	'sleep'
	sesara	*-desara*	'fall, throw down'		*tapuru*	*-dapuru*	'cooked, done'
	sipa	*-dipa*	'catch'		*tia*	*-dia*	'bite'
	suri	*-duri*	'shoot'				
	sisir	*-disir*	'sick'				
	supa	*-dupa*	'spit'				

Table 26. Makalero verbs with initial consonant mutations

	Free	Bound			Free	Bound	
h > -s	*haiʔ*	*-saiʔ*	'finished'	*t > -d*	*tafal*	*-dafal*	'throw away'
	haka	*-saka*	'search'		*taka*	*-daka*	'close'
	hat	*-sat*	'dry'		*tamu*	*-damu*	'name'
	haʔal	*-saʔal*	'fry'		*taru*	*-daru*	'put, place'
	heil	*-seil*	'pull'		*tekih*	*-dekih*	'lean towards'
	hein	*-sein*	'wait'		*teuh*	*-deuh*	'buy'
	heti	*-seti*	'ask'		*teri*	*-deri*	'cut'

(continued)

---

18. Correia (2011: 216) states that in several cases free and bound forms alternate in the same context, suggesting that initial consonant mutations are continuing to break down in Makasae.

Table 26. (*continued*)

Free	Bound		Free	Bound	
heke	*-seke*	'difficult'	*teruʔ*	*-deruʔ*	'shelter'
helar	*-selar*	'big.PL'	*tia*	*-dia*	'sleep'
heman	*-seman*	'take'	*tina*	*-dina*	'cook'
heʔel	*-seʔel*	'jump'	*tiʔal*	*-diʔal*	'kick'
heʔi	*-seʔi*	'cut'	*tiʔ*	*-diʔ*	'pour (liquid)'
hifaʔ	*-sifaʔ*	'catch'	*toʔi*	*-doʔi*	'dig'
hofe	*-sofe*	'know'	*tufa*	*-dufa*	'sweep'
hor	*-sor*	'protect'	*tuku*	*-duku*	'punch'
houn	*-soun*	'plant'	*tula*	*-dula*	'bring, transport'
huma	*-suma*	'angry'	*tule*	*-dule*	'not want'
huri	*-suri*	'shoot, release'	*tupi*	*-dupi*	'pound, thump'
			tuʔil	*-duʔil*	'cook in bamboo'

Fataluku has a similarly sized set of irregular mutating verbs to Makalero, but with more patterns of mutation attested. There are three patterns of initial consonant mutation attested over numerous verbs: *f > -p*, *t > -ts*, and *h > -ts*, as illustrated in Table 27. A fourth pattern, *s > -ts*, is known from one verb *sil > -tsil* 'bind'.

Table 27. Examples of Fataluku (East) initial mutations

	Free	Bound			Free	Bound	
f > -p	*fulu*	*-pulu*	'spit'	*t > -ts*	*tipal*	*-tsipal*	'drum'
	fal	*-pal*	'grab'		*tutef*	*-tsutef*	'blow'
	fetil	*-petil*	'stumble'		*te*	*-tse*	'measure'
	fer	*-per*	'wipe'		*to*	*-tso*	'inside.SG'
	fo	*-po*	'inside.PL'		*taja*	*-tsaja*	'sleep'
	fot	*-pot*	'cut up'		*teku*	*-tseku*	'stir'
	fai	*-pai*	'do'		*tih*	*-tsih*	'cook again'
h > -ts	*hina*	*-tsina*	'plait'		*tomok*	*-tsomok*	'be soft'
	hura	*-tsura*	'spoon'		*tu*	*-tsu*	'feed'
	hit	*-tsir*	'wait'		*tul*	*-tsul*	'sick, lazy'
	here	*-tsere*	'dry'				
	ha	*-tsa*	'warp'				
	huleve	*-tsuleve*	'plant'				

These consonant mutations go back to a locative morpheme **n-* that can still be observed on some vowel initial verbs under the same conditions as the initial consonant mutations. Both Oirata and Fataluku have a sizeable number of (often cognate) verbs that appear with *n-* when prefixed, e.g., Oirata *asi* 'see, look' > *ura-**nasi*** 'look back' and *ihile* 'fly' > *ura-**nihile*** 'fly back', Fataluku *atsi* 'see' > *mucu-**natsi*** 'look

inside', *ipile* 'fly' > *mucu-**nipile*** 'fly inside'. Makalero preserves this *n*- on a small number of verbs also: *umu* 'die' > *afa-**numu*** 'die leaving (so.) behind' (cf. *k-umu* 3-die 'kill so.'), *uta* 'fall (of rain)' > *isi-**nuta*** 'fall on (of rain)', *k-ini* 'make' > *ata-**nini*** 'make in'. The initial consonant mutations seen above are consistent with assimilation to an earlier **n*- prefix that has subsequently been lost except on (some) vowel initial verbs. The mutations most typically involve a change from a voiceless to a voiced consonant at the same place of articulation, a common result of nasalisation.[19] The changes in place of articulation (e.g., *h* > *-s*) are also consistent with the presence of an alveolar nasal.

The reconstruction of **n*- as a locative morpheme is motivated by the fact that the most common trigger for the initial consonant mutations and the *n*- prefix on vowel initial verbs is a prefix with locative semantics. Additional evidence for locative **n*- comes from Bunaq, where a subset of obligatorily possessed nouns show a contrast between locative *n*- and third person agreement marker *g*-.[20] The locative prefix on these items in Bunaq expresses that the referent of the noun it marks has an internal location. For example, on *-iol* 'voice, sound, language', *n*- in (21a) indicates that the noise issues from the engine internal to the motorbike, whereas *g*- in (21b) marks simply that the sound is that of a motorbike, e.g., heard from a distance. Similarly, on *il* 'water', *n*- in (21c) indicates the reference is to the water internal to the nut of the coconut palm, while *g*- in (21d) denotes the fluid circulating through the palm, i.e., its sap.

Bunaq contrast of *n*- and *g*- prefixes on nouns

(21) *n*- form

    a.   *motor*   ***n-iol***
           motorbike LOC-voice
           'internal growl of motorbike engine'

    g- form

    b.   *motor*   ***g-iol***
           motorbike 3-voice
           'motorbike's sound'

---

19. Note that Fataluku /ʦ/ corresponds to Makalero/Makasae /d/ and appears to represent a sound change from an earlier voiced consonant reconstructed by Schapper et al. (2014) as **D*.

20. While it may seem far-fetched to relate an *n*- prefix on nouns in one language with one found on verbs in other languages, recall that in Bunaq, as in other Timor-Alor-Pantar languages, the same paradigm of agreement prefixes on both nouns and verbs. Fataluku also has an *n*- prefix on nouns marking a 3rd person inalienable possessor. Because the semantic relationship between locative and inalienable is not clear to me, I make no claim about whether this *n*- is related to the locative *n*-.

    n- form

c.   *hoza*   **n-il**

    coconut ᴸᴼᶜ-water

    'water contained within a coconut'

    g- form

d.   *hoza*   **g-il**

    coconut 3-water

    'sap of a coconut tree'

Here again we have a situation of prefixes that have been accreted and created new complexity in the form of irregular mutations on some initial consonants on verbs and the availability of an *n-* prefix on other roots.

5.5   Derivational suffixal morphology

Fataluku and Oirata have the largest array of derivational morphology within the Timor-Alor-Pantar languages. Whilst, as already mentioned, derivational morphology does not hold the same status as inflectional morphology for McWhorter (2008: 18–20), the extent of derivational morphology in some of the Papuan languages of Timor is noteworthy.

    Of these, Fataluku has the largest number of documented derivational suffixes, some of which are high-frequency items. Verbs in Fataluku can be nominalised by means of two suffixes: *-n* and *-ana* (or its allomorph *-nana* occurring on vowel-final verbs). The suffix *-n* (often realised as [-nu] at the end of a phonological phrase) only occurs on vowel-final verbs that are not marked with the verbaliser *-e* (examples in Table 28). The nominalising suffix *-ana* can occur on verbs both with and without *-e* (examples in Table 29).[21] These two suffixes can occur on the same verbal root with a semantic difference. Van Engelenhoven (2009b) illustrates the contrast between *-n* and *-ana* with the verb *koso* 'shout': *koso-n* 'shout' refers to the sound produced, while *koso-nana* is an agent nominalisation denoting 'so. shouting/shouter'.

---

21. While van Engelenhoven (2009b) discusses it as a separate morpheme, following Heston (2015), I regard *-ina* as a likely dialect variant of *-ana* chiefly attested in Campagnolo's (1973) material. Van Engelenhoven (2009b) also describes *-ana* as having adjectival functions. This is because property denoting verbs marked with *-nana* can also be used in the following nominal attributive function a N$_{HEAD}$ *i*$_{POSS}$ V-*ana*$_{MOD}$.

Table 28. Examples of Fataluku verbs nominalised with *-n*

Verb		Deverbal noun	
atsi	'see'	*atsin*	'vision'
eru	'lack'	*erun*	'shortage'
kolo	'be mute'	*kolon*	'muteness'
toto	'watch'	*toton*	'view'
lika	'be lean'	*likan*	'leanness'
matse	'eat'	*matsen*	'food'
latsa	'make fence'	*latsan*	'fenced place'

Table 29. Examples of Fataluku verbs nominalised with *-(n)ana*

Verb		Deverbal noun	
atsi	'see'	*atsinana*	'sth. seen, visible'
afile	'slice'	*afilana*	'slice, sth. sliced'
akate	'swollen'	*akatana*	'sth. swollen'
ase	'rough'	*asana*	'sth. rough'
tsulu	'cook'	*tsulunana*	'cooked food'
hoile	'hunt'	*hoilana*	'prey, sth. hunted'
itsane	'fall'	*itsanana*	'sth. falling'
isi	'descend'	*isinana*	'sth. descending'
tahine	'beautiful'	*tahinana*	'so. beautiful'

According to van Engelenhoven and Huber (2020) and Heston (2015), nouns ending in a consonant in Fataluku can be verbalised by means of the suffix *-e*. Examples of such precategorical roots are provided in Table 30.

Table 30. Examples of Fataluku nouns verbalised with *-e* 'VERB'

Noun		Denominal verb	
lapar	'chop'	*lapare*	'chop'
lamak	'crumb'	*lamake*	'crumble'
iʔis	'vomit'	*iʔise*	'vomit'
tsatan	'sign'	*tsatane*	'sign'
laman	'orchard'	*lamane*	'make an orchard'
apat	'tuber'	*apate*	'emerge from soil'
asir	'salt'	*asire*	'put salt on'
kosin	'saddle'	*kosine*	'saddle (horse)'
hit	'hit'	*hite*	'hit with sword'
tupur	'woman'	*tupure*	'be feminine'
inik	'sand'	*inike*	'be sandy'
matar	'stone'	*matare*	'be stony'
lumuk	'mud'	*lumure*	'be muddy'

Alongside verbalising *-e*, Fataluku also has the verbalising suffix *-ne* that occurs on vowel final roots. The final *-ne* suffix is dropped when the item is used prefixally on a predicate. Examples of such items have been given already in Table 23. Van Engelenhoven (2009b) analyses *-ne* as an allomorph of *-e* conditioned by the shape of the root. He explains the appearance of *-e* rather than *-ne* on a small number of apparently vowel final roots as due to the presence of an underlying glottal stop (which is, however, lost in some dialects). By contrast, Heston (2015: 108–109) argues that the glottal stop that appears on suffixation with *-e* is epenthetic (e.g., *utu* 'three' > *utu-e* [utuʔe] 'three-VBLZ', *na-* 'at' > *na-e* [naʔe] 'at-VBLZ') and consequently that the appearance of *-ne* is not morphophonemically predictable. Heston (2015: 24–25) analyses *-ne* as two morphemes *-n* 'LOC' and *-e* 'VBLZ', adducing that the *-n-e* can be replaced with *-p-e* to give a dynamic (motion) reading (e.g., *muʦu-* 'inside', *muʦu-ne* 'be inside', *muʦu-pe* 'go inside'). Whatever the analysis of these suffixes, it is clear that Fataluku makes extensive use of morphological means to derive nouns and verbs (similar morphology is described for Oirata by de Josselin de Jong 1937: 182–183, 188–189).

The other Papuan languages of Timor have far fewer derivational suffixes and they are usually of limited productivity. Makalero, for example, has three derivational suffixes: the first two, *-r* and *-ini,* derive nouns from verbs, illustrated in Tables 31–32.

Table 31. Makalero nominaliser *-ini*

Verb		Noun	
lolo	'say, speak'	*loloini*	'word, conversation'
k-utu	'wear'	*kutuini*	'clothes'
teuh	'buy'	*teuhini*	'gift'
tina	'cook'	*tinaini*	'cooked rice'

Table 32. Makalero nominaliser *-r*

Verb		Noun	
k-ako	'steal'	*akor*	'thief'
nua	'eat'	*nuar*	'food'
umu	'die'	*umur*	'corpse, death'

A further suffix *-ʔ* is used for deriving verbs from nouns, as in the examples in Table 33. This suffix is also found very productively as a verbaliser of verbal prefixes that no longer have an underived verbal form preserved (see Section 5.3).

Table 33. Makalero verbaliser *-ʔ*

Noun		Verb	
uali	'ear'	*ualiʔ*	'hear'
huri	'brush'	*huriʔ*	'brush'
atu	'faeces'	*atuʔ*	'defecate'
teru	'shelter, umbrella'	*teruʔ*	'shelter'

5.6   Morphological and suppletive number marking

The Eastern Timor subgroup of Papuan languages is characterised by lexical classes of verbs and nouns that are marked for number by means of a plural suffix or suppletion. Where a suffix is used to mark plural number, its form is often highly irregular.

Verbal number marking in Eastern Timor languages involves different forms being used depending on whether one or more participants are involved in the action. The simplest systems of verbal number marking are found in Makalero and Makasae. Most verbs in these two languages are invariable, but a small class of intransitive verbs has suppletive forms for singular and plural subjects (Table 34 and 35). The plural forms of these verbs typically end in *-ar ~ -er* (with a few exceptions in each language), but for the most part there is no clear relationship between singular and plural forms of the roots.

Table 34. Makalero suppletive verbs

	SG	PL
'sit'	*mit*	*diar*
'stand'	*nat*	*naser*
'run'	*riaʔ*	*titar*
'lie'	*tia, -dia*	*rou*
'big'	*pere*	*helar*

Table 35. Makasae suppletive verbs

	SG	PL
'sit'	*mi*	*diar*
'stand'	*na*	*nahar*
'run'	*riaʔ*	*ditar*
'be positioned'	*daro*	*doen*
'lie, rest'	*wou*	*rai*
'say'	*lolo*	*lolini*

Fataluku has a much larger set of intransitive verbs marked for singular versus plural. The morphological form the marking takes is unpredictable, as can be seen from the examples in Table 36. The most common form of plural marking is by means of *-re* (on a verb already marked with *-e*) or *-ere*, but many other forms such as *-tere, -care, -oro* are also found. In addition, a subset of these verbs, mostly posture verbs, has suppletive roots for singular and plural. However, as in Makalero and Makasae, the plural is still typically identifiable by an *-r*.

Table 36. Fataluku verbal number marking and suppletion

	Singular subject	Plural subject		Singular subject	Plural subject
'fly'	*ipile*	*ipilere*	'dead, die'	*umu*	*umunoro*
'laugh'	*kele*	*kelere*	'live'	*lauhe*	*lauhoro*
'eat'(intr)	*matse*	*matsere*	'slip, be born'	*suke*	*sukoro*
'be located at'	*nae*	*naere*	'stand upright'	*tsumai*	*tsutoro*
'be placed'	*hitsine*	*hitsinere*	'be inside'	*otoe*	*ofoe*
'disappear'	*molu*	*molure*	'hide' (intr)	*palake*	*pelere*
'full'	*polu*	*polure*	'hang'	*vaiake*	*verire*
'come'	*mau*	*mauere*	'stand'	*nate*	*nehere*
'sleep'	*taia*	*taiatere*	'run, flee'	*tifare*	*helere*
'big'	*lafai*	*lafitsare*	'sit'	*mire*	*tsuare*
'high, long'	*lohai*	*lohitsare*	'lay, lie'	*laku*	*tepere*

De Josselin de Jong's (1937) Oirata materials make it clear that singular ~ plural verbal number marking also exists in that language. Like the other languages of the Eastern Timor group, Oirata has both suppletive (e.g., *mire / rua* 'sit.SG/PL') and suffixal marking with *-(e)re* (e.g., *naaje / naajere* 'swim.SG/PL') for the number of the subject of an intransitive verb. In addition, Oirata uses *-(e)re* on transitive verbs to mark the plural number of the subject. Illustration of the number contrast is given in (22).

Oirata plural suffixation of transitive verbs

(22) Singular subject
  a. *ue in-asi*
      3SG 1PL.EXCL-see
      'He sees us.'

  Plural subject
  b. *ite in-asi-ere*
      2PL 1PL.EXCL-see-PL
      'You see us.'

Alongside plural number marking on verbs, Eastern Timor languages each have a lexically specified class of nouns denoting humans that take a plural suffix. Plural marking is not obligatory in plural reference in these languages, but it is frequent on this class of human nouns. Examples of members of this special plural marked class are given for Oirata in Table 37. The Oirata plural suffix -*ra* has the allomorph -*a* that appears on nouns which end with a final /r/, as can be seen on the form *tuhur*. There is also an irregular plural form *namirara* derived from *namirai* (cf. Fataluku cognate irregular form in Table 39)

Table 37. Examples of the Oirata restricted plural marked class

	SG	PL
'younger brother'	*noo*	*noo***ra**
'elder brother'	*kaka*	*kaka***ra**
'sister'	*leren*	*leren***ra**
'friend'	*hele*	*hele***ra**
'child'	*modo*	*modo***ra**
'daughter'	*modo tuhur*	*modo***ra** *tuhur***a**
'son'	*modo nami*†	*modo***ra** *nami***ra**
'husband'	*namirai*†	*namira***ra**

† *nami* as an independent noun means 'man'.

Makalero has a similar lexical class of human nouns marked with a cognate suffix -*raa*. Examples are given in Table 38. This lexically restricted plural suffix exists in Makalero alongside -*laa*, a plural suffix which can mark any noun, though typically it is only used with humans. Makalero also has a dedicated associative plural suffix -*ara* that occurs on personal names.

Table 38. Examples of the Makalero restricted plural marked class

	SG	PL
'younger sibling'	*noko*	*noko***raa**
'elder sibling'	*nana*	*nana***raa**
'sister-in-law'	*ue*	*ue***raa**
'sibling-in-law'	*mali*	*mali***raa**
'friend'	*pada*	*pada***raa**
'child'	*mata*	*matar* ~ *mata***raa** †
'father'	*upa*	*upa***raa**
'uncle'	*tiu*	*tiu***raa**
'parent-in-law'	*paakin*	*paaki***raa**
'grandparent'	*dada*	*dada***raa**
'widow'	*paardufu*	*paardufu***raa**

† Huber (2011: 118) gives several permutations of the Makalero plural for 'children', including *mata-niki* 'children' existing alongside these forms. In Makasae one plural marked noun has been retained but in an irregular form: *mata* 'child' and *mata-rini* 'children' (Correia 2011) or *mata-kini* (Huber 2008).

Fataluku has a more complex system of plural marking, with multiple small lexical classes of nouns referring to humans. The classes are defined by three different plural suffixes: *-r*, *-ra* and *-afu*. Table 39 presents examples of plural suffix taking nouns. We see here that there are also several irregular forms such as those with *-(r)ara* for 'man' and 'woman'. These restricted plural suffixes exist alongside an enclitic *=ere* that can occur on any noun, including on a noun already marked with a plural suffix.

Table 39. Examples of Fataluku restricted plural classes of nouns

	SG	**PL**
'father's sister'	*tamu*	*tamur*
'younger sibling'	*noko*	*nokor*
'older sibling'	*kaka*	*kakar*
'child'	*motso*	*motsor*
'child'	*moko*	*mokor*
'master, owner'	*otsava*	*otsavar*
'friend'	*lanu*	*lanura*
'parent-in-law'	*painu*	*painura*
'sister'	*lerenu*	*lerenura*
'husband'	*elehu*	*elehura*
'wife'	*jeu*	*jeura*
'man'	*nami*	*namira ~ namirara*
'woman'	*tupur*	*tupurara*
'ancestor'	*tsal*	*tsalafu*
'mother'	*nal*	*nalafu*
'father'	*pal*	*palafu*
'person'	*mar*	*marafu*

Again, the Papuan languages of Timor have complexity in the domain of plural marking. We observed that McWhorter's complex features of suppletion, lexicalisation and irregularity all play a role in the expression of plurality across nouns and verbs in these languages.

6. The comparative picture of complexity in the languages of Timor and surrounds

In the preceding sections, the languages of eastern Timor, both Austronesian and Papuan, have been shown to exhibit a range of complexity in the form of morphology, both productive and fossilised. In arguing for his view of complexity, McWhorter stresses that the comparative benchmark for interpreting the degree of

complexity displayed by a language is other languages of the same family. He hypothesises that "no language without significant non-native acquisition in its history will be especially simplified in comparison with its sisters" and, relatedly, "all languages with significant non-native acquisition in their histories will be significantly less complex according to my metric than their sisters" (2007: 268). In McWhorter's view, variation in complexity will be observed amongst sister languages only within a minor range. In this section, therefore, we will take a comparative look at what the languages of eastern Timor look like in contrast to their relatives in the area.

When compared to their relatives in the region, the Austronesian languages of eastern Timor are morphologically somewhat reduced, but not radically so. Reductions are typically limited to small sets of languages and to particular types of morphology. For example, eastern Timor Austronesian languages for the most part have very similar sets of subject prefixes to the "complex" ones that McWhorter (2007: 244–246) points to in languages of western Timor, Rotinese and Uab Meto. The complete loss of inflectional prefixes is found only in the neighbouring languages of Tokodede and Kemak as well as in the Kawaimina languages. Between these two groups of languages, we see an assortment of subject prefixes that are not strikingly smaller in distinct number of prefixes than West Timor (Map 2). The loss of subject prefixes here has not been abrupt as McWhorter claims. Simplification of prefixes from CV- to simply C- resulted in initial consonant clusters on consonant-initial verbs. A dispreference for these has, in turn, led to the progressive reduction of subject prefixes. This is seen most pointedly in the fact that subject prefixes tend to be more intact on vowel-initial verbs in eastern Timor languages. At the same time,

Map 2. Number of verbal subject prefixes per paradigm
in Austronesian languages of Timor

we see that the innovative agreement prefixes on numerals that have retained their CV-shape are not lost in even the highly isolating Kawaimina languages.

