

JSEALS Special Publication No. 1

ISSUES IN AUSTRONESIAN HISTORICAL LINGUISTICS



Edited by Hsiu-chuan Liao



UNIVERSITY of
HAWAII
PRESS

© 2017 University of Hawai'i Press

All rights reserved

OPEN ACCESS – Semiannual with periodic special publications

E-ISSN: 1836-6821

<http://hdl.handle.net/10524/52405>



Creative Commons License

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

JSEALS publishes fully open access content, which means that all articles are available on the internet to all users immediately upon publication. Non-commercial use and distribution in any medium is permitted, provided the author and the journal are properly credited.

Cover photo courtesy of Alexander Smith.

JSEALS

Journal of the Southeast Asian Linguistics Society

Editor-in-Chief

Mark Alves (Montgomery College, USA)

Managing Editors

Nathan Hill (University of London, SOAS, UK)

Sigrid Lew (Payap University, Thailand)

Paul Sidwell (Australia National University, Australia)

Editorial Advisory Committee

Marc BRUNELLE (University of Ottawa, Canada)

Kamil DEEN (University of Hawai'i at Mānoa, USA)

Gerard DIFFLOTH (Cambodia)

Rikker DOCKUM (Yale University, USA)

San San HNIN TUN (INALCO, France)

Kitima INDRAMBARYA (Kasetsart University, Thailand)

Peter JENKS (UC Berkeley, USA)

Mathias JENNY (University of Zurich, Switzerland)

Daniel KAUFMAN (Queens College, City University of New York & Endangered Language Alliance, USA)

James KIRBY (University of Edinburgh, Scotland)

Hsiu-chuan LIAO (National Tsing Hua University, Taiwan)

Bradley MCDONNELL (University of Hawai'i at Mānoa, USA)

Alexis MICHAUD (CNRS (Le Centre National de la Recherche Scientifique, France)

David MORTENSEN (Carnegie Mellon University, USA)

Peter NORQUEST (University of Arizona, USA)

Pittayawat PITTAYAPORN (Chulalongkorn University, Thailand)

Amara PRASITHRATHSINT (Professor Emeritus of Linguistics, Chulalongkorn University, Thailand)

Joseph (Deth) THACH (Institut National des Langues et Civilisations Orientales, France)

Alice VITTRANT (Aix-Marseille Université / CNRS-DDL, France)

The Journal of the Southeast Asian Linguistics Society publishes articles on a wide range of linguistic topics of the languages and language families of Southeast Asia and surrounding areas. JSEALS has been hosted by the UH Press since the beginning of 2017.

Contents

<i>Introduction from the Volume Editor</i>	<i>iv</i>
<i>About the Contributors</i>	<i>vi</i>
<i>Statement from the JSEALS Editor-in-Chief</i>	<i>viii</i>

Papers

<u>Mora, Vowel Length, and Diachrony: the Case of Arta, a Philippine Negrito Language</u>	<i>Yukinori Kimoto</i>	1-22
<u>Re-evaluating the Position of Iraya among Philippine Languages</u>	<i>Lawrence A. Reid</i>	23-47
<u>Reconstructing Proto Kenyah Pronouns and the Development of a True Five Number System</u>	<i>Alexander D. Smith</i>	48-66
<u>Linguistic Evidence for Prehistory: Oceanic Examples</u>	<i>Malcolm Ross</i>	67-101
<u>Classifying Old Rapa: Linguistic Evidence for Contact Networks in Southeast Polynesia</u>	<i>Mary Walworth</i>	102-122

Introduction from the Volume Editor

The International Conference on Austronesian Linguistics (ICAL) is an open scholarly forum on the languages and linguistics of the Austronesian language family. Since the initial meeting of 1-ICAL in Honolulu in 1974, ICAL has been held regularly in different continents every three to four years.