Possessive suffixes are in some Austronesian languages of eastern Timor noticeably absent, but it is not the case that these languages have fewer suffixes in general. We see in Map 3 that some eastern Timor Austronesian languages have the same number of possessive affixes as Uab Meto in the west. It is accurate to say that the typical eastern Austronesian pattern of suffixes marking inalienable possessors and free elements marking inalienable possessors (see Donohue and Schapper 2008) has broken down in some languages with the result that fewer affixes are involved in the coding of possessors. Systems of possessive classification have been reformed in different ways, with possessive suffixes and free possessor markers entering into paradigmatic relationships in which factors such as phonological shape of the root and person determine the strategy of possessive marking, not simply inalienability. At the same time, a language like Kemak preserves the original alienability split and the associated possessive suffixes, while Dadu'a has developed new prefixal paradigm encoding of possessors. This diverse coding of possessors in the eastern Timor Austronesian languages points to a cycle of morphological loss and creation that McWhorter sees as part of normal language change.

Map 3. Number of possessive suffixes per paradigm in Austronesian languages of Timor

Synchronic metathesis is a notable morphosyntactic feature of the Austronesian languages Uab Meto, spoken in West Timor, and Leti, spoken on an island off the eastern tip of Timor (see Schapper 2015: 137–138 on its areal distribution). In both these languages, metathesis is conditioned by an intricate set of pragmatic and grammatical contexts, and affects a large portion of the lexicon (see Steinhauer 1993, 1996; Edwards 2016; van Engelenhoven 2004). In McWhorter's (2007: 245)

terms, metathesis is a complex feature that requires significant learning of morphosyntactic and morphophonological rules. Whilst synchronic metathesis is not as pervasive in the central Timor languages Mambae and Kemak as it is in Uab Meto and Leti, the fact that it is found with a large number of basic vocabulary items contradicts the idea that these languages are 'undressed'. The existence of a typological rarity such as synchronic metathesis in their grammars rather suggests that the significant erosion of inflectional morphology in the central Timorese languages is not the result of any unnatural interruption to language transmission.

McWhorter (2007: 244) admits that the derivational prefixes of Tetun are similar to those found in Rotinese. We saw in Section 4.2 that, whilst not all Austronesian languages in eastern Timor have the same amount of derivational morphology as Tetun, derivational prefixes are not unusual in the languages, in either productive or fossilised form. What is more, in those languages where derivational prefixes are not productive, they have been accreted and, in several cases, have given rise to new phonemes. The most striking example of this is the Kawaimina languages where the pressure to rid the languages of consonant clusters have caused derivational prefixes to be absorbed to create new consonant phonemes, making one of the largest consonant inventories of the Austronesian family. Crucially, even when fossilised, these prefixes have not been 'shed' from the languages, but have created new complexity in other domains of linguistic structure.

Turning to the Papuan languages, we also do not find a significant difference in complexity between the Papuan languages of Timor and their sisters of the Alor-Pantar subgroup. PTAP only had a single paradigm of inflectional prefixes for person/number, reflexive and reciprocal occurring on both nouns and verbs (Schapper, Huber and van Engelenhoven 2012). This system is still found in the languages of Pantar, but has been expanded in the languages of Alor through the morphologisation of originally free pronominal forms with the shape CV as additional agreement paradigms. Bunaq continues the PTAP system of a single paradigm of agreement prefixes on nouns and verbs, though due to vowel harmony, the number of distinctions marked by prefixal vowels has been lost. In addition, we saw in Section 5.2 that verb prefixation in Bunaq is associated with numerous irregular changes of verb forms and with arbitrary conjugation classes. The agreement system of the Eastern Timor subgroup of Papuan languages shows reduction, with only the PTAP third person prefix having reflexes. The erratic appearance of reflexes, for example, on vowel-initial nouns and verbs in Makasae and Makalero or as an initial vowel of unpredictable shape in Fataluku represents complexity in the form of irregularities that are not found in Alor-Pantar languages.

Whilst Alor-Pantar languages do have more inflectional prefixes than are found in the Papuan languages of Timor, they completely lack the verbal number suppletion or suffixal inflection for plurality found in the Papuan languages of the

Eastern Timor group. Although unique to this TAP subgroup, there is good reason not to consider the feature to be new. Papuan languages on New Guinea, where the ancestor of TAP languages are thought to have originated, frequently have small classes of kin terms that are marked for plural number and of verbs that are suppletive for number, just as in the Eastern Timor Papuan languages. We also find these patterns in the West Bomberai languages (Iha, Mbaham and Kalamang), the relatives of TAP languages on New Guinea. For example, Mbaham has at least two nouns that are marked for number (*namiha* 'man.SG' / *namiata* 'man.PL', *tumbuhar* 'woman.SG' / *tumbuota* 'woman.PL'), while Kalamang has a suffix *-mur* only found on kin terms (Eline Visser p.c.). Verbs suppletive for number have also been attested: *preh-* 'speak.SG' and *ngmbeh-* 'speak.PL' and *tomot-* 'sleep.SG', *tidlh-* 'sleep. PL' in Iha (Donohue 2015), and *wes* 'go.SG', *wuru* 'go.PL', and *mehena* 'sit.SG', *ndigi* 'sit.PL' in Mbaham (Flassy, Ruhukael and Rumbrawer 1984). Whilst none of these are cognate with the suppletive verbs in Eastern Timor languages, we do have what appears to be a cognate of the PET verbal plural marker *-r 'PL' in Kalamang. Here we find optional plural marking by means of a suffix *-r* on two verbs, *melelu* 'sit', *melelur* 'sit.PL', and *na* 'eat/drink', *nar* 'eat/drink.PL' (Eline Visser p.c.). The fact that these morphological patterns in the Eastern Timor subgroup are consistent with patterns of the New Guinea mainland strongly points to their being a retention of an old pattern that is lost in other members of the family.

Some other morphological complexities present in the Papuan languages of Timor are absent in the Alor-Pantar languages and many of them appear to be innovations. The animacy-based gender system of Bunaq is unparalleled among TAP languages (see Schapper 2010b for a description of the wider pattern). The locative and applicative prefixation of the Eastern Timor Papuan languages, particularly in conjunction with the initial root mutations triggered by it, is more elaborate and complex than in any other family members. Alor-Pantar languages often have between one and three locative applicative prefixes, but these are comparatively limited in number and frequency (Willemsen 2015). Derivational suffixes are also more extensive in Fataluku and Oirata than in the other TAP languages. In Alor-Pantar languages, there is some derivational morphology, most notably *-naŋ* in Blagar (Steinhauer 2014: 161–163), but it is not widespread.

On proper consideration of a full data set from the languages of eastern Timor, we find that McWhorter's complexity features of overspecification, structural elaboration and irregularity are well-represented in the morphological properties of the languages. While the eastern Timor languages are sometimes, but certainly not always, morphologically simpler than their nearest relatives, they are by no means without morphology. Indeed, far from 'shedding' or 'dropping', in McWhorter's terms, all their morphology, these languages have significant amounts of accreted morphology that, through lexicalisation and fossilisation, have given rise to

complexities. What is more, the morphological reductions which have undoubtedly occurred in both the Austronesian and Papuan languages of eastern Timor are not the result of drastic stripping, but are explainable with normal processes of progressive phonological and morphosyntactic change.

Finally, morphology and its various manifestations are just one kind of complexity according to McWhorter's metric. It is beyond the scope of this paper to discuss other kinds of complexity, but it is worth pointing out that other complexities are by no means absent in eastern Timor. For example, having more than a two-way demonstrative contrast is complex for McWhorter. In this respect, even the most isolating Timorese languages can be regarded as displaying significant complexity: Naueti has three-way distance marked demonstratives plus a fourth non-familiar demonstrative (Veloso 2016: 47–48); Makalero has a five-way demonstrative distinction involving distance and elevation (Huber 2011: 232–233). In McWhorter's model such features, alongside the morphological complexities discussed at length in the preceding sections, must be taken to reflect the 'oldness' of the languages.

7.   Discussion

In concluding his treatment of the languages of eastern Timor, McWhorter (2007) explicitly deals with potential alternative hypotheses for the isolating word structure that he encounters there. He writes:

> Of course, we might attempt to preserve the chance account in the Timorese […] languages by supposing that the Papuan-related languages developed in this direction because of contact with the reduced Timorese languages. But this would ascribe a highly unusual degree of structural loss to the contact in question. To be sure, Sprachbund effects can occasion loss modeled on one or more of the languages in contact, […], or the loss of concordial morphemes […]. But in all of these cases, a great deal of elaboration was retained: the Balkan languages remain highly inflected, as does Gurnu Baagandji, and the languages of Java retain their derivational morphology and certain inflections. But in the case of the Papuan languages of Timor, we must explain why these four languages have not just lost some inflections of a particular kind, but why they lost so much morphology overall that they stand as strangely analytic, or analytic leaning…   (McWhorter 2007: 248)

As we have seen throughout this paper, the languages of eastern Timor, whether Austronesian or Papuan, are not unnaturally reduced. That is, affixes have not simply been 'shed' from these languages, but have been retained productively and in various reduced and lexicalised forms. I have put it that, contrary to McWhorter's (2007, 2008) claims, the significant amount of accreted morphological complexity that can be observed provides ample evidence of the 'oldness' of the languages of

eastern Timor. McWhorter's claims can only be made because of a massively facile treatment of the data. McWhorter betrays this perhaps most pointedly of all in the following passage, in which he presents the one and only data point from a Timorese Papuan language in all his writings on Timor (McWhorter 2008: 179): "In contrast to typically inflectional Papuan languages is a Fataluku sentence such as *Ana merkadu mara* (I market go) 'I'm going to the market' […]. Behold a Papuan language with the typology of Chinese." While the rhetoric here is appealing, the typology of a language cannot be captured in any accurate way with a single clause. It hardly bears saying that careful examination of the facts of whole language systems is required to sustain arguments about morphological profiles and their historical origins.[22]

Nonetheless, I do not dispute that the languages of eastern Timor do lean towards having an isolating typology overall. What is the reason for this shared characteristic? In eastern Timor languages, morphological reduction has almost certainly gone hand in hand with the phonological changes that we have seen, namely, vowel harmony in unstressed affixal vowels, loss of unstressed vowels in prefixes and in turn the reduction of consonant clusters.[23] The fact that these processes can be observed across Papuan and Austronesian languages in eastern Timor points to just the mechanism that McWhorter himself notes in the above quote. That is, phonological Sprachbund effects between the languages of eastern Timor have occasioned morphological loss. The idea of a Sprachbund centred on eastern Timor is not new. Numerous features such as the absence of a velar nasal phoneme, flexible genitives (i.e., both GEN N and N GEN word orders), nouns for 'face' and 'name' being used as quantifiers of kinds have been noted, among others (Schapper 2011, 2015; Hull 2001). The move towards isolating word structure is adequately

---

22. One reviewer suggests that given the materials available at the time of writing, McWhorter's assumption that the Papuan languages of Timor were isolating was not unreasonable. I do not disagree with describing the overall typology of these Papuan languages as isolating(-leaning). However, that is quite different from the position taken by McWhorter in saying that the languages are morphologically stripped and lacking any features typical of 'old' languages. What is more, some of the morphological patterns which I describe here were indeed present in the early sketches that would have been available to McWhorter, e.g., Fataluku plural inflection on kin terms and derivational suffixes are described in the early sketch of Hull (2005). A lack of engagement with the existing sources and a failure to exercise caution in making, particularly historical, arguments about poorly described languages is indefensible.

23. A further idea to be investigated is that prosodic patterns also contributed to morphological loss. Himmelmann (2010) describes Waima'a as not making phonological use of pitch changes (i.e. lacking lexical tone distinctions as well as postlexical pitch accents), but rather as having phrasal accent on the penultimate syllable. It may be that this prosodic type contributes to phonological erosion at the end of the word.

provided for by convergence of the concentric circles of linguistic isoglosses around Timor. In short, not only is the supporting data not to be found in the languages themselves, but there is no need for recourse to a hypothesis involving radical simplification due to 'heavy non-native acquisition' in relatively recent times to explain the existence of isolating structure in eastern Timor.

Yet, language shift or significant non-native acquisition must have been a factor in the Timor region. Both the Austronesian and Papuan languages of the area are the result of migrations, the former out of Taiwan (Bellwood 1997) and the latter out of New Guinea (Hull 2004; Ross 2005; Schapper 2017; Usher & Schapper ms). The relative timing of the arrival of these two different groups in the Timor area is not clear, but the evidence, such as it is, suggests that they were not separated by a large period of time, both events occurring somewhere in the mid to late Holocene (see Schapper 2015: 141–142 for a summary). But settlement in the region is much older; Timor was populated by modern humans since at least 40,000BP (Hawkins et al. 2017). Genetic studies of human populations in eastern Indonesia show that the dispersal of the Austronesian and TAP language groups was not associated with thorough-going displacement of pre-existing populations (e.g., Richards, Oppenheimer and Sykes 1998; Mona et al. 2007, 2009; Tumonggor et al. 2014; Gomes et al. 2017). As such, the languages of the earliest peoples are likely to have had an impact on those of the incoming groups.

The deeper linguistic prehistory of Timor is, therefore, interesting for understanding the formation of the morphological typology of the languages today. If we compare the morphological properties of Proto-Timor-Alor-Pantar (PTAP), the common ancestor of the Papuan TAP languages, to its relatives on New Guinea, some notable differences in morphology are apparent. In PTAP, case-marking suffixes on nouns and portmanteau person/TAM marking suffixes on verbs are entirely absent, while they are regular features characterising both PTAP's nearest relatives on West Bomberai and Trans-New Guinea languages in general. If we assume such suffixes can be reconstructed to a higher level (cf. the suffixal plural discussion in Section 6), their loss in PTAP or its predecessor shows a drastic move away from a suffixing inflectional type. At the same time, PTAP retains reflexes of the TNG verbal agreement prefixes, but whereas PTNG verbal agreement prefixes marked P (Suter 2012), PTAP verbal agreement prefixes appear to have marked not only P, but also non-active S. These differences between PTAP and its relatives are consistent with PTAP intruding into an area where, as established in the literature, weakly prefixing-to-little inflectional morphology (Gil 2015) and split-S alignment (Donohue 2004) was the norm. Thus, this observation of the profile change of PTAP would add to the wealth of evidence presented in this volume for the ancient origins of isolating word structure across Indonesia.

Acknowledgements

References

Alcantara, Maressa Xavier. 2015. Descrição fonética e fonológica da língua Idaté do Timor Leste. MA thesis, Universidade de São Paulo.

Bellwood, Peter. 1997. *Prehistory of the Indo-Malaysian Archipelago*. Honolulu HI: University of Hawai'i Press.

Bickel, Balthasar & Nichols, Johanna. 2013. Obligatory possessive inflection. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 58. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/58> (1 June 2020).

Blust, Robert. 2008. Is there a Bima-Sumba subgroup? *Oceanic Linguistics* 47(1): 45–113. https://doi.org/10.1353/ol.0.0006

Blust, Robert. 2014. Austronesian. In *The Oxford Handbook of Derivational Morphology*, Rochelle Lieber & Pavol Štekauer (eds), 545–557. Oxford: OUP.

Boarccaech, Alessandro. 2013. *Dicionário Hresuk-Português*. <https://issuu.com/kirstygusmao/docs/dicionario_hresuk_-_portugues> (1 June 2020).

Bowden, John, Hajek, John & Himmelmann, Nikolaus. n.d. *A sketch grammar of Waima'a*. Ms.

Bowden, John. 2010. Metathesis in Helong. *KOLITA* 8: 59–63. Yassir Nasunius (ed.), Jakarta: Pusat Kajian Bahasa dan Budaya, Unika Atma Jaya.

Correia, Adérito J. G. 2011. Describing Makasae: A Trans-New Guinea Language of East Timor. PhD dissertation, University of Western Sydney.

Campagnolo, Henri. 1973. La langue des Fataluku de Lórehe (Timor Portugais). PhD dissertation, Université de Paris V – René Descartes.

da Silva, Eng. Guilherme. 2012. *Disionáriu Wekais-Tetun*. Dili: Secretaria de Estado da Cultura.

Donohue, Mark. 2004. Typology and linguistic areas. *Oceanic Linguistics* 43: 221–239. https://doi.org/10.1353/ol.2004.0008

Donohue, Mark. 2015. Verbal inflection in Iha: A multiplicity of alignments. In *The Oxford Handbook of Inflection*, Matthew Baerman (ed.), 405–418. Oxford: OUP.

Donohue, Mark & Schapper, Antoinette. 2008. Whence the Oceanic indirect possession construction? *Oceanic Linguistics* 47(2): 316–328. https://doi.org/10.1353/ol.0.0017

Edwards, Owen. 2016. Metathesis and Unmetathesis: Parallelism and Complementarity in Amarasi, Timor. PhD dissertation, Australian National University.

Edwards, Owen. 2018a. Top-down historical phonology of Rote-Meto. *Journal of the Southeast Asian Linguistics Society* 11(1): 63–90.

Edwards, Owen. 2018b. Parallel histories in Rote-Meto. *Oceanic Linguistics* 57(2): 359–409. https://doi.org/10.1353/ol.2018.0016

Edwards, Owen. 2019. Reintroducing Welaun. *Oceanic Linguistics* 58(1): 31–58. https://doi.org/10.1353/ol.2019.0002

van Engelenhoven, Aone. 1987. Preliminaire aantekeningen voor een rekonstructie van het proto-Luangisch-Kisarsch. MA thesis, Leiden University.

van Engelenhoven, Aone. 2004. *Leti: A Language of Southwest Maluku*. Leiden: KITLV.

van Engelenhoven, Aone. 2009a. The position of Makuva among the Austronesian languages in East Timor and Southwest Maluku. In *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, K. Alexander Adelaar & Andrew Pawley (eds), 425–442. Canberra: Pacific Linguistics.

van Engelenhoven, Aone. 2009b. On derivational processes in Fataluku, a non-Austronesian language in East-Timor. In *The Linguistics of Endangered Languages: Contributions to Morphology and Morpho-Syntax* [LOT Occasional Series 13], Leo Wetzels (ed.) 333–362, Leiden: Leiden University.

van Engelenhoven, Aone. 2010. Tentatively locating West-Damar among the languages of Southwest Maluku. In *Studia Anthropologica: A Festschrift in Honor of Michael Chlenov*, Svetlana Chlenova & Artem Fedorchuk (eds), 297–326. Moscow-Jerusalem: Gesharim.

Engelenhoven, Aone van & Huber, Juliette. 2020. East Fataluku. In *Papuan Languages of Timor, Alor and Pantar*, Vol. 3, Antoinette Schapper (ed.), 347–426. Berlin: Mouton de Gruyter.

Flassy, Don A. L., Ruhukael, Constantinopel & Rumbrawer, Frans. 1984. *Struktur Mbahaam Maaq / bahasa Bahaam: Morfologi dan Sintaksis*. Jakarta: Departmen Pendidikan dan Kebudayaan.

Gil, David. 2015. The Mekong-Mamberamo linguistic area. In *Languages of Mainland Southeast Asia, The State of the Art*, Nicholas J. Enfield & Bernard Comrie (eds), 266–355. Berlin: De Gruyter Mouton.

Gomes, Sibylle M., van Oven, Mannis, Souto, Luis, Morreira, Helena, Brauer, Silke, Bodner, Martin, Zimmermann, Bettina, Huber, Gabriela, Strobl, Christina, Röck, Alexander W., Côrte-Real, Francisco, Parson, Walther & Kayser, Manfred. 2017. Lack of gene–language correlation due to reciprocal female but directional male admixture in Austronesians and non-Austronesians of East Timor. *European Journal of Human Genetics* 25(2): 246–252. https://doi.org/10.1038/ejhg.2016.101

Grimes, Charles E., Marçal, Carlos & Fereira, Paolino. 2014. *Introductory Dictionary of Mambae (Same): Mambae – English, English – Mambae, Mambae – Indonesia – Tetun Dili, Indonesia – Mambae, Tetun Dili – Mambae*. Darwin: Australian Society for Indigenous Languages.

Hawkins, Stuart, O'Connor, Sue, Ryan Maloney, Tim, Litster, Mirani, Kealy, Shimona, Fenner, Jack N., Aplin, Ken, Boulanger, Clara, Brockwell, Sally, Willan, Richard, Piotto, Elena & Julien Louys. 2017. Oldest human occupation of Wallacea at Laili Cave, Timor-Leste, shows broad-spectrum foraging responses to late Pleistocene environments. *Quaternary Science Reviews* 171(1): 58–72. https://doi.org/10.1016/j.quascirev.2017.07.008

Heston, Tyler M. 2015. The Segmental and Suprasegmental Phonology of Fataluku. PhD dissertation, University of Hawai'i.

Himmelmann, Nikolaus P. 2010. Notes on Waima'a intonation. In *Typological and Areal Analyses: Contributions from East Nusantara*, Michael Ewing & Marian Klamer (eds), 47–69. Canberra: Pacific Linguistics.

Huber, Juliette. 2008. *First Steps Towards a Grammar of Makasae: A Language of East Timor* [Languages of the World/Materials 195]. Munich: Lincom.

Huber, Juliette. 2011. A Grammar of Makalero. PhD dissertation, Leiden University.

Hull, Geoffrey. 1998. The basic lexical affinities of Timor's Austronesian languages: A preliminary investigation. *Studies in the Languages and Cultures of East Timor* 1: 97–174.

Hull, Geoffrey. 2001. A morphological overview of the Timoric Sprachbund. *Studies in Languages and Cultures of East Timor* 4: 98–205.

Hull, Geoffrey. 2002. *Waimaha*. Dili: Instituto Nacional de Linguística.

Hull, Geoffrey. 2003. *Galoli*. Dili: Instituto Nacional de Linguística.

Hull, Geoffrey. 2004. The Papuan languages of Timor. *Studies in Languages and Cultures of East Timor* 6: 23–99.

Hull, Geoffrey. 2005. *Fataluku*. Dili: Instituto Nacional de Linguística.

Jonker, Johan C. G. 1915. *Rottineesche Spraakkunst*. Leiden: Brill.

Josselin de Jong, Jan P. B. 1937. *Studies in Indonesian Culture, I: Oirata, a Timorese settlement on Kisar*. Amsterdam: Noord-Hollandsche Uitgevers-Maatschappij.

van Klinken, Catharina. 1999. *A Grammar of the Fehan Dialect of Tetun: An Austronesian Language of West Timor*. Canberra: Pacific Linguistics.

McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2): 125–166.

McWhorter, John. 2005. *Defining Creole*. Oxford: OUP.