- 1-ICAL (Honolulu, Hawai'i, USA) (Jan. 2-7, 1974)
- 2-ICAL (Canberra, Australia) (Jan. 4-12, 1978)
- 3-ICAL (Denpasar, Bali, Indonesia) (Jan. 19-24, 1981)
- 4-ICAL (Suva, Fiji) (Aug. 13-18, 1984)
- 5-ICAL (Auckland, New Zealand) (Jan. 11-16, 1988)
- 6-ICAL (Honolulu, Hawai'i, USA) (May 20-24, 1991)
- 7-ICAL (Leiden, Netherlands) (Aug. 22-27, 1994)
- 8-ICAL (Taipei, Taiwan) (Dec. 28-30, 1997)
- 9-ICAL (Canberra, Australia) (Jan 8-11, 2002)
- 10-ICAL (Puerto Princesa City, Palawan, Philippines) (Jan. 17-20, 2006)
- 11-ICAL (Aussois, France) (June 22-26, 2009)
- 12-ICAL (Denpasar, Bali, Indonesia) (July 2-6, 2012)
- 13-ICAL (Taipei, Taiwan) (July 18-23, 2015)
- 14-ICAL (Antananarivo, Madagascar) (July 17-20, 2018)

As a tradition of ICAL, the host institution(s) of each ICAL would take charge of the publication of post-conference proceedings or selected papers on specific topics. Following the success of the 13-ICAL meeting held at Academia Sinica, Taipei in July 2015, the 13-ICAL Organizing Committee contacted the editorial team of the international journal *Language and Linguistics (L&L)* for the possibility of publishing selected 13-ICAL papers as a Special Issue 3 of *L&L*. However, the earliest the papers could be published was 2019, a year after 14-ICAL in 2018. Thus, in early August 2015, I contacted the editorial team of the *Journal of the Southeast Asian Linguistics Society (JSEALS)* for the possibility of publishing selected 13-ICAL historical linguistics papers through *JSEALS*. Because of the papers' focus on insular Southeast Asian languages and solid contributors, I received full support to publish selected 13-ICAL historical linguistics papers through the *JSEALS Special Publication series*.

The publication of the 13-ICAL historical linguistics volume faced a new challenge in mid-2016 because of the reorganization of academic departments in the Australian National University (ANU) in July 2016. Because *JSEALS* had been hosted in ANU cyberspace since 2009, we faced uncertainty about the continuing publication of *JSEALS* and the *JSEALS Special Publication series* from ANU. In May 2016, when this issue was raised at the *SEALS 26 Business Meeting*, I suggested that the editorial team contact the University of Hawai'i Press (UH Press) for the possibility of publishing *JSEALS* through

them. In late 2016, Mark Alves, the JSEALS editor-in-chief, informed me that the JSEALS editorial team had signed a contract with the UH Press for the publication of JSEALS. Even though most of papers for the 13-ICAL historical linguistics volume were ready for publication in late 2016, Mark and I jointly decided to postpone the publication of the 13-ICAL historical linguistics volume until the partnership between JSEALS and UH Press became effective in January 2017.

Since August 2015, I have been working closely with each contributor, reviewer, and Mark for the publication of the 13-ICAL historical linguistics papers. It is highly significant that the 13-ICAL historical linguistics volume will become the first JSEALS special publication to be published by UH Press.

As with papers in regular issues of JSEALS, all papers in this JSEALS Special Publication were reviewed by at least two specialists in the relevant subfields. I would like to thank the many scholars who served as reviewers of papers for the volume, among whom were Robert Blust, Ross Clark, Jeff Marck, Lawrence Reid, William Wilson, and David Zorc. Finally, I would like to thank each contributor, the JSEALS editorial team from 2015 to 2017 (Mark Alves, Nathan Hill, Peter Jenks, Sigrid Lew, and Paul Sidwell), and the UH Press. Without their support, the publication of this volume would not have been possible.

Hsiu-chuan Liao

May 25, 2017

Honolulu, Hawai'i

About the Contributors

The Editor

HSIU-CHUAN LIAO (hcliao@mx.nthu.edu.tw) is an Associate Professor at the Institute of Linguistics, National Tsing Hua University, Taiwan. She received her PhD in Linguistics from the University of Hawai'i at Mānoa in 2004. Her primary research interests include historical linguistics, Austronesian languages (especially Philippine languages and other western Austronesian languages), and morphosyntactic typology and theories. She serves as an Associate Editor for *Oceanic Linguistics* and is a member of the Editorial Advisory Committee for the *Journal of the Southeast Asian Linguistics Society* and is a member of the Board of Editorial Consultants for the *Philippine Journal of Linguistics*.