McWhorter, John. 2007. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780195309805.001.0001

McWhorter, John. 2008. Why does a language undress? Strange cases in Indonesia. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94], Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 167–190. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.12mcw

Mills, Roger F. 1991. Tanimbar-Kei: An Eastern Indonesian subgroup. In *Currents in Pacific Linguistics: Papers on Austronesian Languages and ethnolinguistics in Honour of George W. Grace*, Robert Blust (ed.), 241–263. Canberra: Pacific Linguistics.

Mona, Stefano, Tommaseo-Ponzetta, Mila, Brauer, Silke, Sudoyo, Herawati, Marzuki, Sangkot & Kayser, Manfred. 2007. Patterns of Y-chromosome diversity intersect with the Trans-New Guinea Hypothesis. *Molecular Biology and Evolution* 24(11): 2546–2555. https://doi.org/10.1093/molbev/msm187

Mona, Stefano, Grunz, Katharina E., Brauer, Silke, Pakendorf, Brigitte, Castrì, Loredana, Sudoyo, Herawati, Marzuki, Sangkot, Barnes, Robert H., Schmidtke, Jörg, Stoneking, Mark & Kayser, Manfred. 2009. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Molecular Biology and Evolution* 26(8): 1865–1877. https://doi.org/10.1093/molbev/msp097

Penn, David. 2006. Introducing Dadu'a. Honours dissertation, University of New England, Australia.

Richards, Martin, Oppenheimer, Stephen & Sykes, Bryan. 1998. mtDNA suggests Polynesian origins in eastern Indonesia. *American Journal of Human Genetics* 63: 1234–1236. https://doi.org/10.1086/302043

Ross, Malcolm. 1998. Possessive-like attribute constructions in the Oceanic languages of north-west Melanesia. *Oceanic Linguistics* 37(2): 234–276. https://doi.org/10.2307/3623410

Ross, Malcolm. 2005. Pronouns as a preliminary diagnostic of grouping Papuan languages. In: *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide (eds), 15–65. Canberra: Pacific Linguistics.

Sampson, Geoffrey, Gil, David & Trudgill, Peter (eds). 2009. *Language Complexity as an Evolving Variable*. Oxford: OUP.

Schapper, Antoinette. 2009. Possession in Kemak. Paper presented at the 11th International Conference on Austronesian Linguistics, June 21–25. Aussois, France.

Schapper, Antoinette. 2010a. Bunaq, a Papuan Language of Central Timor. PhD dissertation, The Australian National University.

Schapper, Antoinette. 2010b. Neuter gender in eastern Indonesia. *Oceanic Linguistics* 49(2): 407–435.

Schapper, Antoinette. 2011. Crossing the border: Historical and linguistic divides among the Bunaq in central Timor. *Wacana, Journal of the Humanities of Indonesia* 13(1): 29–49. https://doi.org/10.17510/wjhi.v13i1.8

Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel* 90: 99–151. https://doi.org/10.4000/archipel.371

Schapper, Antoinette. 2017. Farming and the Trans-New Guinea family: A consideration. In *Language Dispersal beyond Farming*, Martine Robbeets (ed), 155–182. Amsterdam: John Benjamins. https://doi.org/10.1075/z.215.07sch

Schapper, Antoinette. 2019. Build the wall! Village fortification, its timing and triggers in southern Maluku. *Indonesia and the Malay World* 47: 220–251. https://doi.org/10.1080/13639811.2019.1554778

Schapper, Antoinette, Huber, Juliette & van Engelenhoven, Aone. 2012. The historical relation of the Papuan languages of Timor and Kisar. In *History, Contact and Classification of Papuan Languages*, Harald Hammarström & Wilco van den Heuvel (eds). Special issue of *Language and Linguistics in Melanesia* 2012: 194–242. Port Moresby: Linguistic Society of Papua New Guinea.

Schapper, Antoinette, Huber, Juliette & van Engelenhoven, Aone. 2014. The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In *Alor-Pantar Languages: History and Typology*, Marian Klamer (ed.), 99–154. Berlin: Language Science Press.

Schapper, Antoinette & de Vries, Lourens. 2018. Comparatives in Melanesia: Concentric circles of convergence. *Linguistic Typology* 22(3): 437–494. https://doi.org/10.1515/lingty-2018-0015

Steinhauer, Hein. 1993. Notes on verbs in Dawanese (Timor). In *Topics in Descriptive Austronesian Linguistics*, Ger P. Reesink (ed.), 131–158. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

Steinhauer, Hein. 1996. Morphemic metathesis in Dawanese (Timor). In *Papers in Austronesian Linguistics*, Hein Steinhauer (ed.), 217–232. Canberra: Pacific Linguistics.

Steinhauer, Hein. 2014. Blagar. In *Papuan languages of Timor, Alor and Pantar*, Vol. 1, Antoinette Schapper (ed.), 147–218. Berlin: Mouton de Gruyter.

Suter, Edgar. 2012. Verbs with pronominal object prefixes in Finisterre-Huon languages. In *History, Contact & Classification of Papuan languages*, Wilco van den Heuvel & Harald Hammarström (eds). Special issue of *Language and Linguistics in Melanesia* 2012: 23–58.

Taber, Mark. 1993. Toward a better understanding of the indigenous languages of Southwestern Maluku. *Oceanic Linguistics* 32(2): 389–441. https://doi.org/10.2307/3623199

Tumonggor, Meryanne K., Karafet, Tatiana M., Downey, Sean, Lansing, J. Stephen, Norquest, Peter, Sudoyo, Herawati, Hammer, Michael F. & Cox, Murray P. 2014. Isolation, contact and social behavior shaped genetic diversity in West Timor. *Journal of Human Genetics* 59: 494–503.   https://doi.org/10.1038/jhg.2014.62

Usher, Timothy & Schapper, Antoinette. n.d. The Greater West Bomberai Family. A demonstration. Ms.

Veloso, Alexandre. 2016. A Grammar Sketch of Naueti, a Language of East Timor. MA thesis, Leiden University.

Wellfelt, Emilie. 2016. Historyscapes in Alor. Approaching Indigenous Histories in Eastern Indonesia. PhD dissertation, Linnaeus University.

Willemsen, Jeroen. 2015. Function and Grammaticalisation in Alor-Pantar Asymmetrical Verb Serialisation. MA thesis, Leiden University.

Becoming Austronesian

Mechanisms of language dispersal across southern Island Southeast Asia and the collapse of Austronesian morphosyntax

Mark Donohue and Tim Denham

Living Tongues Institute for Endangered Languages /
The Australian National University

We examine the spread of Austronesian languages as a process that proceeded in different ways at different times, even in the same locale. We examine the many ways a language can show 'Austronesian traits', and confront this with the known presence of pre-Austronesian languages across Island Southeast Asia, and the inferred similarity of social processes between mainland and Island Southeast Asia. We argue that many languages which are classified as Austronesian are indeed exemplary Austronesian languages, but that many others should be considered to be the outcome of creolisation processes, and yet others show the traces of scenarios involving (imperfect) language shift from earlier non-Austronesian languages. Indeed, many of the languages should be considered to be non-Austronesian languages ('Papuan') with (in some cases minimal) Austronesian (lexical) veneers.

Keywords: language contact, creolisation, substrate, family profile, typology, Island Southeast Asia, Austronesian

1. Introduction: The spread of Austronesian languages across Island Southeast Asia

In this paper we examine the spread of Austronesian languages across southern Island Southeast Asia.[1] In contrast to most previous portrayals (eg., Bellwood, Fox & Tryon 1995; Bellwood 2007), we focus on the spread of Austronesian languages

---

1. We use the term 'Island Southeast Asia' to refer to the islands found in a triangle with points at Taiwan, Sumatra, and New Guinea. This approximates Solheim's 'Nusantao' region (1984–1985, 2006). While all these terms are ambiguous, we will concentrate on the southern part of the range, approximately the area in which the modern states of Malaysia, Singapore, Indonesia and Timor Leste are found.

through this region as a process distributed in time even more than in space. The process is, in many senses, ongoing and needs to be considered in terms of the broader interaction within the region between mainland Southeast Asia and New Guinea. We illustrate the diversity of ways a language can 'be Austronesian', with particular reference to those languages with a creole-like typology. In doing so we suggest that the distribution of different types of languages in eastern Indonesia is best explained by assuming a linguistic mosaic such as that described for the languages of mainland Southeast Asia (eg., Enfield 2005), and that a range of social contact scenarios similar to those attested and evidenced in mainland Southeast Asia is also relevant across Island Southeast Asia.

The dispersal of Austronesian languages is an historical process, for which the precise antiquity, human migratory implications and cultural associations are not always clear (compare Bellwood 2005; Bulbeck 2008; Donohue & Denham 2010; Denham and Donohue 2012a). The dispersal of Austronesian linguistic forms from begun after 4000 years ago, based on archaeological evidence across the Batanes Strait dating to this period (eg, Piper et al. 2009). However, the claimed association between material cultural linkages and linguistic dispersal is still fundamentally an assumption, despite recurrent attempts to provide a robust foundation (eg, Pawley & Green 1973; Shutler & Marck 1975; Pawley 2007). Consequently, although the dispersal of Austronesian languages southward across Island Southeast Asia from Taiwan is generally considered to date to c. 3800–3500 years before present, the precise antiquity is uncertain, and dates in regional locales are even less firm.

Additionally, discussions of Austronesian dispersal are repeatedly confused by the conflation, obfuscation and misapplication of categories that are not intrinsically interchangeable (see discussion in Oppenheimer 2004; Terrell 2004; Donohue & Denham 2010; Denham & Donohue 2012a). Although the term 'Austronesian' is often considered polyvalent, its application should be specific. 'Austronesian' refers unambiguously only to languages and neither to people nor to material culture. Archaeological remains from Taiwan should be considered Taiwanese material cultural traits, as opposed to 'Austronesian' in the sense that is relevant to the history of islands to the south of Taiwan, with other regions contributing to the long-term history of Island Southeast Asia. Similarly, the exact provenance of 'Asian' genetic characteristics within Island Southeast Asia and beyond is not clear; they likely include a small Taiwanese component, as well as components from other places (Jinam et al. 2012). Additional work (eg., Hill et al. 2007; Soares et al. 2011; Tumonggor et al. 2013) suggests that, regardless of the provenance of the 'Asian' genetic characteristics, any Taiwanese component represented is only minor. Indeed, Tumonggor et al. (2013: 170, 172) note that the genetic data is "consistent with a rapid expansion from Taiwan to the Philippines and Indonesia, but population

dispersals in the opposite direction are equally likely" and that "many aspects of culture – notably the widespread dispersal of Austronesian languages – are not obviously associated with genetics". Importantly, the dispersal of linguistic, genetic and material cultural traits across Island Southeast Asia should not be assumed to correspond. Languages, genes and cultures may have dispersed separately and together at different times and in different places within the late Holocene (the last 4000 years) history of Island Southeast Asia; indeed, various components of each are also likely to have distinct temporal and geographical patterns of dispersal (discussed in Denham & Donohue 2012a).

In this paper we argue that the dispersal of Austronesian languages across Island Southeast Asia should be considered against the broader historical backdrop evidencing widespread and long-term linguistic commonalities and interactions across the southern Southeast Asian region, including the Southeast Asian mainland. Just as on the mainland, extensive language contact and the widespread, stable co-existence of languages of different families must be posited across Island Southeast Asia. The emphasis of our argumentation is upon the temporal processes through which people became Austronesian, at least in terms of the languages they spoke.

We shall highlight a typological characteristic, morphological isolation, that is found in many languages of the Southeast Asian region, both mainland and island, but very rare from a global perspective. After a quick summary of some of the socio-historical similarities and differences between mainland and Island Southeast Asia, we shall remark on the difference in distributions of language families in the two regions. We then present compelling reasons to regard the proposed linguistic genealogical homogeneity across much of Island Southeast Asia as an error of analysis. A more nuanced account will reveal the same kind of diversity and the same kind of geographic intermeshing of language family ranges within Island Southeast Asia as we find on mainland Southeast Asia. The intention is not to criticise the whole field of Austronesian historical linguistics, but rather to critique the often monodirectional nature of historical linguistic research and the search for language affiliations along single dimensions. Since Austronesian languages demonstrably represent a wide variety of types, in typological and historical linguistic terms, we believe that it is best to think of the process of 'becoming Austronesian' as not being a single, uniform process, but rather different processes involving different starting points, different trajectories, and different time spans in different societies (even those in the same local area). These differential ways of becoming Austronesian are not simply asserted: they can be measured in multiple dimensions.

2.   The unnaturalness of isolation

Classification of languages according to scales of morphological character has been employed for a long time (see discussion in Sapir 1921). Here we focus on the nature of so-called isolating languages. This section will show that the cross-linguistically rare trait of 'isolation' is concentrated in the complex contact area that is mainland Southeast Asia, and is also found in numerous languages of southern Island Southeast Asia (see other chapters in this volume).

Isolation is defined as the absence of large amounts of bound morphology, such that the morpheme:word ratio approaches 1. Examples (1)–(3) show, with the English translations, sentences in four languages expressing the same meaning. In the Iha example, from the western edge of New Guinea, the two words contain six morphemes, with the focus on the verb; the Tukang Besi example, from central ISEA shows three words with seven morphemes. The English translation uses six words with eight morphemes, and the Papuan Malay in (3) has six morphemes in its six words.[2] The measures of average morphemes/word for the different languages are: Iha, 3; Tukang Besi, 2.3; English, 1.3, and Papuan Malay, 1.0. If we examined just the amount of morphology on the verb, the locus of marking in most languages, we have Iha: 5, Tukang Besi: 4, and Papuan Malay 1.

Iha

(1)   ***Komoh ni-ndo-nwe-mb-ih.***
      taro   1PAT-CAUS-eat-YESTERDAY.PST-3
      'They made us eat taro yesterday.'

Tukang Besi

(2)   ***No-pa-manga=kami   te=opa   dinggawi.***
      3R-CAUS-eat=1PAUCAL CORE=tuber yesterday
      'They made us eat taro yesterday.'

(3)   ***Dong kasi makan kladi kitong kemarin.***
      3PL   give   eat   taro   1PL   yesterday
      'They made us eat taro yesterday.'

We can further note the nature of the bound morphology. In Iha and Tukang Besi we find agreement for two arguments on the verb; Iha has additional inflectional morphology in the form of the tense marking suffix, and both of these languages have a bound derivational prefix marking causative. Further, the Tukang Besi example shows a core case marker on the nominal object of the clause, *opa*. The English translation shows more than one morpheme per word only with the suppletive

---

2.   Donohue and Sawaki (2007) discuss the possibilities of cliticisation of forms such as *dong=* in Papuan Malay. The language also allows non-bound forms, as shown here.

forms *make* and *us*, which include information, respectively. In the Papuan Malay sentence, there is exactly one morpheme per word. Just these three examples show us that measures of isolation are sensitive to both inflectional and derivational possibilities; to agglutinative and suppletive forms; and, to possible differences between verbal and nominal morphology. We could easily add other widely-attested features to a discussion of the up 'isolating'. Citing features coded in the World Atlas of Language Structures (WALS) database (Haspelmath et al. 2005), the features listed in Table 1 all bear on the measure of degree of morphological isolation (all of them, monitoring as they do bound morphology, would count towards a language being classified as less isolating).

Table 1. Morphosyntactic features bearing on the notion of 'isolating' (from WALS)

Affixed gender	Affixed demonstratives
Affixed tense/aspect/mood	Affixed subordinating
Affixed (in)definiteness	Core case affixation
Affixed possession	Subordinating structures
Adpositional agreement	Verbal agreement

Verbal affixation for: negation, optative, epistemic and situational possibility, evidentiality, reciprocals, passive, antipassive, applicative, causative, interrogative, desiderative

Verbal suppletion for: tense, aspect, number, imperative, hortative

Map 1 shows the distribution of isolating languages, as coded by Bickel and Nichols in the WALS (see Bickel & Nichols 2011 for a discussion on what this feature represents linguistically). On this map solid black circles represent languages with only isolating or tonal morphological profiles, and grey circles are those languages coded as having 'isolating/concatenative' profiles. From this map, it is clear that the distribution of languages with a predominantly isolating character is neither random, nor widespread. Bickel and Nichols (2011) write that 'Languages with isolating formatives, or traces of isolating structure in mixed types, are mostly confined to the Sahel Belt of West Africa and to Southeast Asia and the Pacific.'

While useful, Map 1 is based on only 165 languages coded for the feature '± isolating'. Due to this lack of resolution, it is not straightforward to determine the location of the 'isolating' regions. Of concern to us here, the region of 'Southeast Asia and the Pacific' covers a lot of ground. In Map 2 we see those languages with a low (grey) or very low (black) level of morphological elaboration, essentially calculated from the collection of features listed in Table 1, coded for a database of 2371 languages.[3] From Map 2 it is clear that while the Sahel is an

---

3. The map excludes known pidgins and creoles; the grey and black dots mark languages with a morphological complexity that is 1 (grey) or 2 (black) standard deviations lower than world average.
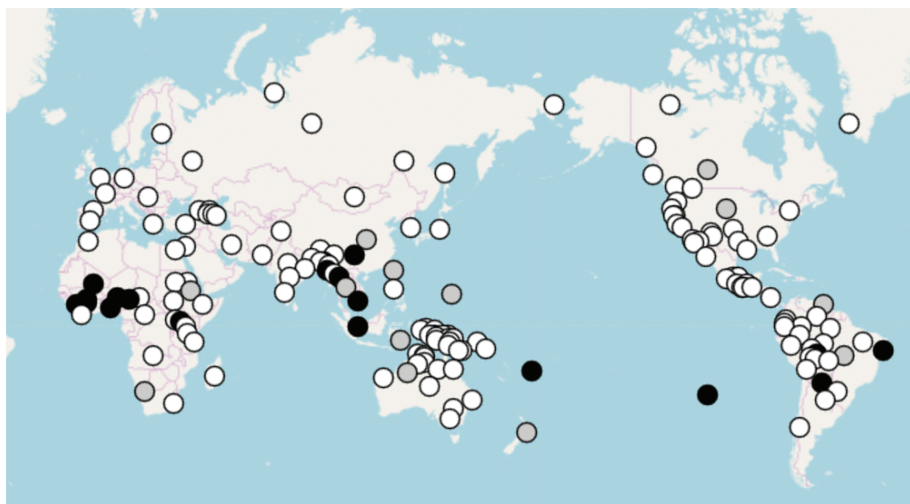
Map 1. Isolating languages (as coded by Bickel and Nichols 2011 in WALS)

area with languages showing little morphological elaboration, it is at nothing like the density or extremity found in mainland Southeast Asia.[4] While 'the Pacific' hosts a number of additional isolating languages, most of them are found west of New Guinea, or on the (western) New Guinea mainland, joining this region to mainland Southeast Asia. It is clear that, while an isolating profile *can* arise in a language from almost any part of the world, it is rare for a region to be dominated by isolating languages.[5]

---

4.   The visual impression is borne out by statistical analysis. Using Autotyp areas (Bickel and Nichols 2002), 20% of West African languages are coded as isolating, compared to 60% of Southeast Asian languages (n = 203 and 302, respectively, with 'Southeast Asia' including those languages of mainland Southeast Asia as far north as China, and Island Southeast Asia as far as Borneo), a statistically significant difference ($p < 0.0001$, chi-squared test). The average morphological complexity (on a 0–1.0 scale) for a West African language in the sample is 0.39, while the Southeast Asian languages in the sample show an average value of 0.18, a difference that is again statistically significant ($p < 0.0001$, two-tailed t-test). The languages east of the Southeast Asian languages, but west of New Guinea, are less likely to be coded as 'isolating' (only 9%; n = 135), but have an average morphological complexity of 0.29, also significantly lower ($p < 0.0001$) than the values found in West Africa.

5.   Other than the Sahel and (mainland and Island) Southeast Asia, the only parts of the world where very low levels of morphological elaboration are found are two regions of northern New Guinea and one in the south-east; Vanuatu and Polynesia in the Pacific; and the Khoisan region in southern Africa.
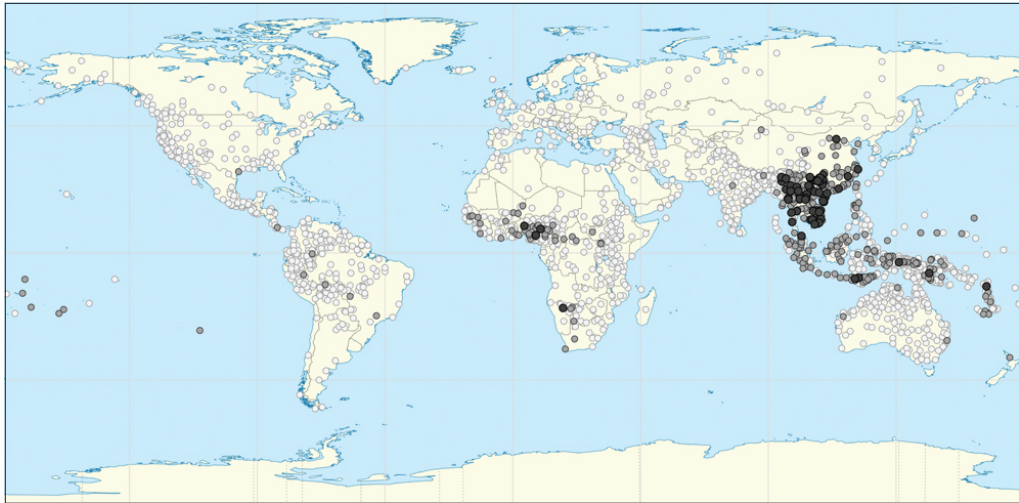
Map 2. Languages with low levels of morphological elaboration

Map 2 is an improvement over Map 1 in terms of coverage, but it suffers from the same collapsing of multiple dimensions that could be used to measure isolation. Does a language count as isolating if it displays an isolating setting for only one of the parameters mentioned in Table 1? Do the different parameters have independent existences, or is there a morphological conspiracy? Maps 3–8 present the distribution of the features discussed with respect to examples (1)–(3), and the addition of '± subordinating morphology'; the frequencies of these different features are shown in Table 2.[6] Note that there is a significant degree of correlation between the different values; if the distribution of the five features (shown in Maps 3–7) was truly random we would expect only 0.4% of the languages in the sample to show all the features together (the product of the frequencies in rows 2–6).

Table 2. Morphosyntactic features shown in maps 3–8

Map	Feature	Global frequency	Sample size
1, 2	isolating	11.4%	1384
3	no verbal agreement	43%	2378
4	no core case marking	43%	2373
5	no tense marking	27%	2162
6	no bound causative	32%	1979
7	no subordinating morphology	26%	1923
8	Sum of features 3–7	8%	1649

An examination of these features shows that the Sahel belt is less prominently featured, and that the correlation between the features in Table 2 is geographical. The 'Southeast Asia and the Pacific' described in WALS contracts or expands according to the feature examined, and is not evenly distributed across all of that region. Of note is the fact that for some features we do *not* see the disjunct distribution found in Map 2, with mainland Southeast Asia separated from the (western) New Guinea region. Map 8 combines the features shown in Maps 3–7 showing how many are present in an individual language (with the darker shades indicating more of the five features present). From Map 8 we can see that there is a clear connection between mainland Southeast Asia and western New Guinea. For some features this trail

---

6. The features are similarly drawn from, and defined in, WALS. The interested reader is referred to chapters 66, 67, 98, 99, 102, 111, 125, 126 and 127 (Dahl & Velupillai 2011a, 2011b; Comrie 2011a, 2011b; Siewierska 2011;, Song 2011; Cristofaro 2011a, 2011b and 2011c). The feature 'no verbal agreement' follows chapter 102; 'no tense marking' is based on the union of negative values for both of WALS features 66 and 67; 'no bound causative' follows chapter 111, and 'no core case marking' is coded for languages showing negative values for chapters 98 and 99. The feature 'no subordinating morphology' is calculated for languages that show 'balanced' values for all of WALS chapters 125–127, and no 'deranked' values.
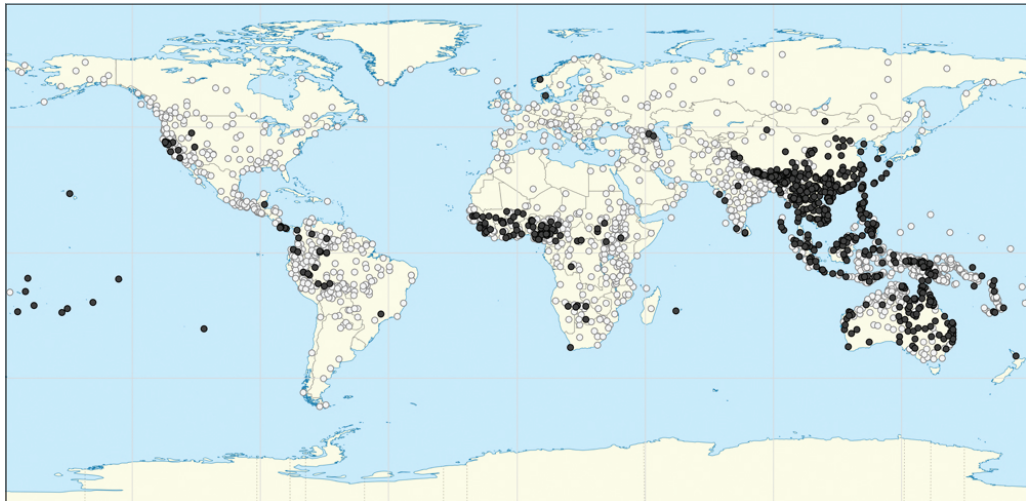
Map 3. No verbal agreement
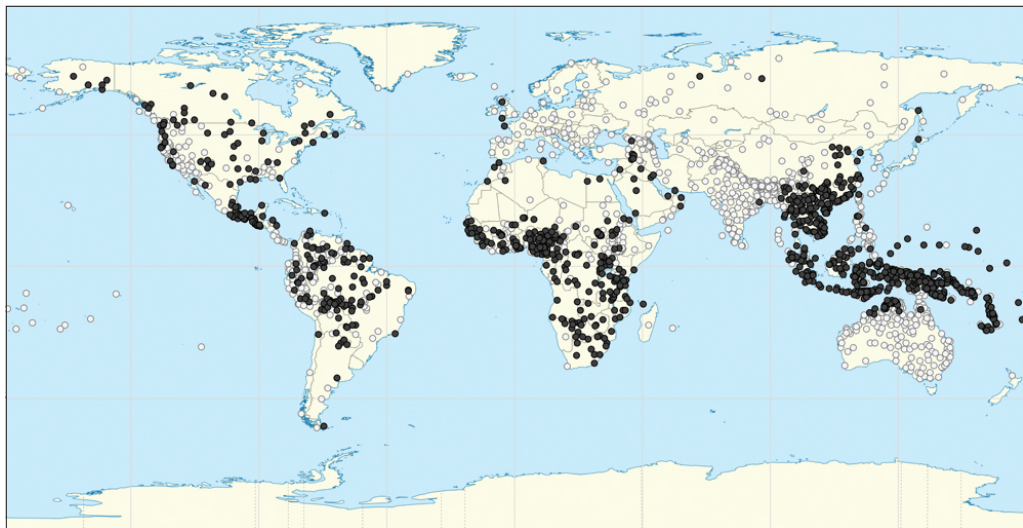
Map 4. No core case marking
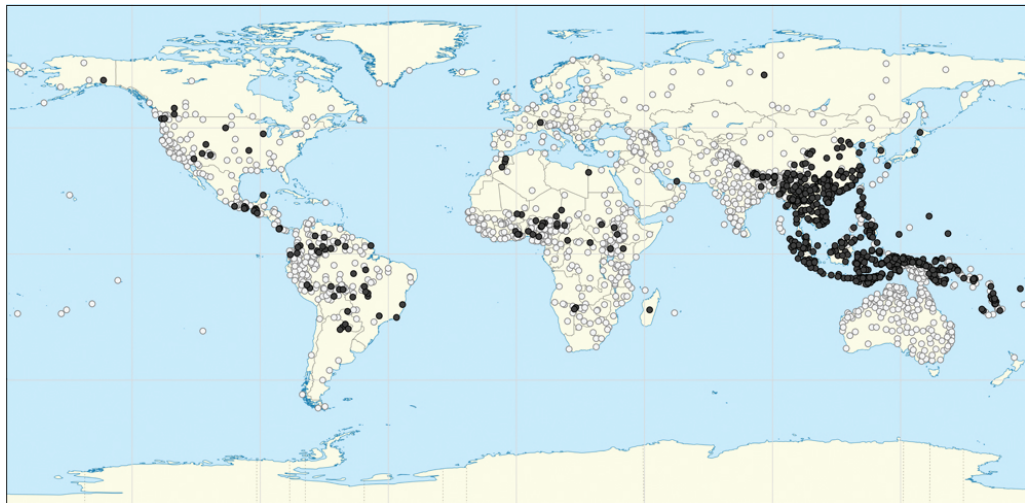
Map 5. No tense marking
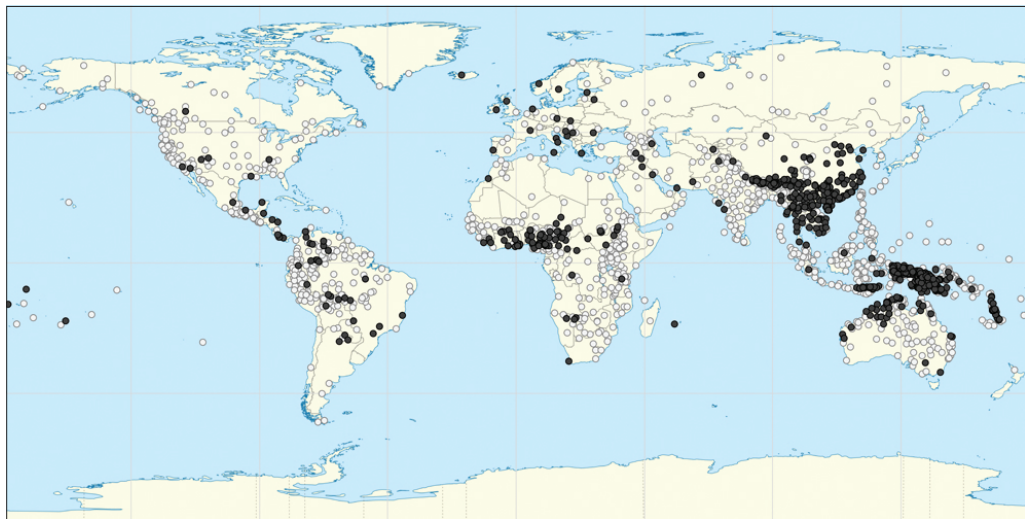
Map 6. No bound causative
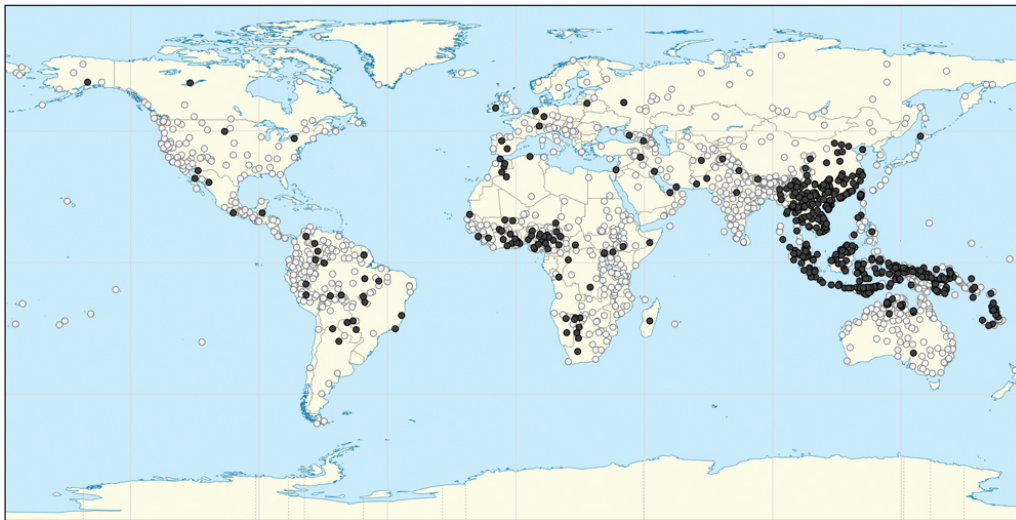
Map 7. No subordinating morphology
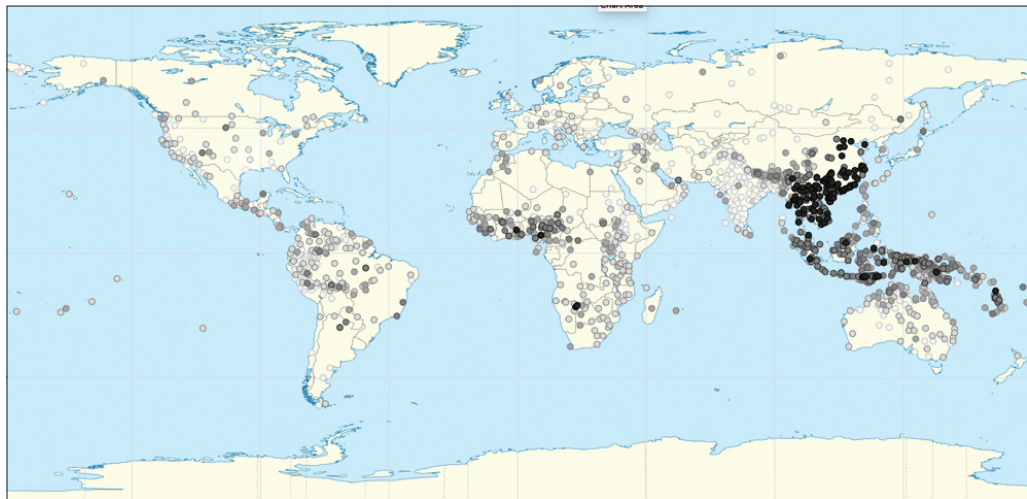
Map 8. Combined features from Maps 3–7

has been muddied, probably by the spread of language typologies from the north, but it remains detectable and shows how this region stands out against the sparse distribution of isolating languages elsewhere in the world.

Judging from the information we have seen in Table 2 and Maps 3–8, we can state that 'isolation', measured as a single parameter or decomposed into separate variables, is not a normal state of affairs for most languages or language areas around the world. Rather, some degree of inflectional (and derivational) morphology is the norm. In one area, however, stretching from the (eastern) Himalayas to (western) New Guinea (depending on the feature examined), isolating characteristics are modal with a much higher than expected frequency, as can be informally judged from the maps. The same area is remarkable for the widespread adoption of SVO order in Austronesian languages, independent of any recognised subgroup that includes these languages, a region that also excludes Austronesian languages to the north in Taiwan and the Philippines (Donohue 2005, 2007; Austronesian languages are universally acknowledged to have originated on Taiwan, with verb-initial word order). We shall examine this distribution in terms of the processes that are known to have brought about such widespread language contact in mainland Southeast Asia, compared to our assumptions about language and population dispersals in Island Southeast Asia.

3.   Southern Southeast Asia

In this section we will point out some of the similarities, and some of the differences, between Mainland and Island Southeast Asia. We shall combine this exposition of differences, often superficial, with the many similarities that we have seen in Section 2, and the social histories that are implied by those data.

The area of 'Southeast Asia' has been described in terms of two subparts, in many ways unified and yet frequently treated disparately. In the west on the mainland, the area most proto-typically referred to without qualification as 'Southeast Asia' (eg., Enfield 2005) is centred approximately on Thailand and includes surrounding regions. Enfield defines it as 'the region encompassing Vietnam, Laos, Cambodia, and Thailand, with some extension west into Burma, south into Peninsular Malaysia, and north into southern China' (2005: 182). This practical definition explicitly excludes the other subpart, Island Southeast Asia (also known as Maritime Southeast Asia, or Insular Southeast Asia, Indo-Malaysia or (earlier) the Malay archipelago). This second region can be characterised as the islands lying between the Malay peninsula and New Guinea; whether a northern extension to include the Philippines is counted or not varies from author to author, and for the purposes of this paper we will take the more restricted definition, that roughly corresponds to the insular

part of the modern states of Malaysia, Indonesia and Timor Leste. For clarity, we shall refer to Island Southeast Asia and mainland Southeast Asia when we want to differentiate the two regions.

There are many similarities between these two regions. We can summarise just a few (eg., Bellwood 2007; Bellwood & Glover 2004; Bellina et al. 2010; Osbourne 2010; Wright et al. 2013):

–   Both areas have a long history of modern human occupation, especially around coastal regions.
–   Both regions host multiple ecological zones, with high mountainous hinterlands leading to coastal plains that in some cases extend far inland.
–   Both regions have seen long-distance trade and interaction spheres, in which goods from one zone are transported to another zone, or out of the region entirely. These kinds of long-distance interactions are repeated throughout the Holocene period and into modern times.
–   Both regions have complex proto-historic traditions, including the arrival of Indic states in the last 2000 years (eg., Coedes 1968), with the concomitant presence of Hindu architectural styles and cosmologies, and the traces of Sanskrit vocabulary in languages of the regions.

At the same time, substantial differences can be ascribed to the two regions.

–   language family distributions
    Mainland Southeast Asia is a mosaic of intertwined language families. Austroasiatic, Austronesian, Hmong-Mien, Tai-Kradai and Tibeto-Burman languages are all found, with massively disjunct distributions throughout most of the region. It is not possible, just looking at the distribution of different families, to identify exclusive 'heartlands' for the different families in the region. In some cases communities are distinguished from their geographically close neighbours by altitude and lifestyle, in others only by local custom and tradition. By contrast, the map of Island Southeast Asia is dominated by Austronesian languages, which are the only languages found in the Philippines and on most of the islands of Indonesia west of New Guinea. There are pockets of non-Austronesian languages, but the overall level of diversity is nothing like that accepted for mainland Southeast Asia.
–   written traditions
    Both regions were the recipients of Indic written traditions during the period of Indic empires (starting circa 2,200 years ago). This led to the development of local scripts, which flourished in local states. In mainland Southeast Asia these scripts continued in use and in development into modern times, allowing us to examine written histories spanning centuries if not millennia. In Island

Southeast Asia the use of the Indic-derived scripts, which never spread over much of the region, was either abandoned earlier on or never so prevalent in the epigraphic record, leading to fewer historical records with the kind of detail that is found on the Asian mainland (until the later arrival of cultures bringing Arabic and Roman scripts).

– political traditions

Both regions saw the rise of local states of various degree of durability. In mainland Southeast Asia some of these states, particularly away from the coast, involved extreme control of the local societies that they dominated, and ended up driving nonconformist members of their populations into more marginal environments (see Scott 2010). In Island and mainland Southeast Asia, the independence of food production and personal mobility offered by coastal locations created historical situations whereby state control was never as pervasive as it was for inland locations on the mainland.

The question that we must address is why these differences in language distributions are found in two otherwise similar areas. It is trivial, but possibly relevant, to note that Island Southeast Asia consists of islands, while mainland Southeast Asia is a single land mass; but in what ways did these different geographic facts affect the social milieux of the two regions? In Maps 9 and 10 we can see the difference in scale between one small area of northern Laos, and the vast area in which the Austronesian languages are described. Map 9 shows a region in northern Laos, only 125km from west to east, containing languages from four different families: Tibeto-Burman (grey), Tai-Kadai (diagonal stripes), Hmong-Mien (vertical stripes) and Austroasiatic (square hatching for Khmuic, black for Palaungic) (based on Epprecht et al. 2018). The area occupied by Map 9 is shown in the top left of Map 10, which outlines the areas in which Austronesian languages of Island Southeast Asia are described. It is true that there are areas on the fringes of the Austronesian world, such as around the coasts of mainland Southeast Asia and New Guinea, where similar complexities can be found, and that there are a few pockets of non-Austronesian languages between New Guinea and the Southeast Asian mainland, but the lack of described family diversity, compared to mainland Southeast Asia, is striking. What is also striking is that *all* areas of mainland Southeast Asia where there are hills that could offer refuge from lowland states show the sort of diversity depicted in Map 9. It is clear that the Austronesian world overwhelmingly consists of islands, which are highly accessible to a maritime culture. It must, however, be remembered that islands, too, have hinterlands, which have the potential to be areas of refuge, where societies can escape the control of large states; indeed, some of these islands are major land masses, such as Borneo, with extensive interiors and major ethnographic divisions between coastal and interior peoples. Further, since travel

between two islands is most economical without stopping at intervening islands, even small islands have the potential to remain (in some ways) outside the social control of a surrounding polity, if they wish (see Section 4 for more discussion). As with the mainland, the geography of Island Southeast Asia offers the opportunity for different socio-political entities that overlap in space.
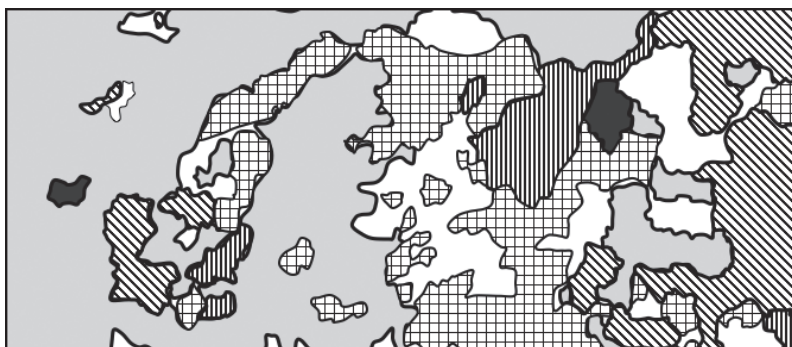
Map 9. Language families in one small area of northern Laos
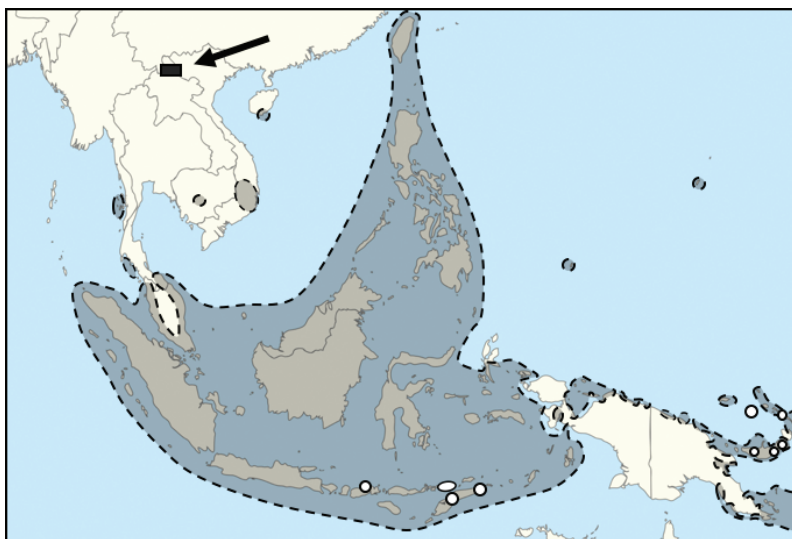
Map 10. The spread of Austronesian languages in Island Southeast Asia

Seeing the differences between the two maps, we must ask how such a large area of the world was so radically 'Austronesianised', without wholesale genetic replacement (see Section 1, and also Donohue & Denham 2011; Denham & Donohue 2012b).

4. Accounting for language distributions

Mainland SEA exhibits many levels of linguistic integration: while there are many small, loosely structured minority groups, there is also clear evidence for small elites achieving dominance in most of the lowland states (eg., Munoz 2006, Osborne 2010 and others).

The spread of polities, or trading circles, is advantageous in many ways to many of the people who live within the bounds of the newly stable social unit. That does not, however, mean that the spread of a new social domain will be uniformly influential. Scott (2010) describes the decision to live on the margins of mainland Southeast Asian empires by numerous small groups that chose to escape the control of the plains' polities. In the Indo-Malaysian archipelago the decision to not be included is an easy one: simply living on an island that is not connected to the others in a chain is enough to be left alone, given the way maritime travel can bypass islands more easily than stopping at all possible waypoints. This means that 'controlled' areas can be interspersed with regions that are outside the 'control' of a polity. Unlike the mainland, where nonconformist groups had to physically relocate to less accessible locations in highland regions (typically on the fringes of the lowland states), it is possible for a much more diverse, and less bounded, group of societies to occupy the same time and space in an archipelago region. In a typical mainland Southeast Asian setting (eg., Lebar et al. 1964 and others), emerging polities expand along river valleys and flat terrain. Ethnic groups which become refugees or discontents from these polities frequently come to occupy altitudinally differentiated zones, as shown schematically in Figure 1 (similar behaviour is also described for the Caucasus; Nichols 1998). While these zones lack the ease of movement of the flat lowlands, they allow the hill-dwellers to exploit a wider range of ecological zones, and so have access to resources that are denied to
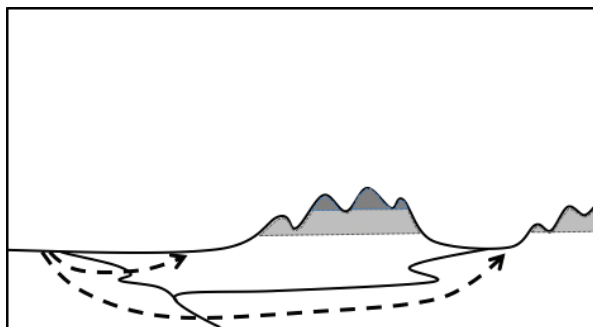
Figure 1. Social barriers (land mass)

the valley-dwellers. From the valley-dwellers' perspective travel between points in valleys, even when that requires a considerable detour from a straight-line path, is easier than negotiating the social and geographic hurdles imposed by the hills that lie between valleys. This continued contact makes for yet more cohesion along the valleys and reinforcement of the contiguous, though frequently interrupted, geography of the lowland polity.