Authors

YUKINORI KIMOTO (yk.kimoto@gmail.com) is a postdoctoral fellow of the Japan Society for the Promotion of Science hosted by Nagoya University, Japan. He received his PhD in Human and Environmental Studies from Kyoto University in 2017 (Dissertation title: *A Grammar of Arta: A Philippine Negrito Language*). His primary research interests include Philippine languages (particularly the description of the Arta language spoken in Luzon), historical linguistics, linguistic typology, cognitive linguistics, and sociolinguistics.

LAWRENCE A. REID (reid@hawaii.edu) is an Emeritus Researcher with the University of Hawai'i and a Research Fellow with the National Museum of the Philippines. His current research activities focus on the historical development of Philippine languages, e.g. 'Accounting for variability in Malayo-Polynesian pronouns: Paradigmatic instability or drift?' (Reid 2016—*Journal of Historical Linguistics* 6: 130-164). He is also assisting in the documentation of Isinay, a Philippine language that is no longer spoken by young people, helping the elders in the development of a Community Dictionary of Isinay. His *Talking Dictionary of Khinina-ang Bontok* is available on-line.

MALCOLM ROSS (malcolm.ross@anu.edu.au) is an Emeritus Professor with the Australian National University. He specializes in Austronesian and Papuan historical linguistics, as well as wider issues in historical linguistics, particularly methodology and contact-induced change. Together with Andrew Pawley and Meredith Osmond, his ongoing work with The Oceanic Lexicon Project has contributed to a valuable reconstruction of the lexicon of Proto Oceanic (Ross, Pawley, and Osmond 1998, 2003, 2008, 2011, 2016), the language ancestral to most of the Austronesian languages of Melanesia, Polynesia, and Micronesia. In addition to the Proto-Oceanic lexicon project, his main involvements have been in various aspects of Formosan (Taiwan) and Papuan (New Guinea) historical linguistics in the last few years. He is currently building a Trans

New Guinea lexical database to see if a more thorough examination of TNG lexicon casts brighter light on the history of the family.

ALEXANDER D. SMITH (smithad@hawaii.edu) recently received his PhD in linguistics from the University of Hawai'i at Mānoa in spring 2017. His dissertation, *The Languages of Borneo: A Comprehensive Classification*, is the most complete classification of the languages of Borneo to date. His primary research interests include comparative Austronesian linguistics, historical phonology, the languages of Borneo, language documentation and description, and linguistic fieldwork.

MARY WALWORTH (maryewalworth@gmail.com) is a Postdoctoral Researcher in Linguistics in the Department of Linguistic and Cultural Evolution at the Max Planck Institute for the Science of Human History. She received her PhD in Linguistics from the University of Hawai'i at Mānoa in 2015. Her primary research interests center on historical linguistics, language contact related issues, and documentation of Oceanic languages (especially of Polynesian languages).

From the JSEALS Editor-in-Chief

This is the first JSEALS special publication since JSEALS became a University of Hawai'i Press publication as of January 2017. The goal of JSEALS special publications is to share collections of linguistics articles, such as select papers from conferences or other special research agendas, as well as to offer a way for linguistic researchers in the greater Southeast Asian region to publish monograph-length works.

In this instance, Hsiu-chuan Liao, the primary editor of this publication, requested that papers from the 2015 13-ICAL meeting be published through JSEALS, which we were able to complete in a timely manner. The five papers include works by prominent names in the field of Austronesian historical linguistics.

Subsequent special publications are already being planned. These include a collection of articles resulting from a special Southeast Asian linguistic workshop held at Chulalongkorn University and proceedings from an Austronesian linguistics conference (AFLA).

We are very pleased that JSEALS is able to contribute to the sharing of quality linguistic research in both mainland and insular Southeast Asia.