In an archipelagic society the expansion of a new polity is not constrained by the need to avoid particular terrain types when travelling. An emerging society (or social network) in a maritime environment can easily by-pass islands that are not included within its sphere, due to the ease of water-travel (the same is also true in other environments; see McAndrews et al. 1997; Stanish 2003; Shillington 2005 for examples in the Andes and the Sahara). In an archipelagic society it is quite possible, even likely, that at different times the expanding polity (or social network) would expand past areas that were not incorporated due to the ease of travelling across the sea and the ease of avoiding non-incorporated societies. Expressed from the alternative perspective, a non-incorporated society, the social equivalent of the hill-dwellers described earlier, can exist on an island that lies directly on the path between two archipelagic-societies settlements. On larger islands there may well be a hinterland that is not part of the coastal society, just as in the scenario summarised in Figure 1, but importantly travel between economically powerful incorporated coastal communities on different islands does *not* reinforce the archipelagic society, since the terrain traversed in such travel is sea and does not have to involve passing through like communities. In Island Southeast Asia the polities of Srivijaya and Majapahit, and later Malacca, were of this nature (Munoz 2006; Borschberg 2010). Figure 2 shows a schematic representation of this situation; different islands (circles
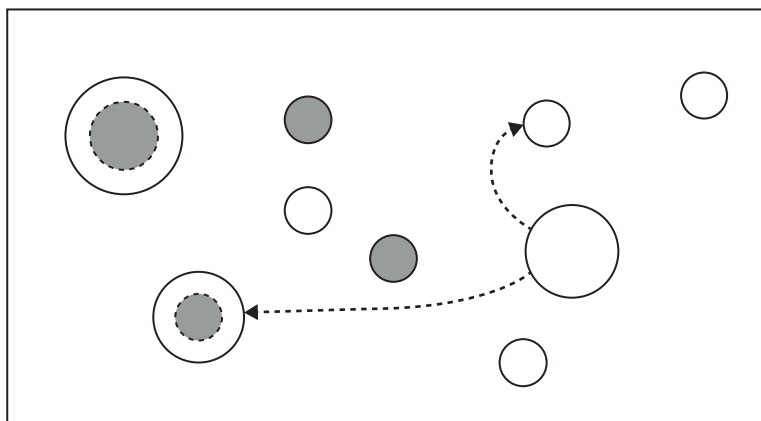
Figure 2. Social pluralism without barriers (archipelago)

in the figure) can be occupied by societies with different allegiances (shown as white vs. grey). Importantly, these different social groupings do not have to be contiguous; depending on the winds and ocean currents at different times of year, it can easily be the case that it is simpler to travel a longer distance past unincorporated societies than a short distance to a fully incorporated group. The large white-coloured island in Figure 2 is closer to a fully grey island than to the mixed islands on the left of the figure, but intervening reefs or adverse currents might mean that travel is simpler, and so colonisation of the coasts of the two large islands to the left of the figure is a simpler matter than occupation of the smaller and closer islands that remain fully grey (reminiscent of the barriers imposed by altitude schematised in Figure 1).

Within Island Southeast Asia, we frequently talk of the widespread Austronesian family, and it is indeed incredibly widely spread, much more so than any other pre-modern language family. On the other hand, the very atypicality of Austronesian, with an incredibly diverse typological profile and a very conservative basic lexicon, leads us to examine what it means to be a member of this family and whether membership in 'Austronesian' is of a similar nature to membership in (for instance) 'Indo-European' or 'Nakh-Daghestanian'.

The field of historical linguistics recognises two kinds of criteria in establishing language relations (elaborating on Noonan 2010):

–   The existence of regular correspondences (in the phonological system, or in morphological paradigms);
    (importantly, correspondences in the lexicon are not of themselves sufficient criteria, as the lexicon is too borrowable; for recent discussion, see Donohue, Denham and Oppenheimer 2012a, 2012b; Greenhill & Gray 2012)
–   Typological matches in terms of the 'nature of the language'
    (while this is neither a sufficient or necessary feature to establish genealogical relatedness, it is cited to exclude creoles from inclusion in the family of their lexifying parents – eg., see discussion in Thomason & Kaufman 1988; Sebba 1997)

The nature of the sound correspondences in Austronesian languages has been discussed in Donohue (2013a). To summarise the results presented there, the Austronesian languages of Indo-Malaysia sometimes show the same kind of profile for regular sound correspondences that characterises language families like Indo-European. Other individual languages, however, especially those found east of Sulawesi, show significantly lower levels of regularity. Maps 11 and 12 (reproduced from Donohue 2013a) show the locations of the extreme low-regularity languages; these are languages with such a lack of regularity in sound correspondences (and in most cases also a lack of reflexes of reconstructible morphological forms) that

Map 11. Below 50% regular
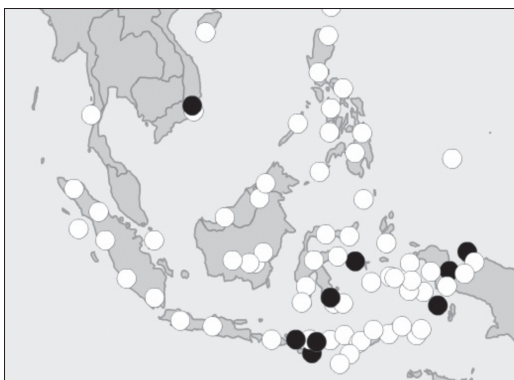
Map 12. Below 70% regular

they fall significantly below the standards of regularity attested in Indo-European languages.[7] (Donohue and Denham 2010 show the distribution of languages coded for the level of retention of Proto-Malayo-Polynesian etyma.)

Two possibilities suggest themselves as explanations for this failure to achieve regular sound correspondences (the two possibilities may be combined). The first solution is that much of the basic vocabulary of these Austronesian languages should be considered to be loanwords (from unidentified Austronesian sources).

---

7.   The languages marked on the maps are, from least regular to most are: Bima (48%) (in Map 11), Soboyo (62%), Ngadha (63%), Chru (67%), Kei (68%), Wandamen (68%), Biak (68%), Muna (69%) and Kambera (69%). Plenty of Austronesian languages in the Philippines and western Indonesian show regularities above 90%, including 100% regular languages (Donohue 2013a). A similar study of European languages results in most languages showing a regularity level above 90%. See footnote 9 for further discussion of the role of bound morphology in historical comparison.

This acknowledgement of lower levels of regularity is in keeping with the tradition of detailed studies of 'speech strata' described in Austronesian languages such as Ngaju Dayak (Dyen 1956), Rotuman (Biggs 1965) and Tiruray and Thao (Blust 1992 and Blust 2009, respectively). This would reduce the levels of retained Austronesian lexicon even further than reported in Donohue & Denham (2010), and would suggest that 'lexical innovation' (as viewed from the perspective of Proto-Austronesian) was (erratically) rife in some parts of the Austronesian world. Given that one of the characteristics of the Austronesian family is the extreme conservatism of the lexicon (Wichmann forthcoming, Donohue & Denham 2010), this is a surprising finding. The alternative possibility is that the languages in question have undergone large amounts of erratic and unconditioned sound change. Either option suggests a language history that is not the ideal one envisioned by advocates of the Neogrammarian hypothesis of regular sound change.

Objectively assessing the 'nature of the language family' is not trivial, but neither is it an insurmountable challenge. At the risk of missing many interesting and noteworthy quirks, we can use the WALS database (Haspelmath et al. 2005) and the World Phonotactics Database (Donohue et al. 2013) to test for features that occur with a significantly higher (or lower) frequency in the family compared to its neighbours. For example, Indo-European, with all its internal diversity, can be shown to be characterised by a number of typological traits which differ significantly in their frequency when compared to other languages of Eurasia, the continental-scale comparison (Table 3).[8]

Given that it is possible to characterise a language family as sharing certain broad typological tendencies, what does this method yield when applied to the Austronesian languages, compared to their region (Southeast Asia and the Pacific)? Without needing to resort to a table, we can summarise the distinguishing characteristics quite simply: Austronesian languages, taken as a group, show a tendency (compared to languages in other families in Southeast Asia and the Pacific) for head-initial syntax (as was first pointed out in Foley 1998). This is manifested by VO order at the clause level, prepositions rather than postpositions, and a N-modifier order in NPs. There are many violations of this, but even in those Austronesian languages with SOV clausal word order, nominal modifiers (other than the genitive) tend to follow the noun. In contrast to the thirteen morphosyntactic traits that are

---

8. Similar lists of defining traits can be assembled for other language families as well, though they are not presented here. Traits are only listed when the distribution of those traits in Indo-European is (statistically) significantly higher (or lower) than their distribution in all other language families of Eurasia, as determined by t-tests. The databases used were the 2400 language sample of morphosyntactactic properties described in Section 2, with ~700 languages from Eurasia, and phonological properties of 6500 languages, of which 2100 are from Eurasia (building on Donohue et al. 2013).

Table 3. Traits distinguishing Indo-European from other language families in Eurasia ($p < 0.0001$)

Features	Indo-European higher	Indo-European lower
Phonology	complex onsets	tonal contrasts
	unrestricted codas	velar nasals
	large fricative inventories	glottal stops
	voicing contrasts in stops	aspiration contrasts in stops
	large liquid inventories	
Morphosyntax	accusative alignment	numeral classifiers
	relative pronouns	N Rel order in NPs
	suffixing	(modifer N order in NPs)
	gender contrasts	
	plural suffixes	
	passive	
	agreement syncretisms	
	suppletive verbs (tense, aspect)	
	tense marking	
	initial Wh-words	

distinctive for Indo-European, covering alignment, valency, headedness and others, there is only one distinctive trait for Austronesian, head-initiality. In contrast to the nine phonological traits that are typical for Indo-European languages, not a single phonological trait characterises the Austronesian family as a whole.

However we choose to think about it, it is clear that the Austronesian family does not constitute a language family that coheres, lexically, structurally or typologically, in the same way that a well-established language family like Indo-European does. At the same time, the fact that 'well-behaved' languages are interspersed with erratic languages (from the perspectives examined here), without regard to geographic or subgrouping continuity, implies that there have been multiple pathways leading to language communities becoming Austronesian. In some cases we see the spread of language along with language communities; in others it is more parsimonious to think of the dispersal of Austronesian language traits to existing communities, which have acquired 'Austronesianness' to different degrees and in different ways. In some cases there are clear traces of a pre-Austronesian stratum, such as is seen in the distribution of agreement in Austronesian languages in and near New Guinea. For instance, 90% of Austronesian languages on or near New Guinea show verbal agreement, compared to 29% outside this region; this represents a highly significant difference ($p < 0.0001$) in the sample ($n = 227$). In other cases the absence of inflectional morphology, as discussed in Section 2, combined with the failure of the language to show regular sound correspondences at an 'adequate' level (Donohue 2013a), suggests that creolisation played a significant role in the dispersal of Austronesian traits.

Where creolisation is not obviously implicated in the formation of the modern languages, we can still observe that language shift was a common process involved in the dispersal of Austronesian languages. Far from being exceptional, language shift was normal, though sporadic, involving at times (and places) imperfect learning of the new Austronesian languages by speakers of languages from the different pre-Austronesian lineages, depending on local social circumstances. Evidence of this putative substrate varies from language to language, and can be (not exhaustively) illustrated by examining three very crude dimensions of variation: the lexicon, the phonology and the morphosyntax. We have mentioned that head-initiality is the only trait that can be used to identify Austronesian languages in their geographical context. If we are to examine just those features that are typical of (reconstructed) Proto-Austronesian and Proto-Malayo-Polynesian, and of several languages high in the genealogical tree, we can obtain an idea of what is and what is not typologically typical of an Austronesian language, as much as this is a valid approach. The different dimensions of the cube shown in Figure 3 are defined as follows:

Lexicon
– Examining a basic wordlist gives us an idea of the degree to which the basic lexicon has been retained in any given language (eg., Donohue & Denham 2010). Combined with the regularity of sound correspondences (as described in Donohue 2013a), this tells us how directly the lexicon of the language reflects Austronesian etymological sources.[9]

Phonology
– The phonological system of a language can be compared to the phonological system of Proto-Malayo-Polynesian, and any unexpected (or diachronically irregular) phoneme series (such as the voiceless prenasalised series /mp nt ntʃ ŋk/, or in other languages the loss of manner contrasts in stops, or the acquisition of tone or register contrasts) can be considered to be evidence of a non-Austronesian character. We examined ~200 phonological traits per language (see Donohue et al. 2013).

---

9. Under 'lexicon' we include the lexical entries for productive, bound morphology, the retention of which is at least as indicative of origins as is that of free lexemes. In the case of non-productive, or fossilised, forms that show forms similar to productive morphemes in other languages, the analysis is less certain because we cannot assume that non-productive morpheme look-alikes have been directly inherited, rather than borrowed. Many languages of eastern Indonesia and Timor Leste show few, if any, retentions of reconstructed Austronesian bound morphemes, or else exhibit these morphemes with non-productive functions. The large number of languages with agreement prefixes, the form and function of which are likely to be cognate with Austronesian genitive morphemes (see Wolff 1996 and others), are witness of a strong and enduring Austronesian influence, but the fact that these prefixes are not reconstructed in Austronesian makes them equivocal as evidence.

Morphosyntax

– Given what is known of the syntax of those languages high in the Austronesian tree, in Taiwan, and the syntax of the languages in question, various traits (for example, OV word order, agreement, possessive classes, or an isolating nature (as per Section 2) lacking case marking or valency changing morphology) are traits of a non-Austronesian character; just as the loss of the Austronesian voice system evidences a less Austronesian character. We used the morphosyntactic traits found in the WALS database (Haspelmath et al. 2005) as a basis for comparison, after decomposing them into ~200 mostly binary variables.

These three dimensions allow us to visualise the different extents to which different languages can be said to be Austronesian. In Figure 3, along with Map 13, we can see how positions along these different dimensions of 'Austronesian-ness' are occupied by a selection of different languages from across southern Island Southeast Asia.[10] Colloquial Javanese, Bima and Kei are low on all three of the scales examined, while Sangir (and most of the languages of the Philippines to the north and northern Borneo to the west) are high on all scales. Old Javanese has the same overall lexical and phonological profile that is seen in (modern) colloquial Javanese (Conners, this volume), but shows a considerably more conservative (from an Austronesian perspective) morphosyntax. Both of these attested varieties of Javanese are lexically and phonologically highly innovative. Languages such as Rote, Lio and Sika have 'well-behaved' lexicons, in that basic semantic fields are well-populated with reflexes of Proto-Malayo-Polynesian reconstructions, but show very non-Austronesian phonological and morphosyntactic traits. Gayo, in the top left rear corner in Figure 3, is as exemplary as Sangir structurally, but has a lexicon in which substantial basic elements have been acquired by indirect inheritance, or borrowing, or else have no recognisable widespread Austronesian etymology. In Kowiai the phonology and the lexicon are unsurprising, but the morphosyntax is highly aberrant, from a conservative Austronesian perspective. In the figure and map we can see that the languages discussed in other chapters of this volume all show low levels of 'Austronesian-ness' in at least one dimension,

---

10. The languages of the Philippines fill the same position as Sangir in this cube, being exemplary conservative Austronesian languages in most ways. Note that Figure 13 only illustrates the points made for Austronesian languages discussed here. The key to Figure 3 is: 1: Colloquial Javanese, Bima, Kei (W > E); 2: Cham, Kéo, Sika (W), Rote (E); 3: Riau Indonesian, Buru, Kowiai, Papuan Malay (W > E); 4: Sangir, Philippine-type lgs; 5: (Old) Javanese; 6: Banggai; 7: Gayo; 8: Muna; 9: Wolio; 10: Bugis; 11: Minangkabau, Iban; 12: Ma'ya. The location of these languages is indicated on Map 13. Donohue (2013b) elaborates on social scenarios leading to different outcomes, illustrating more of the possible outcomes than are exemplified here.

and are widely dispersed across Island Southeast Asia. Other languages could be added to this cube, including many at non-peripheral points within the cube, though we note that there is a dearth of languages that would be placed near the rear left lower corner: languages without significant lexical or morphosyntactic traits typical of Austronesian, but with an Austronesian phonology. This most likely reflects the implausibility of a social scenario in which non-Austronesian influence was strong enough that both the lexicon and the morphosyntax were disrupted, without leading the phonological system to acquire non-Austronesian features. Note that Ma'ya, located in the middle of the lower front face of the cube, shows extreme divergence from the Austronesian norm in many ways, and also occupies approximately the position in which Proto-Oceanic would be placed, following the metrics employed here.

Figure 3. The many dimensions of Austronesian-ness

Map 13. The many dimensions of Austronesian-ness

Importantly for the model we discuss, the 'aberrant' and 'exemplary' Austronesian languages in Island Southeast Asia are *not* separated discretely by geography. In the Philippines and in much of western Island Southeast Asia we generally find 'exemplary' languages, but within Wallacea we find an apparently random distribution of exemplary and aberrant languages; the two (non-discrete) categories mixing and intermingling on the contemporary map.

5.   Conclusions

We conclude that social admixture was the norm, not the exception, in the dispersal of Austronesian languages across Island Southeast Asia. Rather than a wave of Austronesianisation rolling out over a technologically simple and unsophisticated pre-Austronesian social milieu, an assumption that pervades much of the literature on the 'Austronesian dispersal',[11] an already integrated region saw the appearance of a new and attractive linguistic code, and that code was variously adopted or not adopted by different societies, as they saw fit. As with all social innovations, the spread was not uniform and did not follow uniform pathways. In some areas there was wholesale language replacement, presumably where the social pressure or social incentives for acquiring the new linguistic code were strong; in other areas the Austronesian content was adopted only much later and much more haphazardly. We can infer that in some communities the process of 'Austronesianisation' took generations and followed an almost reluctant path that did not see the wholesale loss of local structural or typological characteristics, nor the adoption of Austronesian forms from a single Austronesian parent.

We propose that languages listed as Austronesian form a spectrum of different non-discrete types, and different characteristics can be used as evidence of some of the different social processes that led to 'becoming Austronesian'. The logical conclusions are that the Languages Classified as Austronesian (LCA) deserve re-examination, with a number of different outcomes (vertex references refer to the vertices found in Figure 3):[12]

---

11. For example Klamer (2019), summarises much previous work (specifically Bellwood's "early farming" model) as asserting that "with their new food-producing technologies, the [Malayo-Polynesian] migrators were able to colonize and replace the preexisting hunter-gatherer (forager) populations of ISEA". Klamer rejects this characterisation, later concluding that "[t]he original populations of ISEA were not (only) hunter-gatherers but had among them sea-faring groups and agriculturalists". See also Latinis (2000), Oliveira (2008), and the debate in Donohue & Denham (2010).

12. Note that we are not proposing a 1:1 mapping relation: the cube representation conceals the different kinds of data that fit into the broad categories of lexicon, phonology and morphosyntax, each of which logically has its own history. Without more detailed work we cannot arrive at more precise sets of hypotheses about the linguistic and social histories of the different communities.

–   Some LCAs are members of the Austronesian family in the sense normally accepted by practitioners of the Comparative Method (vertex 4);

–   Some LCAs show all the structural characteristics of creole languages, and should be considered to be creoles, rather than the outcome of regular, uninterrupted inter-generational transmission (vertices 1, 2 and 3, and admitting the imprecise use of the term 'creole' generally);

–   Some LCAs can be considered to be members of the Austronesian family, but with such substantial pre-Austronesian substratal structural properties that they must be considered to have been remodelled via metatypy (eg., Ross 2006). In some cases languages of this type can perhaps better be considered mixed languages, or non-Austronesian languages that have been relexified (vertices 1, 2, 5 and 6);

–   Some LCAs show the right structural characteristics for a member of the Austronesian family, but with clear evidence that the vocabulary was not derived from a single source, implying that regular inter-generational transmission was not the process of the formation of the modern language (vertices 5 and 7);

–   Some LCAs show so few characteristics of Austronesian languages, with such irregular sound and morphological correspondences, that they should be considered *not* to belong to the Austronesian family, but rather should be thought of as being the vestiges of pre-Austronesian families once dispersed about Island Southeast Asia that have been affected by the influx of Austronesian lexical items and, in some cases, Austronesian structural characteristics (vertex 1).

The creole (re-)classification of many Austronesian languages means that we should reconsider what the 'minimal requirements' are to be counted as Austronesian, or that we should more readily consider creole languages as members of the language families of their (primary?) lexifiers. This would however miss the point that such languages have multiple antecedents, and that a classification that truly reflects a language's social history should include information about the different sources of the different modules of that language. It also misses the point that showing *some* evidence for inclusion in a particular language family does not imply that membership in that language family is the only possible classification of that language, and that the reality, as implied by the positions exemplified in Figure 3 and Map 13, is that there are intermediate positions in all classifications.

The more isolating languages present a particularly interesting perspective on the nature of the dispersal of Austronesian linguistic structures across Island Southeast Asia. We know that a strongly isolating morphosyntactic character is not an archaic Austronesian characteristic. We have described three possible broad social scenarios that might underlie the more extremely isolating languages (located at vertices 1–2, and some of those in 3 in the cube); these languages are implicated in creolisation, heavy substrate influence, and/or misclassification. If we are dealing

with creolisation scenarios we can simply appeal to the well-known process of simplification due to imperfect second-language acquisition by adults. If we are discussing LCAs that are 'really' (or, 'better classified as') non-Austronesian, or if we are supposing a heavy substrate, then we have to account for the repeated emergence of isolating characteristics in the Austronesian languages across Island Southeast Asia, but not outside this region.

Of course, the important point is not that this is a characteristic of the *Austronesian* languages of the region, but rather a characteristic of languages of the broader Southeast Asia region, including a tendency for languages to show isolating behaviour (as described in Section 2), which fades towards greater morphological complexity as the languages approach New Guinea (in the east) and South Asia (in the west).[13] If we project the contemporary linguistic situation back, which seems reasonable, to judge from the arguments presented in Section 3, then we would have to assume that many of the languages of the region that were spoken in place prior to the dispersal of Austronesian languages also showed an isolating profile. The reasons for this can be speculated upon; it might be that sustained interaction over a long time led to continual creolisation scenarios, but that cannot be ascertained with certainty, nor is it particularly relevant to a discussion of the contemporary languages. What is relevant is the reasonable assumption that many, or at least some, of the pre-Austronesian substrata across much of Island Southeast Asia had a large number of isolating characteristics. This means that we can suppose that at least some of the more isolating characteristics of many of the contemporary Austronesian languages (some of which are described in this volume) can be attributed to contact with languages that were already strongly isolating (or, of course, that some of the contemporary languages in question *are* those strongly isolating, pre-Austronesian languages, with an Austronesian veneer applied). This in turn gives us a model that allows us to hypothesise about the observed trend towards an isolating profile in colloquial varieties of large languages in Island Southeast Asia. If there is already a (long-standing) linguistic ecology that promotes isolating characteristics, then multilingual speakers will see that as a model that might in some cases be explicitly targeted and will certainly not be avoided.

The reclassification of many of the LCAs in Island Southeast Asia, particularly in eastern Indonesia (see Map 12), as not belonging to the Austronesian family will make the language maps of that part of the archipelago more closely resemble

---

13. Delancey (2010, 2014) offers perspectives on the genesis of many of the subgroups of Tibeto-Burman that are highly congruent with the conclusions reached here. Also, see McWhorter (2011), where he identifies Southeast Sulawesi as a region of high morphological complexity in eastern Indonesia, and papers by numerous authors identifying areal features in eastern Indonesia.

those drawn for mainland Southeast Asia in terms of topological complexity (see Map 9),[14] and in terms of mixed languages that defy conventional classification (e.g., Jiamao, as discussed by Thurgood 1992; Norquest 2015). The same patterns of intermingled families, lack of contiguous territories, and extensive bilingualism (leading to much shared vocabulary and many shared structural characteristics) are common between mainland Southeast Asia and Island Southeast Asia. While the written traditions of the two regions are different, the other points of difference noted in Section 3, namely those involving different written and political traditions, and differences in the description of the distribution of language families, can in part be the result of different traditions and methodologies in the practice of linguistic classification rather than reflecting deeply-rooted social differences between Mainland and Island Southeast Asia.

Acknowledgements

Funding

References

Bellina, Beatrice, Bacus, Elizabeth A., Pryce, Thomas O. & Wisseman-Christie, Jan. 2010. *50 Years of Archaeology in Southeast Asia: Essays in Honour of Ian Glover*. Bangkok: River Books.

Bellwood, Peter. 2005. *First Farmers: The Origins of Agricultural Societies*. Oxford: Blackwell.

Bellwood, Peter. 2007. *Prehistory of the Indo-Malaysian Archipelago*. Canberra: ANU E. Press. <http://epress.anu.edu.au/titles/prehistory-of-the-indo-malaysian-archipelago> (1 June 2020). https://doi.org/10.22459/PIMA.03.2007

Bellwood, Peter & Glover, Ian. 2004. *Southeast Asia: From Prehistory to History*. Abingdon: Routledge.

Bellwood, Peter, Fox, James & Tryon, Darrell (eds). 1995. *The Austronesians: Historical and Comparative Perspectives*. Canberra: Department of Anthropology, Australian National University. <http://epress.anu.edu.au/titles/comparative-austronesian-series/austronesians_citation> (1 June 2020).

---

14. This is not to imply that many of these languages could be reliably classified *with* known non-Austronesian language families, merely that they could be classified as *not* displaying adequate evidence for an Austronesian classification.

Bickel, Balthasar & Nichols, Johanna. 2002. Autotypologizing databases and their use in field-work. In *Proceedings of the International Language Resources and Evaluation Workshop on Resources and Tools in Field Linguistics, Las Palmas, 26–27 May 2002*, Peter Austin, Helen Dry & Peter Wittenburg (eds). Nijmegen: ISLE and DOBES.

Bickel, Balthasar & Nichols, Johanna. 2011. Fusion of selected inflectional formatives. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 20. Munich: Max Planck Digital Library. <http://wals.info/chapter/20> (21 July 2013).

Biggs, Bruce. 1965. Direct and indirect inheritance in Rotuman. *Lingua* 14: 383–415. https://doi.org/10.1016/0024-3841(65)90053-7

Blust, Robert. 1992. On speech strata in Tiruray. In *Papers in Austronesian Linguistics*, Vol. 2, Malcolm D. Ross (ed.), 1–52. Canberra: Pacific Linguistics.

Blust, Robert. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.

Borschberg, Peter. 2010. *The Singapore and Melaka Straits. Violence, Security and Diplomacy in the 17th Century*. Singapore: National University of Singapore Press. https://doi.org/10.2307/j.ctv1ntgc9

Bulbeck, David. 2008. An integrated perspective on the Austronesian diaspora: The switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. *Australian Archaeology* 67: 31–51. https://doi.org/10.1080/03122417.2008.11681877

Coedes, Georges. 1968. *The Indianized states of Southeast Asia*, ed. by Walter F. Vella, transl. by Susan Brown Cowing. Canberra: Australian National University Press.

Comrie, Bernard. 2011a. Alignment of case marking of full noun phrases. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 98. Munich: Max Planck Digital Library. <http://wals.info/chapter/98> (20 August 2013).

Comrie, Bernard. 2011b. Alignment of case marking of pronouns. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 98. Munich: Max Planck Digital Library. <http://wals.info/chapter/98> (20 August 2013).

Cristofaro, Sonia. 2011a. Purpose clauses. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 125. Munich: Max Planck Digital Library. <http://wals.info/chapter/125> (25 August 2013).

Cristofaro, Sonia. 2011b. 'When' Clauses In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 126. Munich: Max Planck Digital Library. <http://wals.info/chapter/126> (25 August 2013).

Cristofaro, Sonia. 2011c. Reason clauses. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 127. Munich: Max Planck Digital Library. <http://wals.info/chapter/127> (25 August 2013).

Dahl, Östen & Velupillai, Viveka. 2011a. The past tense. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 66. Munich: Max Planck Digital Library. <http://wals.info/chapter/66> (25 August 2013).

Dahl, Östen & Velupillai, Viveka. 2011b. The future tense. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 67. Munich: Max Planck Digital Library. <http://wals.info/chapter/67> (25 August 2013).

Delancey, Scott. 2010. Language replacement and the spread of Tibeto-Burman. *Journal of the South East Asian Linguistics Society* 3(1): 40–55.

Delancey, Scott. 2014. Creolization in the divergence of the Tibeto-Burman languages. In *Proceedings from the 16th Himalayan Languages Symposium*, Nathan Hill & Tom Owen-Smith (eds), 41–70. London: School of Oriental and African Studies.

Denham, Tim & Donohue, Mark. 2012a. Reconnecting genes, languages and material culture in Island Southeast Asia: Aphorisms on geography and history. *Language Dynamics and Change* 2: 184–211. https://doi.org/10.1163/22105832-20120207

Denham, Tim & Donohue, Mark. 2012b. Lack of correspondence between Asian-Papuan genetic admixture and Austronesian dispersal in eastern Indonesia. *Proceedings of the National Academy of Sciences of the United States of America* 109(39): E2577. doi:10.1073/pnas.1209616109 (2 June 2020).

Donohue, Mark. 2005. Word order in New Guinea: Dispelling a myth. *Oceanic Linguistics* 44(2): 527–536. https://doi.org/10.1353/ol.2005.0033

Donohue, Mark. 2007. Word order in Austronesian: From north to south and west to east. *Linguistic Typology* 11(2): 351–393. https://doi.org/10.1515/LINGTY.2007.026

Donohue, Mark. 2013a. Towards a Papuan history of languages. *Language and Linguistics in Melanesia* 31(1): 24–41.

Donohue, Mark. 2013b. Who inherits what, when? Contact, substrates and superimposition zones. In *Language Typology and Historical Contingency. In honor of Johanna Nichols* [Typological Studies in Language 104], Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds), 219–240. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.104.09don

Donohue, Mark & Denham, Tim. 2010. Farming and language in Island Southeast Asia: Reframing Austronesian history. *Current Anthropology* 51(2): 223–256. https://doi.org/10.1086/650991

Donohue, Mark & Denham, Tim. 2011. Language and genes attest different histories in Island Southeast Asia. *Oceanic Linguistics* 50(2): 536–542. https://doi.org/10.1353/ol.2011.0026

Donohue, Mark, Denham, Tim & Oppenheimer, Stephen. 2012a. Uncoupling inheritance and diffusion: A lexical-based methodology detects social distance. *Diachronica* 29(4): 502–522.

Donohue, Mark, Denham, Tim & Oppenheimer, Stephen. 2012b. Consensus and the lexicon in historical linguistics. *Diachronica* 29(4): 538–546. https://doi.org/10.1075/dia.29.4.06don

Donohue, Mark, Hetherington, Rebecca, McElvenny, James & Dawson, Virginia. 2013. World phonotactics database. Department of Linguistics, The Australian National University. Previously available at <http://phonotactics.anu.edu.au>.

Donohue, Mark & Sawaki, Yusuf. 2007. Papuan Malay pronouns: Forms and functions. *Oceanic Linguistics* 46(1): 253–276. https://doi.org/10.1353/ol.2007.0017

Dyen, Isadore. 1956. The Ngaju-Dayak 'Old speech stratum'. *Language* 32: 83–87. https://doi.org/10.2307/410657

Enfield, Nicholas J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34: 181–206. https://doi.org/10.1146/annurev.anthro.34.081804.120406

Epprecht, Michael, Bosoni, Nicholas, Ehrensperger, Albrecht, Nagasawa, Hiroko, Lu, Juliet, Studer, David, Vollmar, Patrizia & Vilaysook, Sisoulath. 2018. *Socio-Economic Atlas of the Lao PDR: patterns and trends from 2005 to 2015*. Bern, Switzerland, and Vientiane, Lao PDR: Centre for Development and Environment, University of Bern, and Lao Statistics Bureau, Lao PDR, with Bern Open Publishing.

Foley, William A. 1998. Toward understanding Papuan languages. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia: Proceedings of the Conference, Leiden, 13–17 October 1997*, Jelle Miedema, Cecilia Odé & Rien A. C. Dam (eds), 503–518. Amsterdam: Rodopi.

Greenhill, Simon J. & Gray, Russel D. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica* 29: 523–537. https://doi.org/10.1075/dia.29.4.05gre

Haspelmath, Martin, Dryer, Matthew S., Gil, David & Comrie, Bernhard (eds). 2005. *The World Atlas of Language Structures*. Oxford: OUP.

Hill, Catherine, Soares, Pedro, Mormina, Maru, Macaulay, Vincent, Clarke, Dougie, Blumbach, Petya B., Vizuete-Forster, Mathieu, Forster, Peter, Bulbeck, David, Oppenheimer, Stephen & Richards, Martin. 2007. A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics* 80: 29–43. https://doi.org/10.1086/510412

Jinam, Timothy A., Hong, Lih-Chun, Phipps, Maude E., Stoneking, Mark, Ameen, Mahmood, Edo, Juli, HUGO Pan-Asian SNP Consortium and Saitou, Naruya. 2012. Evolutionary history of continental southeast Asians: 'Early train' hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Molecular Biology and Evolution* 29: 3513–3527. https://doi.org/10.1093/molbev/mss169

Klamer, Marian. 2019. The dispersal of Austronesian languages in Island Southeast Asia: Current findings and debates. *Language and Linguistics Compass*. Wiley Online Library. https://doi.org/10.1111/lnc3.12325

Latinis, D. Kyle. 2000. The development of subsistence system models for Island Southeast Asia and Near Oceania: The nature and role of arboriculture and arboreal-based economies. *World Archaeology* 32(1): 41–67. https://doi.org/10.1080/004382400409880

Lebar, Frank M., Hickey, Gerald C. & Musgrave, John K. 1964. *Ethnic Groups of Mainland Southeast Asia*. New Haven CT: Human Relations Area Files Press.

McAndrews, Tim, Albarracín-Jordan, Juan & Bermann, Marc. 1997. Regional settlement patterns in the Tiwanaku valley of Bolivia. *Journal of Field Archaeology* 24: 67–83.

McWhorter, John. 2011. *Linguistic Simplicity and Complexity: Why do Languages Undress?* Berlin: Walter De Gruyter. https://doi.org/10.1515/9781934078402

Munoz, Paul Michel. 2006. *Early Kingdoms of the Indonesian Archipelago and the Malay Peninsula*. Singapore: Editions Didier Millet.

Nichols, Johanna. 1998. An overview of languages of the Caucasus. <http://popgen.well.ox.ac.uk/eurasia/htdocs/nichols/nichols.html> (15 August 2013).

Noonan, Michael. 2010. Genetic classification and language contact. In *The Handbook of Language Contact*, Raymond Hickey (ed.), 48–67. Malden MA: Wiley Blackwell. https://doi.org/10.1002/9781444318159.ch2

Norquest, Peter. 2015. *A Phonological Reconstruction of Proto-Hlai*. Leiden: Brill. https://doi.org/10.1163/9789004300521

Oliveira, Nuno Vasco. 2008. Subsistence Archaeobotany: Food Production and the Agricultural Transition in East Timor. PhD dissertation, The Australian National University.

Osbourne, Milton. 2010. *Southeast Asia. An Introductory History*. London: Allen & Unwin.

Oppenheimer, Stephen. 2004. The 'Express Train from Taiwan to Polynesia': On the congruence of proxy lines of evidence. *World Archaeology* 36(4): 591–600. https://doi.org/10.1080/0043824042000303773

Pawley, Andrew. 2007. The origins of early Lapita culture: The testimony of historical linguistics. In *Oceanic Explorations: Lapita and Western Pacific Settlement*, Stuart Bedford, Christophe Sand & Sean P. Connaughton (eds), 17–49. Canberra: ANU E Press. https://doi.org/10.22459/TA26.2007.02

Pawley, Andrew & Green, Roger. 1973. Dating the dispersal of the Oceanic languages. *Oceanic Linguistics* 12: 1–67. https://doi.org/10.2307/3622852

Piper, Philip J., Hung, Hsioa-Chun, Campos, Fredeliza Z., Bellwood, Peter & Santiago, Rey. 2009. A 4000 year-old introduction of domestic pigs into the Philippine Archipelago: Implications for understanding routes of human migration through Island Southeast Asia and Wallacea. *Antiquity* 83: 687–695. https://doi.org/10.1017/S0003598X00098914

Ross, Malcolm D. 2006. Metatypy. In *Encyclopedia of Language and Linguistics*, Keith Brown (ed.), 95–99. Oxford: Elsevier. https://doi.org/10.1016/B0-08-044854-2/00199-1

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York NY: Harcourt, Brace and Company.

Scott, James C. 2010. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven CT: Yale University Press.

Sebba, Mark. 1997. *Contact Languages: Pidgins and Creoles*. London: Macmillan. https://doi.org/10.1007/978-1-349-25587-0

Shillington, Kevin. 2005. *History of Africa*. New York NY: Macmillan.

Shutler , Jr., J. Richard & Marck, Jeffrey C. 1975. On the dispersal of the Austronesian horticulturalists. *Archaeology and Physical Anthropoogy in Oceania* 10: 81–113.

Siewierska, Anna. 2011. Verbal person marking. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 102. Munich: Max Planck Digital Library. <http://wals.info/chapter/102> (25 August 2013).

Soares, Pedro, Rito, Teresa, Trejaut, Jean, Mormina, Maru, Hill, Catherine, Tinkler-Hundai, Emma, Braid, Michelle, Clarke, Douglas J., Loo, Jun-Hun, Thomson, Noel, Denham, Tim, Donohue, Mark, Macaulay, Vincent, Lin, Marie, Oppenheimer, Stephen & Richards, Martin. 2011. Ancient voyaging and Polynesian origins. *American Journal of Human Genetics* 88: 239–247. https://doi.org/10.1016/j.ajhg.2011.01.009

Solheim, William G. 1984–1985. The Nusantao hypothesis: The origin and spread of Austronesian speakers. *Asian Perspectives* 26:77–88.

Solheim, William G. 2006. *Archaeology and Culture in Southeast Asia: Unraveling the Nusantao*. With contributions from David Bulbeck and Ambika Flavel. Manila: University of the Philippines Press.

Song, Jae Jung. 2011. Nonperiphrastic causative constructions. In *The World Atlas of Language Structures Online*, Matthew S. Dryer & Martin Haspelmath (eds), chapter 111. Munich: Max Planck Digital Library. <http://wals.info/chapter/111> (25 August 2013).

Stanish, Charles. 2003. *Ancient Titicaca: The Evolution of Complex Societies in Southern Peru and Northern Bolivia*. Berkeley CA: University of California Press. https://doi.org/10.1525/california/9780520232457.001.0001

Terrell, John Edward. 2004. 'Austronesia' and the great Austronesian migration. *World Archaeology* 36(4): 586–590. https://doi.org/10.1080/0043824042000303764

Thomason, Sarah & Kaufman, Terence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.

Thurgood, Graham. 1992. The aberrancy of the Jiamao dialect of Hlai: speculation on its origin. In *Papers from the First Annual Meeting of the Southeast Asian Linguistic Society*, Martha S. Ratliff & Eric Schiller (eds), 417–433. Tempe AZ: Arizona State University.

Tumonggor, Meryanne K., Karafet, Tatiana M., Hallmark, Brian, Lansing, J. Stephen, Sudoyo, Herawati, Hammer, Michael F. & Cox, Murray P. 2013. The Indonesian archipelago: An ancient genetic highway linking Asia and the Pacific. *Journal of Human Genetics* 58: 165–173. https://doi.org/10.1038/jhg.2012.154

Wichmann, Søren. Forthcoming. Neolithic linguistics. In *Language and Prehistory of the Indo-European Peoples: A Cross-disciplinary Perspective*, Gojko Barjamovic, Irene Elmerot, Adam Hyllested, Benedicte Nielsen & Bjørn Okholm Skaarup (eds). Budapest: Archaeolingua.

Wolff, John W. 1996. The development of the passive verb with pronominal prefix in Western Austronesian languages. In *Reconstruction, Classification, Description: Festschrift in Honor of Isidore Dyen*, Bernd Nothofer (ed.), 16–40. Hamburg: Abera-Verlag.

Wright, Duncan, Denham, Tim, Shine, Denis & Donohue, Mark. 2013. An archaeological review of western New Guinea. *Journal of World Prehistory* 26(1): 25–73.
https://doi.org/10.1007/s10963-013-9063-8

Concluding reflections

John McWhorter

1. Introduction

1.1 Purpose

This volume's topic was inspired by a long-standing, friendly debate between David Gil and me, over whether languages can become isolating without the intervention of adult acquisition, sparked by Gil's thorough and fascinating response to an article I wrote about creole languages and grammatical complexity (McWhorter 2001). I am eternally thankful for the opportunity to comment on the contents of this book.

I do not think it would be useful to anyone for me to comment in a general sense on each chapter in succession, especially as I do not have career training as an Austronesianist. Rather, my purpose here is to comment on how the chapters relate to, support, or disconfirm aspects of my claim that Riau Indonesian, the languages of Central Flores, some languages of East Timor and some others of the region can be reconstructed as having undergone heavy adult acquisition despite that written records do not exist to confirm this.

As such, this chapter will cover a certain few chapters disproportionately, as they address my hypothesis most directly. However, all of the chapters will figure to some extent, and I benefitted massively from reading all of them.

1.2 Isolating languages as unnatural

When a linguist specialises in languages that are analytic or isolating, they may have little occasion to consider how extraordinary, in the literal sense of the term, such languages are. One can even internalise a sense that analyticity or isolating structure is the norm of how languages are structured and that synthetic languages are departures from it. I know this as a specialist in creoles, for example. It might seem that a language like Mandarin, Keo, Riau Indonesian, Tokodede, or my language of specialty Saramaccan Creole, is in a certain sense the norm, while languages like Russian and Navajo are inflected beyond said norm. I openly admit that as an

English speaker I am, in the depths of my cognition, the segment unaffected by linguistic training, subject to this impression.

Yet while one might ask "Why is Russian so inflected?", just as apt a question would be "Why is Mandarin *not* inflected?" This is because, as Donohue & Denham get across so aptly in this volume, analyticity and isolating structure in the cross-linguistic sense are rare. The linguistics textbook describes the isolating Mandarin, agglutinative Turkish, fusional French, and polysynthetic Mohawk alongside one another, with an implication that the world's 7000 languages might be divided almost evenly between these types. However, the reality is that the isolating language is an oddity – Donohue & Denham's Map 8 is iconically useful here – to the point that as they state outright, it is "unnatural."

There would seem to be, for example, no indigenous languages of North or South America of the truly isolating profile. That covers one entire hemisphere. Then, there are no such Indo-European, Uralic, or Afro-Asiatic languages reported, nor any of the Caucasus. Dravidian and Altaic include no such languages, nor is any "Paleosiberian" language anything less than massively complex inflectionally. No indigenous Australian language is isolating. In Africa, Nilo-Saharan languages and Khoi-San languages are all inflected (some Khoi-San more than others, but none have the typology of Riau Indonesian), as are all of the roughly 1000 Niger-Congo languages except a mere handful spoken on or near the Bight of Benin. Then, in Austronesian, neither Formosan, Philippines, or Oceanic languages are isolating, nor are most Western or Central Malayo-Polynesian languages. The languages of Papua are, as a whole, highly inflected. This covers the entire globe; I have of course left out isolates, but I am aware of none that are isolating.

Only a small number of languages of Papua (Paauw 2007) are isolating, all spoken in the northwest rather than distributed by chance, and they are joined by Chinese, most languages of Southeast Asia, many dialects of Malay/Indonesian, and a few languages of Flores and Timor. Combined with the handful of isolating Niger-Congo languages, we are considering at most a few hundred of the world's 7000 languages. Upon this, most of that few hundred – i.e. the Southeast Asian ones of the Austroasiatic, Tai-Kadai and Hmong-Mien families – are agreed by most specialists to have acquired this profile via contact with Chinese. That is, under this analysis they did not become isolating just by chance.