Mark J. Alves

May 26, 2017

Rockville, Maryland

MORA, VOWEL LENGTH, AND DIACHRONY: THE CASE OF ARTA, A PHILIPPINE NEGRITO LANGUAGE

Yukinori Kimoto
Nagoya University
<yk.kimoto@gmail.com>

Abstract

This paper attempts to provide an explanation for the diachronic development of long vowels in Arta, a Negrito language spoken in Nagtipunan, Quirino Province, the Philippines. In Arta, a large number of lexical roots and morphologically complex words have long vowels in them, but the items with a long penultimate vowel which are shared with other Philippine languages that retain an older accentual system are reflected as short vowels. Thus, the long vowels seen in Arta should be separated from inherited accents. It is argued that these vowels developed independently in the language by compensatory lengthening and vowel fusion, after the loss of *k, *q, and *h. Since both compensatory lengthening and vowel fusion crucially involve the principle of mora count conservation, the phonological changes which occurred in Arta indicate that the mora has played a significant role in the language.

Keywords: Arta, vowel length, mora, compensatory lengthening, vowel fusion
ISO 639-3 codes: atz

1 Introduction

One of the interesting typological features observed in many Philippine languages involves the realization of contrastive word accents (or stresses) as vowel length on the penultimate open syllable. This phonetic manifestation shows a clear typological difference from English (stress accent) and Japanese (pitch accent).¹ The following minimal pairs from Tagalog, Ilokano and Bikol illustrate the point.

- (1) Tagalog *áso* [ʔa:so] ‘dog’ vs. *asó* [ʔaso] ‘smoke’
Ilokano *bára* [ba:ra] ‘hot’ vs. *bará* [bara] ‘lung’
Bikol *bága* [ba:ga] ‘ember’ vs. *bagá* [baga] ‘truly!’

(Zorc 1993:18)

This kind of vowel-length contrast is also found among various Philippine languages, including Aklanon, Balangao, Cebuano, Hanunoo, Ibanag, Ifugao, Isnag, Kalinga, Kapampangan, Sambal (Zorc 1979:241). Based on the fact that cognate forms among these languages share the same length on the penult, Zorc (1979) argues that the contrastive word accent system may be attributed to “Proto-Philippines”,² as shown in Proto-Philippines (PPH) *da:Raʔ ‘blood’, cf. Isnag *da:ga*, Ilokano *da:ra*,

¹ This article is a revised version of the following talks I presented: ‘Synchronic and diachronic phonology of the Arta language’ in The 3rd MINPAKU Linguistics Circle, Osaka, Japan, Sept. 2014, ‘Development of the phonological system in Arta’ in Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, Japan, 2015, and ‘The role of mora in phonology: A case for Arta, a Northern Luzon language’ in 13th International Conference on Austronesian Linguistics, Academia Sinica, Taipei, Taiwan, July, 2015. My thanks go to Ritsuko Kikusawa, Hsiu-chuan Liao, Naonori Nagaya, Lawrence A. Reid, and David Zorc for comments on the presentations and/or an earlier version of the paper. All remaining errors are my own.

² Although Zorc labelled the hypothetical proto-language as “Proto-Philippines” in his previous publications, and this paper cites his reconstructed data with the same label Proto-Philippines for the purpose of comparing reconstructed forms with Arta forms, he no longer holds to a Proto Philippines. (Zorc

Ifugao *da:la*, Kapampangan, Sambal *da:yaʔ*; PPH **da:lan* ‘path, trail’, cf. Aklanon, Balangao, Cebuano, Bikol, Hanunoo, Ibanag, Isnag, Kapampangan, Ilokano *da:lan*.

This is not the case in Arta, a Northern Luzon Negrito language, however. In spite of the abundance of items with long vowels (e.g. *ka:man* ‘big, large’, *bu:ru* ‘new’, and *a:na:* ‘children’), the etyma with long penult in the reconstructed language are all reflected with short vowels, as in (2).

- (2) PPH **si:ku* > Arta *siku* ‘elbow’
PPH **tu:bu* > Arta *tubu* ‘grow’

The aim of this paper is thus to explain the historical development of long vowels in Arta. Three sources of long vowels are identified: (i) onomatopoeia, (ii) borrowing from Ilokano and Yogad, and (iii) sound changes in inherited forms from Proto-Malayo-Polynesian (henceforth PMP). It is argued that the *mora* was responsible for the third pattern, playing a significant role in the historical phonology in Arta.