1.3   Grounds for my hypothesis

Thus my more specific question is whether a mere few dozen isolating languages – i.e. those without the aforementioned history of contact with Chinese – can be classified as a normal result of language change uninterrupted. It is incontestable

that languages of this kind *can* emerge via pidginisation and creolisation. My claim is that given the rarity of such languages otherwise, we assume that the few Niger-Congo ones and Chinese, the colloquial Malays, the isolating languages of Flores and Timor, and the "Papuan" examples are also products of adult acquisition.

Syntheticity, I suggest, is the inevitable state of human language over time. I propose that not out of some sense of Indo-European languages as "normal," but based on an assumption that cross-linguistically, grammaticalisation and compounding are universal and unceasing, and that therefore, isolating structure can only emerge via adult acquisition.

My motivation here is one of scientific economy. Malay is a long-time lingua franca. The few isolating Niger-Congo languages, the Chinese languages, and the languages of Central Flores harbor many features *other than loss of affixation* that are diagnostically typical of adult acquisition (McWhorter 2016, 2019). These facts seem to suggest a unified explanation. I lack an explanation for the Papuan cases at present, but scientific economy, again, coaxes me to try to bring them under a single explanatory umbrella. (It bears mentioning that both Scott Paauw ([May 2007 p.c.] and William Foley [September 2019, p.c.] have found my approach plausible.)

Donohue & Denham take the same approach, concluding that Austronesian languages of the isolating profile are the products of what they designate as, admitting the imprecision of the term, "creolization." Klamer in this volume addresses the isolating character of Alorese with the same framework. It is considered beyond discussion that Tetun Dili's especially isolating character resulted from its use as a lingua franca. Tetun Terik is much less isolating, to a degree that would be less likely to motivate an adult acquisition analysis (cf. Van Klinken 1999). Williams-Van Klinken & Hajek's chapter treats this analysis of Tetun Dili as uncontroversial, as do all sources I am aware of, presumably because Tetun Dili's use as a lingua franca is observable in the present day.

Yet various authors in this volume counter that isolating typology can emerge not only via adult acquisition, but by chance, via ordinary grammar-internal processes. More specifically, while they concur with Donohue & Denham's and Klamer's analyses, they find my application of their same approach to colloquial Malay/Indonesian, the languages of Central Flores, and some languages of East Timor scientifically mistaken.

I have learned a great deal from these authors' work despite its opposition to mine. However, I am not convinced that they have refuted my approaches, and I will address them in turn.

2. Riau Indonesian

For twenty years at this writing, David Gil and I have carried on a bracingly civil debate over the origins of Riau Indonesian. David formerly argued that its isolating structure emerged as a natural, grammar-internal development, but more recently has proposed that Riau Indonesian is but one manifestation of a "Mekong-Mamberamo" Sprachbund within which isolating structure is one trait – displacing the grammar-internal account of isolating structure to the entire Sprachbund, I assume. I, however, argue that a grammar can only attain the state of Riau Indonesian through the interrupted transmission of adult acquisition.

Gil's presentation of his position in this volume is his most comprehensive since 2001. I find it replete with fascinating linguistic data and sociohistorical information, eminently fair and attentive to all of my writings on the subject, overflowing with careful, serious thought – and ultimately unconvincing.

2.1 Proto-Malayic and Proto-Malayo-Polynesian affixation is fatal to the Mekong-Mamberamo scenario

A sterling sign of Gil's integrity as a scholar and colleague is that he is honest enough to point out something which decisively cripples both his proposition of a Mekong-Mamberamo Sprachbund and the proposition that Riau Indonesian represents the original state of Malay.

Namely, we must ask why, if Austronesian entered this Sprachbund and immediately took on the isolating Mekong-Mamberamo profile, Proto-Malayic had no fewer than 11 affixes traceable further back to Western Malayo-Polynesian or even Malayo-Polynesian itself. To this we must add nine other affixes which may have been free forms at some point within these earlier stages – but may also not have been (they may have only been free in Proto-Austronesian). These earlier stages of Malayic evolution existed millennia into the period when the Mekong-Mamberamo isolating tendency supposedly already reigned over the languages, and are hardly plausible as the source of dialects like Riau Indonesian.

Gil fashions a clever strategy to get past this problem, reconstructing Proto-Malayic as once an analytic Mekong-Mamberamo language which then borrowed all of these affixes. He goes on to note various cases of affix borrowing among languages and dialects of the area. To be sure, claims that morphology is rarely borrowed have been vastly overstated: for instance North Germanic lost a group of derivational morphemes which the Mainland Scandivanian languages then borrowed back from Low German.

However, Gil appeals to borrowing to a degree that I suspect most evaluators will find to strain plausibility. The proposal entails that Malayic first lost its affixes because of Sprachbund pressure, but that subsequently some varieties, including the one that became the standard, somehow borrowed not just an affix or two, but an entire dozen or more, which only by chance left it indistinguishable from an ordinary descendant of the Proto-Austronesian ancestor. Meanwhile, for some reason the colloquial dialects did no such thing. (I apologise if I am misinterpreting Gil's scenario.)

For all of the erudition and reasoning power that this proposal entails, I find it *ad hoc*. Rather, Proto-Malayic's and Proto-Malayo-Polynesian's syntheticity are fatal to the ideas of Austronesian shedding its affixes and "going Mekong-Mamberamo," as well as to the idea of colloquial Indonesian analyticity as an original state.

2.2   Why is Malayic so modestly inflected overall?

I claim that colloquial Indonesian's isolating character resulted from Malay/Indonesian's use as a lingua franca; Gil situates that character within his Sprachbund proposal. In support, he notes that Malayic languages in general are markedly modest in grammatical complexity, suggesting the operation of a factor broader than that which has affected only Malay/Indonesian. The question is why Minangkabau and Iban, for example, are barely less elaborated than Malay, and the question becomes even more urgent with Crouch Revington's demonstration that colloquial Minangkabau's verbal affixation is much less obligatory than the standard variety described in written sources (such as Moussay 1981 which I referred to in earlier work).

I have elsewhere suggested (2007: 240–241) that Malay has affected the surrounding languages (including Javanese; see below) over time, creating a kind of Sprachbund in which analyticity is a shared trait. However, another way of approaching this issue is to conceive of the Malayic languages as a single "stock" rather than as separate "languages." The difficulty of drawing lines between them is notorious, and the variation among them would seem to be akin to that between varieties of the "language" called German, and less than that between the varieties of "language" called Italian. Crouch Revington's chapter is relevant here, in demonstrating the porousness of the boundary between Minangkabau and Indonesian. With different imperatives, it would hardly be implausible to imagine scholars arguing that the notion of the Malayic varieties as separate languages is a colonial imposition, at odds with a more complex, dynamic reality involving only continua and fuzzy categories.

Under this analysis, it is less the single variety Malay than the single variety Malayic that has been affected by adult acquisition. A great many dialects of Malayic – some having come for sociohistorical reasons to be called by names other than "Malay" – have co-existed for millennia in intimate relationships amidst constant population movements. The rampant affix-borrowing that Gil points to for other reasons would seem to confirm the roiling intimacy over time and space of the relationship between these varieties.

As such, the adult acquisition of importance within the Malayic orbit would have been among speakers of related varieties, adjusting to the differences between them, the result being isolating varieties not only of "Malay/Indonesian" but Minangkabau, Iban, and other varieties. To wit, Malayic overall, not just "Malay," would have been streamlined by what linguists elsewhere term koineization. That concept has not been widely applied to Malayic, but I propose bringing Malayic into the traditional language contact orbit here.

In such koineization circumstances, speakers of equally elaborated varieties create less elaborated lingua franca varieties, in which forms of similar function but different form between two varieties are often eliminated in the interlanguage that results. The analogy would be the encounter between Old Norse and Old English, where two languages with three genders and rich verbal affixation yielded modern English with neither, or Kituba, the product of interactions between speakers of various dialects of Kikongo, which has vastly less affixation than other Bantu varieties. This analysis has the advantage of, once again, appealing to generally known processes of language change and contact rather than novel speculations.

Gil's alternative is to imagine Riau Indonesian as paralleling the emergence of Daghestani Russian or interlanguage between Czechs and Slovaks, where second-language usage (in the Russian case) and mutual adjustment (in the Slavic case) do not eliminate grammatical machinery. I have riposted that the Russian case differs in that it is taught in school; I might add today that one must also include the more general factor of high rates of literacy and the especially deep penetration of the written language into the culture on all levels. Gil now adduces reports that Daghestani Russian children already speak "some" Russian before they go to school, suggesting that education is not a distinguishing factor. However, I meant not that Daghestanis only encounter Russian in the schoolroom, but that school – and literacy and the degree of saturation of the written language into daily life – exposes them to the *standard* language. The emergence of the colloquial Indonesians would be analogous to what Daghestani Russian would be like if children never had gone to school – i.e. "some" Russian – and in a culture where writing played a relatively marginal role in daily life. (Of course, I refer to Indonesia of the past, not the modern nation.)

As for Czech and Slovak not yielding a reduced interlanguage version, it remains to be seen whether this would happen under circumstances of low literacy

and ongoing immediate need – which is likely never to happen. More to the point, Gil's presentation neglects the very existence of koine varieties like Fiji Hindustani (where speakers of Hindi dialects of similarly, but differently, elaborated morphosyntaxes met and "undid" them into "creolish" structures [Siegel 1987]) and Kituba, and by extension, the emergence of even Mainland Scandinavian languages, born of an encounter between Low German and Old Norse, or English.

2.3 If Riau Indonesian is a Sprachbund language, why is it so unmixed?

If today's colloquial dialects of Indonesian were just ordinary results of language contact, then based on the Uniformity Principle, we would expect that *all* of them would be language mixtures.

That is, all of them would be like Baba Malay, Uruk Lawoi and other indisputably hybrid varieties of Malay/Indonesian. Instead, there are dozens of Malay/Indonesians that are saliently "simpler," in Gil's terminology, than the standard, but harbor little or no transfer from indigenous languages. Papuan Malay is not a blend of Malay and languages like Hatam. Unsurprisingly one can find some influence from languages like Hatam in it – but nothing on the order of the copious and undeniable Hokkien traits in Baba Malay. Papuan Malay is, mainly, just streamlined Malay. This contrasts crucially with the rest of the world, where, as language contact researchers endlessly remind linguists, rampant grammatical mixture is the norm.

So – even if all of these Malay/Indonesian dialects are the way they are because they have converged on Mekong-Mamberamo traits, then via the Uniformity Principle we would expect all of them to *also* have incorporated traits from the indigenous languages their speakers also use, or have used. So often, however, they do not. Instead, they are simply Malay/Indonesians with much less of the machinery of the standard, and largely just that. If these dialects are, as Gil proposes, the original state of Malay, their unmixed character is utterly confounding. What would make these speech varieties so mysteriously impermeable to the influence of other languages for so very long?

My thesis spares us the question. Papuan Malay is not a Hatam-infused Malay to any significant degree because it emerged as incompletely acquired Malay passed on to future generations. No documentation can confirm this, but we can know, based on how Sprachbunds work, that if the issue were merely Malay/Indonesian interacting with Hatam over time, then the result would be a "Hatam Indonesian," blending traits of both languages. That language doesn't exist, because what happened instead was Hatam speakers either learned Malay/Indonesian incompletely, or learned Malay/Indonesian from the basis of a pidgin Malay *à la* Bazaar Malay.

2.4 Are there actually dialects of other Indonesian languages as structurally reduced as Riau Indonesian?

I observe that one indication that Riau Indonesian and similarly isolating Malay/Indonesians are products of adult acquisition is that there do not seem to be varieties of other Malayic (or even Indonesian) languages of such a high degree of analyticity. Malay's long-term status as a lingua franca, I argue, provides an explanation for why so many of its dialects are so isolating.

Gil ripostes that there are in fact similarly isolating varieties of other Indonesian languages, and in this volume, Conners on Javanese and Crouch Revington on Minangkabau demonstrate the point.

2.4.1 *Javanese*

With Conners the issue of degree is key. His point is well-taken that we ought to think of typical Javanese not as the standard variety but as the less inflected peripheral dialects. However, they clearly do not represent anything approaching the stunningly isolating nature of Riau Indonesian. They are modestly inflected varieties, which seem to have shed some of the equipment of standard Javanese via ordinary processes of phonetic erosion, overgeneralization and analogy.

Then, Conners argues that these dialects represent Javanese's original state and that the standard alone took on various inflectional complexities. This, however, takes a page from Gil's proposition of this kind about Malayic, which I have addressed critically above. We must ask just why this and only this dialect complexified in this way, and are unlikely to find an answer.

In contrast, an alternate argument would be that Standard Javanese conserves antique features because of the retardative effects on change of print, literacy, and prestige, which the peripheral dialects have been less subject to. Again, this is a well-attested source of difference between standard and nonstandard varieties of languages. In contrast, the scenario where a standard variety just happens to take on inflectional complexities by chance while other dialects do not is unfamiliar in language contact theory.

Of course the question is why the colloquial Javanese dialects favored loss over gain so much. Here we might allow that there is a Mekong-Mamberamo Sprachbund effect – but then Javanese is also spoken contiguously to Malayic. More to the point, Javanese has been included in a clade as part of Nothofer's (1975) "Malayo-Javanic." A possibility is that the "Malayic" tendencies have occurred within an even larger complex, due to rampant bilingualism between Malayic and Javanese over the past two millennia. Here, it is relevant that sociohistorical work has analyzed "Malayo-Javanic" as not just a linguistic grouping but a long-standing cultural one: Landmann (2017) is especially enlightening here. The question would

be: is it an accident that Javanese dialects, spoken alongside and bilingually with Malayic ones for a very long time, have taken on this profile while languages further away like Lamaholot and Kambera have not?

2.4.2 *Minangkabau*

Crouch Revington, too, adopts a framework of Gil's that I question. Standard Indonesian's voice marking is grammatically conditioned; Riau Indonesian's is conditioned by semantics. Gil more recently has titled the latter case a "Sundic" typology, and argues that it was Indonesian's original state, later developing into the grammatically conditioned "Indonesian" typology.

It is certainly possible that semantically conditioned marking can become grammatically conditioned over time. However, I doubt that it happened in Indonesia, for three reasons:

1. Malayo-Polynesian reconstruction so strongly suggests that the grammatically conditioned voice marking was primary. The issue is not just the reconstructed affixes – which could presumably have been used "Sundically" from the outset – but that they are used grammatically in the Phillipines, in Sulawesi, and so often in "Indonesian-type" languages beyond Malayic so consistently. Diachronic principle teaches us to reconstruct this as the original state.
2. The idea that this reconstruction is irrelevant because Malayic incorporated its affixation only via latter-day borrowing founders upon the abovementioned implausibility of this very scenario from a theoretical perspective.
3. Crucially, Gil elides a component of the issue: the contrast here is not solely between the grammatical and the semantic, but the obligatory and the optional. I have argued (first in McWhorter 2007: 226–228) that it would contravene diachronic directionality to suppose that voice-marking began as obligatory and then became optional. Grammaticalization obligatorifies (Lehmann 1985); adult acquisition undoes. The optionality of the affixation in "Malayic" varieties, then, is typical of adult acquisition, having all of the marks not of grammar-internal development but of "undressing."

In this light, Crouch Revington's data fit more gracefully into the conception of "Malayic" as a single entity that I mentioned above. Minangkabau would not be a distinct Malayic language following a natural pathway from "Indonesian" to "Sundic," but one of several varieties of a single complex of varieties called "Malayic," in which obligatory (and pragmatic) marking has been transformed into optional (and semantic) marking in the same fashion as it has among other Malayic varieties such as Riau Indonesian, of the kind which so many Minangkabau speakers have used alongside their native variety.

2.4.3   *Other cases?*

Gil notes other cases of purportedly isolating language varieties of Indonesia. However, over the years, Gil's conception of what isolating structure consists of has evolved considerably.

Gil first drew attention to Riau Indonesian as a speech variety unusually light on affixation of any kind, to the point of questioning there being any meaningful distinction between constituent classes. This is different from a language that simply tends to have a high percentage of monomorphemic words, a metric he uses here often on the basis of small collections of sentences. After all, a passage of English can have quite a high percentage of monomorphemic words, and yet the language has vastly too much affixation to be considered isolating.

Along those lines, of the languages Gil notes, few linguists would concur with the idea that Acehnese, Madurese, or Maanyan, all represented by substantial grammatical descriptions, resemble Riau Indonesian in degree of isolating structure. All three have rich affixal batteries, obligatory to a degree foreign to Riau Indonesian, and including a degree of inflection.

Rejang is only scantily described at present, but even in its few pages, McGinn (1982) reveals a language with infixation and other affixes as well as richer morphophonemic processes than anything in Riau Indonesian. The Nasal language source (Anderbeck & Aprilani 2013) includes a mere 30 brief sentences rather broadly glossed, hardly a basis for a claim that the language is as isolating as Riau Indonesian.

As for Kenyah and Onya Darat, Gil shows two sentences with one-to-one word/morpheme mapping. However, he admits that the languages have "some" affixes. Here, a general theme in this volume of the dangers of how we represent languages on the basis of scanty data is germane. Consider this from English:

> Twinkle, twinkle, little star.
> How I wonder what you are.
> Up above the world so high,
> Like a diamond in the sky,
> Twinkle, twinkle, little star.
> How I wonder what you are.

We would not deem English as "isolating" on the basis of this sample.

2.5   A note on the Jambi varieties

I have specified that a refutation of my hypothesis would be:

> several nonstandard Malay varieties spoken in regions with a long tradition of interethnic mixture, that have ample overspecifications, structural elaborations, and/or irregularities absent in the standard. I refer to a hypothetical Malay in which, perhaps, there are obligatory subject-marking prefixes, a three-way distinction in demonstratives, a good dozen numeral classifiers in regular use, and imperative affixes.

Gil proposes that various Malayic varieties of Jambi province constitute this refutation, with their words occurring in absolute/oblique pairs determined by numerous phonological rules and used according to various syntactic constraints.

I am hardly unaware of varieties like Kerinci, which I have often cited as an example of what otherwise seems anomalously absent in colloquial Malay varieties. That is, the Jambi varieties give a glimpse of what the uninterrupted rendition of Malayic would be like.

Just a glimpse, however. The very fact that I have often referred to Kerinci demonstrates that I do not consider it an example of my hypothetical refutation above. Rather, the Jambi varieties display a single trait, in the absolute/oblique distinctions, that qualifies as the exception that proves the rule. This trait is not joined by other developments that would make them more like languages of the Philippines or Sulawesi, such as the ones I hypothesised in the passage above.

Claims that my dismissal here is arbitrary would be understandable, but then Gil himself refers to the Jambi varieties constantly in his own piece *as typical "Mekong-Mamberamo" languages beyond this one novel feature*. Even he would have to admit, then, that these varieties are typically isolating colloquial "Malayic," with the exception of a single, albeit fascinating, morphophonemic quirk. The general pattern amidst colloquial Malay/Indonesian remains, and our issue is the pattern, not the quirk.

2.6   An alternative story

Gil leaves it to speculation why the Mekong-Mamberamo typology is so isolating. Yet it must be clear that this Sprachbund's defining features are not a disparate assemblage in the fashion of the Balkan Sprachbund. Rather, 8 of the 17 features are ones typical of isolating languages, such as low differentiation of adnominal attributives, weakly developed grammatical voice, low grammatical-morpheme density, optional thematic role-flagging, optional TAM marking, SVO order, and what Gil titles "verby adjectives."

That is, a linguist unaware that this list of features was intended as a Sprachbund description might mistake it as mostly a characterization of the isolating language, in the same way that a description of a landscape as having wet soil, droplets on tree branches, a damp smell in the air, and worms emerging from the soil would be readily treated as indicating that it had just rained, rather than as a description of a happenstance set of circumstances.

Gil's position is that the reasons this isolating profile emerged are likely unrecoverable, and it would seem that he considers this relatively unimportant in the grand scheme of these issues. The epistemological grounding of this indifference may be, if Gil's earlier works (1994) are an indication, his skepticism of the Chomskyan framework and its assumption that (overt) inflection of various kinds is inherent to the language faculty. For Gil, it is in a way "convenient" that even without a break in transmission, isolating languages could emerge via ordinary processes of transformation, perhaps recapitulating what the "natural" state of language is.

Whatever the wisdom of this skepticism of Chomskyan syntax and the assumption that language emerged as isolating (both of which I share), it is quite compatible with a proposition that nevertheless, isolating typology is only due to adult acquisition. To wit, inflection-rich languages hardly need be the product of innate neurological specifications for inflection, and syntactic theory moves ever further away from any such assumption by the year. My assumption is that inflection was an inevitable but posterior development, occurring after language hit the ground complete with its genetic specification, and thus cannot be founded upon equipment evolved to produce and process inflectional morphology (McWhorter 2008a).

There certainly is, however, a strong tendency towards isolating structure to various degrees in the zone Gil deems a Sprachbund. I see this as the product of two developments, rather than as the manifestation of a single Sprachbund. Donohue & Denham's observation of the high degree of language contact between families in this area is well-taken, but I seek somewhat more specificity, in that this kind of contact has occurred between families in, for example, the Amazon and much of North America without significant affixal erosion.

First, Chinese was rendered isolating via adult acquisition and then spread its isolating character into Hmong-Mien and then into Southeast Asia via language contact. That Chinese was the culprit seems clear from the fact that just where its speakers did not reach in the south, the Aslian Austroasiatic languages are still synthetic (despite that some of their affixation is from Malay). Even in Hmong-Mien, there are more remnants of earlier affixation in the Hmongic languages, which are less affected by Chinese (Ratliff 2006).

It remains to account for the southerly area of what Gil treats as a Sprachbund. Proto-Austronesian appears to have been rather modestly inflected. I am agnostic as to whether the reason for this was chance or external intervention – my claim is not that any degree of inflection less than Tsez or Navajo's (to choose what Gil regards as my favorite examples!) signals adult acquisition.

However, this modest degree of inflection meant that when adult acquisition indeed happened, it left a higher degree of analyticity than it did in, for example, Algonquian or Nilo-Saharan. One result of this was Malay itself, as the result of its use as a lingua franca. Another result was when this language was, in turn, acquired heavily by adults – or koineised among speakers of related dialects of it and/or other Malayic varieties – in various locations, yielding dialects so analytic as to seem (superficially) "like creoles." Central Flores and East Timor are similar examples: the same degree of adult acquisition of a Germanic language, for example, yielded English and Afrikaans rather than Keo and Tokodede. Germanic had more inflectional morphology to start with.

Because most of these developments happened unrecorded, there is an inevitable amount of speculation necessary. Gil's and my accounts can be seen as competing detective stories of a sort. However, I venture that mine accounts for the facts better than the idea that Proto-Malayo-Polynesian lost almost all of its affixation in a particular region only for some dialects of the area – including today's standard Indonesian and Javanese – to for some reason later borrow much of it back. (I sincerely apologise if I misunderstand the reasoning here.)

I ask: why is Riau Indonesian so isolating? Gil's answer, in a sense, is "Because all the languages around it were." This would seem to merely resituate my question in time and space. If Gil can ground this proposal in cross-linguistically familiar processes of language change and contact, I will be quite receptive. This will be, however, an imposing task, and I in the meantime consider my position on Riau Indonesian and similar dialects unrefuted.

3. Flores

Elias proposes that speakers of an early Austronesian language encountered speakers of a "Papuan" language of the isolating Mekong-Mamberamo profile, whose rendition of that Austronesian language displayed structural interference from the Mekong-Mamberamo one, thereby taking on its isolating typology. This account has two serious problems.

3.1   The Mekong-Mamberamo scenario leaves more questions than answers

One is that Elias founds his account on the Mekong-Mamberamo isolating profile, the problems with which I have outlined above. As such:

> My proposal:
> Central Flores languages' isolating typology is due to a mechanism observable worldwide.
>
> Elias' proposal:
> Central Flores' isolating typology is due to transfer from a hypothetical language that harbored this typology for no specified reason.

It is unclear that Elias' account is scientifically preferable.

3.2   Why are West and East Flores languages more inflected?

Second, Elias must account for why languages westward and eastward of the Central Flores ones are so much more heavily inflected. Here, he offers only unconstrained speculation, such as that the Mekong-Mamberamo languages westward and eastward were, despite that we can never know what they were like, for some reason less isolating.

My account entails a reason that the Flores languages westward and eastward are more inflected. Namely, there is no explanation necessary, *per se*, because I propose that these languages represent a normal state of Austronesian. Only the Central Flores languages require explanation.

Again, Elias' account requires simply moving past an anomaly that I propose an actual mechanism for.

3.3   Mekong-Mamberamo traits and transferred numerals are compatible with adult acquisition

To be sure, the Mekong-Mamberamo profile includes features other than isolating ones. However, none of them affect my analysis. Numeral classifiers are indeed an overspecification under my complexity metric – but I have not claimed that Central Flores languages have no overspecification of any kind, and I stress that I do not classify them as, for example, creole languages (e.g. McWhorter 2011: 245–246). Rather, the crucial contrast is between the grammars of these languages overall in comparison to other Austronesian ones, including Central Malayo-Polynesian ones spoken contiguously. Classifiers are but one feature: the contrast beyond the classifiers between Keo and Tukang Besi, Lamaholot or even Bimanese, remains.

Then beyond this, we can hardly say that a few other features such as a dental click gesture and *sun* rendered as "eye day" constitute refutations of an argument that a language is the product of adult acquisition. These features could easily have been present in a language not especially analytic, and then survived an episode of widespread adult acquisition.

This includes the fascinating data on numerals that Elias presents. They, too, are quite compatible with an adult acquisition account, in that transfer from an earlier language via ordinary long-term bilingualism (Dixon's [1997] linguistic equilibrium) can well have happened before an encounter with different adults later. The Celtic "sheep-counting" numerals in regional Englishes that Elias uses as a comparison are in fact equally useful for my account. The numbers were transferred from Celtic, *after* which Scandinavian invaders acquired English imperfectly via adult acquisition. Klamer's treatment of numbers in Alorese exemplifies my approach – the transfer hardly refutes an adult acquisition account.

3.4   Central Flores languages are not pidginised Sulawesi ones

Finally, I am perplexed that Elias sees the absence of Sulawesi borrowings in Flores languages as counterevidence to my scenario when he later observes that Uralic profoundly impacted Russian grammatically while leaving little lexical evidence – especially when I myself have made the same observation (McWhorter 2019: 195).

Moreover, Elias supposes that I assume that Central Flores speakers learned the language of the Sulawesi invaders. This, however, was Hull's (1998) scenario, not mine, and I disavowed it explicitly starting in my earliest papers on Flores (specifying, for example, that we would expect Sulawesi-derived grammatical items in Central Flores languages upon this basis, when they are in fact absent [McWhorter 2011: 252–255]).

In earlier work, I indeed adopted Hull's observation that Central Flores lexical items seem anomalously close to Tukang Besi ones' phonetically, but I reconstructed this within an account under which the Sulawesi speakers acquired the Flores language. I speculated that Tukang Besi speakers had contributed their cognates to their rendition of the Central Flores languages. However, I have abandoned that hypothesis in later publications, and concur with Elias' objections. For one, the data does not meaningfully demonstrate the argument. In addition, that account, like Gil's that Proto-Malayic borrowed its affixes after emerging as an isolating language, is based on no generally known mechanism of language change and contact.

In sum, neither the central Flores languages' Mekong-Mamberamo traits nor their likely substrate-derived numeral systems refute my adult acquisition argument.

4. East Timor

Schapper, on languages of East Timor, takes especial issue with "the idea that these languages are simple," with the assumption that I have made such a claim. This misrepresents my approach, which addresses *relative* complexity and has in no publication of any kind designated a language "simple." I have made this clear in my articles on these Timor languages (McWhorter 2011: 245–246):

> *My claim is not that the affixless languages of Flores and the near-affixless ones of Timor are "creoles,"* to the extent that creolization represents the sharpest kind of break in transmission (Thomason & Kaufman 1988) and the most radical degree of grammatical simplification in natural language (McWhorter 2001). Affixation is but one kind of grammatical complexity; *these languages contain ample amounts of other kinds.* (Italics mine)

Importantly, here I note that acquisition of these languages has not even been as interrupted as that of creoles, languages whose own complexities, moreover, I have stressed, as in my grammatical description of Saramaccan (McWhorter & Good 2012). The idea that I have designated the languages of East Timor "simple" is misrepresentative.

My claim, then, is that certain languages of East Timor are languages of ordinary complexity which, nevertheless, display signs of adult acquisition in their morphosyntax that render them less grammatically complex than many of their sister and cousin languages.

4.1 Fossilised derivation

As such, the fact that Timorese languages often have rather large phonological inventories, grammatical metathesis, and alienable possessive marking is in no way incompatible with the analysis, as interesting and invaluable as the data Schapper has elicited is.

As Schapper herself even notes, ample derivational morphology, too, is compatible with my account, which stresses the erosion of inflectional morphology, uncontroversially treated as diagnostic of second language acquisition, creolization and related processes (cf. Pienemann 1998; Plag 2008; Matras 2009: 153–7). This includes that the erosion of derivation has left behind various phonological reflexes in Timorese languages. This is also true of Chinese, where the question nevertheless arises as to why there are no similar indications of previous inflectional, as opposed to derivational, morphemes, and motivates the reconstruction of adult acquisition in work such as DeLancey (2013) and McWhorter (2007, 2016). I merely extend that question to certain Timorese languages.

Schapper's observations actually touch upon the fact that in my work I have less stressed isolating structure – something Gil especially stresses in his work – than, more specifically, what I have termed *radical analyticity*: absence (or all but absence) of inflectional marking indicated by affixation, tone, or vowel changes in quality or length (McWhorter 2016). Radically analytic languages may have ample derivational affixation (or derivation marked by tone). The absence of inflection, specifically, is what I argue is key to adult acquisition. As such, a language with elaborate metathesis processes marking derivation still fits under my rubric if it nevertheless lacks inflectional morphology.

4.2   The inflection question

Thus it remains remarkable, despite the richness of the data Schapper presents, that Waima'a verbs are not variant according to person and number via phonetic transformations of initial consonants conditioned by the erosion of erstwhile prefixes. This is typical of other Timor and Central Malayo-Polynesian languages. It is reasonable to propose that in the Waima'a case – as well as in Naueti, Tokodede and Kemak – the prefixes were eliminated rather than gradually assimilated phonologically.

Schapper attempts to address this anomaly by noting that in Waima'a, in one instance a subject prefix was rebracketed as part of the following causative prefix, and then subject to the change that this portmanteau underwent. However, this is one affix in one language. In so many related languages, the subject prefixes have undergone transformations distinguishable from those of the following derivational prefixes.

4.3   Signs of adult acquisition in Waima'a, Naueti, Tokodede, and Kemak

Overall, even beyond the subject prefixes, Schapper's presentation leaves Waima'a, Naueti, Tokodede and Kemak plausible as products of adult acquisition, for the same reasons that the languages of Flores – or English, Swahili, Mandarin, and Yoruba – are: the difference between them and their relatives.

a.   The portmanteau subject prefix/causative in Waima'a, for example, began as a third person plural one but generalised to all persons and numbers, in contrast to a language like Galoli in which a similar portmanteau occurs in distinct allomorphs for all persons in the singular and two in the plural.

b.   Schapper's demonstration of grammatical metathesis in many Timor languages, intended to show that the languages are not "simple," is important. Yet, it remains relevant that in addition to (a) above, Waima'a, Naueti and

Tokodede do not have this feature (Kemak Marobo does, but to a lesser extent than close relatives).

c.   Waima'a and Naueti, *in addition to (a) and (b) above*, have largely lost the inalienable possessive marking, the machinery now only distinguishing third from the other persons. In its differing distinctions from language to language, inalienable possessive marking entails a degree of irregularity that must be retained. A mere person distinction is regular, and in that, less "complex" according to my metric.

d.   Naueti has just four numeral classifiers. It and Waima'a have a separate animacy distinction with numerals, but overall, this contrasts with the much richer array of noun class distinctions that Central Malayo-Polynesian languages make (including the Central Flores ones).

Crucially, Schapper does not indicate features emerging in these languages that "match" in number or complexity (according to my metric or anyone's) the ones lost. It would seem that loss has predominated considerably over gains – and not just in terms of the subject prefixes. This, I suggest, can be a sign of adult acquisition of a language.

4.4   Papuan languages

Finally, while I fully accept Schapper's correction of my depiction of the Papuan languages of Timor as highly isolating, it must be clear, for one, that I referred to these languages in passing in a mere few paragraphs (McWhorter 2008b: 178–179), and did not in any sense depict them as "lacking any features typical of 'old' languages," as Schapper claims. She criticises my comparing Fataluku to Mandarin – but to read this as a claim that Fataluku is a "simple" language would seem to suggest certain questionable assumptions about Mandarin on Schapper's part, not mine.

Crucially, however, I wrote the paragraphs in 2005, when little data was available on most of the languages. Almost all of the data Schapper cites is from her own fieldwork years later. Josselin de Jong (1937) on Oirata does not present data contradicting a basic claim that the language is starkly more isolating than the typical Papuan language – as opposed to showing that Oirata is not maximally "simple," a claim I did not dispute in 2005 anymore than I do now. My characterization of these languages was based on what was available to me at the time, such as the characterization of the languages in Hull (1998), and the fact that no linguist informed about the area questioned my assumption that these languages were as isolating as some of the Austronesian Timorese ones.

Schapper's data is invaluable. Moreover, it is quite compatible with my hypothesis. The languages are spoken in the east, and remain strikingly less morphologically

elaborate than almost all other of the several hundred Papuan languages. They suggest that the languages of the eastern half of the island were subject to heavy adult acquisition.

I also submit that Schapper's passing dismissal of Hull's (1998: 154–164) proposition that the Timorese languages were incompletely acquired by invaders from Ambon undercovers the substance of the argument that Hull made. It will remain to future evaluators to decide its plausibility; one suspects that the rarity of the publication has hindered its wider assessment.

4.5 Different paths to the same mountaintop?

The degree in these Eastern Timor Austronesian languages of analyticity, as well as loss of other grammatical features, remains striking compared to other Austronesian languages including ones of West Timor. Schapper openly acknowledges the contrast, methodically addressing the relatively low level of inflection in many of East Timor's languages, allowing that a contact account with "Papuan" languages is plausible. One might suppose that she and I are addressing the same issue in the same basic way.

Yet Schapper clearly does not think so, and what mainly seems to lead her to see my approach as incompatible with hers is an impression that I have claimed that the languages in question have outright pidgin grammars. I have not. I have stated that certain languages of Eastern Timor are highly analytic, to a degree unusual among their relatives, and that adult acquisition is the most likely cause of that difference.

I propose, to wit, that precisely what Klamer documents as having happened to Alorese due to contact with Lamaholot, and then also as even actually observed presently in contact with Adang based on Moro (2019), would have been what happened to some languages in East Timor (as well as Flores). Schapper has presented nothing incompatible with that proposal, and I stand by it.

5. A note on Chamic

Brunelle argues that evidence of Chamic varieties as products of non-native acquisition of Austronesian (Thurgood 1999) is less clear than often thought, because there is no robust evidence of heavy contact between Mon-Khmer and Austronesian speakers in Chamic's Classical period. Brunelle prefers what would seem to be an assumption shared by many in this volume: that large-scale morphological erosion may, but need not, be due to non-native acquisition.

As such, he seems agnostic as to which factor left Chamic so low on the Malayo-Polynesian derivational and inflection inheritance (cf. Ross 2002) even by

the time of its earliest inscription. Rather, he focuses on a convincing argument that Chamic's monosyllabicity, a later development, was only indirectly due to contact with Vietnamese, and was basically an internal development driven by ordinary phonotactic changes.

Obviously, Brunelle's agnosticism on the "first wave" of change in Chamic is incompatible with my hypothesis. The issue is what we accept as proof, which will always be a thorny question with sociohistorical developments that occurred without written record. However, the evidence is at least compatible with the idea that Mon-Khmer had a stronger impact on early Chamic than archaeology alone would suggest.

The large component of Mon-Khmer vocabulary is one indication, as is the fact that there is such a component in Chamic's "prodigal son" member – or at least separate but closely-related offshoot "cousin" – Acehnese spoken in Sumatra. Since Acehnese can only have incorporated the Mon-Khmer element while its speakers were still in Southeast Asia, this pushes back the Mon-Khmer contact even closer to Chamic's emergence as a group.

This lexical mixture could indeed have happened without incomplete acquisition, as it did with the French incursion into English. However, Chamic languages other than Acehnese are not only mono- or sesquisyllabic, and many are either tonal or have register contrasts. This suggests more intimate contact with Mon-Khmer languages than the lexical items do, and possibly less complete acquisition of Austronesian. An Austronesian rendered mono- or sesquisyllabically, whether by Austronesians or Austroasiatic speakers, could be analyzed as an incomplete Austronesian, and the loss of so much affixation would be evidence, when alongside this phonotactic transformation, of an abbreviated acquisition.

Moreover, a toneless L2 (i.e. an Austronesian variety) rendered with contrastive tone is a transformation to such a degree that qualifies almost by definition as incomplete acquisition of the original language amidst replacement of its lexical contrastive machinery by that of another family. It is this kind of factor that informs Thurgood's (1999) proposition that Chamic is essentially Austronesian "in" Mon-Khmer.

Brunelle is obviously correct that Vietnamese influence, specifically, on Chamic has been a latter-day effect, and there is indeed no reason to suppose that Chams have started rendering their language monosyllabic (or even sesquisyllabic) on the model of a second language they happen to speak. This would leave the question as to why Tukang Besi speakers have not been modeling their language on the much less synthetic Indonesian, or why Bantu speakers of largely analytic varieties like Kituba and Lingala have not been creating new analytic versions of their indigenous vernaculars spoken alongside.

However, the evidence can be taken as suggesting that contact with Mon-Khmer was more substantial in the past, such that Mon-Khmer speakers created an L2 version of Austronesian gradually infused with Mon-Khmer style phonology including tone, as well as at least the beginnings of Mon-Khmer phonotactics. Otherwise, we have no explanation for the contrast in morphological loss between Tagalog and Cham, between Makasar and Cham, or even between Lamaholot and Cham.

6. Conclusion

Is it a methodological weakness to seek to bring data into an overarching model, as opposed to addressing each language as an individual case? Among some linguists, claims that a single language's degree of analyticity traces to adult acquisition are well received, but a larger proposal that such a degree of comparative analyticity be treated as *diagnostic* of adult acquisition is considered untenable.

Of course one must work with caution. However, a general impatience with "models" is a more unusual position in itself than it may seem. Is it not a somewhat peculiar approach to science to distrust attempts to systematise, to make predictions?

While I have certainly worked from a "model" in my work on language contact, those who prefer to address each situation individually and distrust overarching models are also working from a "model." This latter approach, wary of patterns and stressing idiosyncrasy, is one that many scientists of other kinds would find curious.

To wit:

If:
1. languages simplify morphosyntactically amidst the heavy adult acquisition central to pidgin and creole formation;

and:
2. no languages have *been explicitly documented* to simplify this way via mere grammar-internal change;

then:
3. languages starkly less morphosyntactically elaborate than their relatives may be reconstructable as having undergone heavy adult acquisition as well.

This hypothesis may turn out not to be true, but outsiders might wonder why it would be considered careless.

Rather, the rub would seem to be with my particular "model." I sense, for one, that I may have seemed in my wording to be proclaiming a truth rather than putting

forward a hypothesis. That is my fault, and I must specify that I have intended to put forth a model with clarity, not to hubristically declare a verity.

That model is similar to the perspectives of, in this volume, Donohue & Denham and Klamer. My point is not that languages like Keo and Tokodede have no grammar. However, my proposal – not declaration – is indeed that languages do not settle into the grammatical typology of those languages without adult acquisition.

I come away from these enlightening chapters in awe of their erudition, and yet even more convinced of my position than I was before.

References

Anderbeck, Karl & Aprilani, Herdian. 2013. *The Improbable Language: Survey Report on the Nasal Language of Bengkulu*. Dallas TX: SIL International.

DeLancey, Scott. 2013. The origins of Sinitic. Increased empiricism. In *Recent Advances in Chinese Linguistics* [Studies in Chinese Language and Discourse 2], Jing-Schmidt, Zhuo (ed.), 73–100. Amsterdam: John Benjamins.

Dixon, Robert M. W. 1997. *The Rise and Fall of Languages*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511612060

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17: 179–200.
https://doi.org/10.1017/S0332586500003000

Hull, Geoffrey. 1998. The basic lexical affinities of Timor's Austronesian languages: A preliminary investigation. *Studies in the Languages and Cultures of East Timor* 1: 97–174.

de Josselin de Jong, J. P. B. 1937. *Studies in Indonesian Culture. Oirata: A Timorese Settlement on Kisar*, Vol. 1. Amsterdam: Noord-Hollandische Uitgevers-Maatschappij.

Landmann, Alexandra. 2017. Can we reconstruct a "Malayo-Javanic" law area? In *Spirits and Ships*, Andrea Acri, Roger Blench & Alexandra Landmann (eds). Singapore: ISEAS.
https://doi.org/10.1355/9789814762779-006

Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20: 303–318.

Matras, Yaron. 2009. *Language Contact*. Cambridge: CUP.
https://doi.org/10.1017/CBO9780511809873

McGinn, Richard. 1982. *Outline of Rejang Syntax*. Jakarta: Badan Penyelenggara Seri Nusa, Universitas Atma Jaya.

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(3–4): 125–156.

McWhorter, John H. 2007. *Language Interrupted*. Oxford: OUP.
https://doi.org/10.1093/acprof:oso/9780195309805.001.0001

McWhorter, John H. 2008a. Inflectional grammar and universal grammar: Post hoc versus propter hoc. In *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, Regine Eckhardt, Gerhard Jäger & Tonjes Veenstra (eds), 337–373. Berlin: Mouton de Gruyter.

McWhorter, John H. 2008b. Why does a language undress? Strange cases of Indonesia. In *Language Complexity: Typology, Contact, Change* [Studies in Language Companion Series 94], Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds), 167–190. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.12mcw

McWhorter, John H. 2011. *Linguistic Simplicity and Complexity*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781934078402

McWhorter, John H. 2016. Is analyticity normal? Implications of Niger-Congo and Sino-Tibetan for typology and diachronic theory. In *Cyclical Change Continued* [Linguistik Aktuell/Linguistics Today 227], Elly van Gelderen (ed.), 49–91. Amsterdam: John Benjamins. https://doi.org/10.1075/la.227.03mcw

McWhorter, John H. 2019. The radically isolating languages of Flores: A challenge to diachronic theory. *Journal of Historical Linguistics* 9: 177–207. https://doi.org/10.1075/jhl.16021.mcw

McWhorter, John H. & Good, Jeff. 2012. *A Grammar of Saramaccan Creole*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110278262

Moro, Francesca. 2019. Loss of morphology in Alorese (Austronesian): Simplification in adult language contact. *Journal of Language Contact* 19: 378–403. https://doi.org/10.1163/19552629-01202005

Moussay, Gerard. 1981. *La langue minangkabau*. Paris: Archipel.

Nothofer, Bernd. 1975. *The Reconstruction of Proto-Malayo-Javanic*. The Hague: Martinus Nijhoff.

Paauw, Scott. 2007. A North Papua linguistic area? Paper presented at the First Workshop on the Languages of Papua, Manokwari.

Pienemann, Manfred. 1998. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.15

Plag, Ingo. 2008. Creoles as interlanguages: Inflectional morphology. *Journal of Pidgin Creole Languages* 23: 114–135. https://doi.org/10.1075/jpcl.23.1.06pla

Ratliff, Martha. 2006. Prefix variation and reconstruction. In *Variation and Reconstruction* [Current Issues in Linguistic Theory 268], Thomas D. Cravens (ed.), 165–178. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.268.09rat

Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In *The History and Typology of Western Austronesian Voice Systems*, Fay Wouk & Malcolm Ross (eds), 17–78. Canberra: Pacific Linguistics.

Siegel, Jeff. 1987. *Language Contact in a Plantation Environment*. Cambridge: CUP.

Thomason, Sarah J. & Kaufman, Terence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*. Honolulu HI: University of Hawaii Press.

Van Klinken, Catharina. 1999. *A Grammar of the Fehan Dialect of Tetun, an Austronesian Language of West Timor*. Canberra: Pacific Linguistics, C-155.

Index

Many Austronesian languages exhibit isolating word structure. This volume offers a series of investigations into these languages, which are found in an "isolating crescent" extending from Mainland Southeast Asia through the Indonesian archipelago and into western New Guinea. Some of the languages examined in this volume include Cham, Minangkabau, colloquial Malay/Indonesian and Javanese, Lio, Alorese, and Tetun Dili.

The main purpose of this volume is to address the general question of how and why languages become isolating, by examination of a number of competing hypotheses. While some view morphological loss as a natural process, others argue that the development of isolating word structure is typically driven by language contact through various mechanisms such as creolization, metatypy, and Sprachbund effects. This volume should be of interest not only to Austronesianists and historians of Insular Southeast Asia, but also to grammarians, typologists, historical linguists, creolists, and specialists in language contact.

John Benjamins Publishing Company