This paper is organized as follows. §2 provides basic information about the Arta language. In §3, it is argued that the *mora* plays a significant role in synchronic phonology in Arta. §4 deals with the matter as to how long vowels in Arta are developed historically. After the discussion of the first two sources of long vowels in Arta in §4.1 and §4.2, §4.3 observes sound changes and some factors which might have motivated the vowel lengthening, with special reference to the *mora* count. §5 deals with the interactions between the *mora* and other factors which may affect current reflexes in Arta.

2 The Arta language

2.1 Its speakers and sociolinguistic profile

Arta is an Austronesian language currently spoken by eleven Arta people living in the municipality of Nagtipunan, Quirino province, the Philippines. The speakers of Arta belong to a larger group called Negrito, who are commonly characterized as having shorter stature, curled hair, and darker skin. They are considered to be a descendant of the people who had settled in the Philippine archipelago over tens of thousands years before the speakers of an Austronesian language migrated into the islands; since Negrito people switched from their original languages to Austronesian languages, the current language that the Arta people speak clearly belongs to the Austronesian family.³

The northern part of Luzon is home to a number of Negrito groups as well as non-Negrito groups (see Fig. 1). The longest river in the Philippines, the Cagayan River, runs from south to north, forming the Cagayan Valley. The valley is sandwiched between the Sierra Madre mountains in the east and mountainous district in the west. The east side of Sierra Madre, that is, the eastern coast of northern part of Luzon, is home to various Negrito groups such as Dupanangan Agta, Pahanan Agta, Casiguran Agta, and Dinapigue Agta. The valley of the Cagayan River is also occupied by other Negrito groups, Pamplona Atta, Faire-Rizal Atta, Pudtol Atta in the northwestern side of the valley, and Central Cagayan Agta in the northeastern side. Quirino Province is located on the upper reaches of the Cagayan River, which is occupied by Nagtipunan Agta as well as Arta (Fig. 2). The area was formerly covered with a thick rainforest, traditionally occupied by Nagtipunan Agta and a non-Negrito group Ilongot (or Bugkalot) as well as Arta, but the area is currently inhabited by a large number of immigrants from outside of the province to reclaim the forested area. The Arta people formerly lived in Cordon and Alicia in Isabela, but they moved to Disubu in the municipality of Aglipay in Quirino around four to five decades ago; they are currently settled in the municipality of Nagtipunan, Quirino.

pers.comm.); in fact, recent studies (Reid 1982, Ross 2005) point out the lack of phonological evidence to justify Proto-Philippines as a distinct proto-language from Proto-Malayo-Polynesian.

³ It is interesting that many of the Negrito languages are conservative in grammatical and/or phonological characteristics. Arta, for example, retains the old pronominal form =*myu* (second person plural genitive form), which could date back to Proto-Northern Luzon (Reid 1979), and perhaps to PMP (see Reid 2009, but Ross 2006 proposes the formatives *=ihu, *=mu-ihu for the second person plural genitive).

Arta is a severely endangered language compared with other Negrito languages which are also seen as “minority” languages (see Headland 2003 for the demographic data on Philippine Negrito languages). The number of fluent speakers of Arta is 10, with 35–45 people barely understanding the language but unable to speak properly; most of the speakers are over 40 years old, except one young fluent speaker at the age of 29. In the last two decades, Arta people moved from Aglipay and/or Maddela to Nagtipunan, being merged into the speech communities whose majority are Nagtipunan Agta.⁴ This seems to have caused the increase of the intermarriage between Arta and Nagtipunan Agta, to the extent that there is currently no “pure” Arta family. This social structure inevitably forces the Arta language, a “minority” language in the community, to be out of use; in my fieldwork, I have been unable to find any family in which Arta is spoken. They usually communicate in Nagtipunan Agta within the community, in Ilokano outside the community, and in Arta with an older generation whose first language is Arta; in fact, some Arta people use the language when they talk to their siblings, but not to their children.

Figure 1: Northern part of Luzon



Figure 2: Quirino Province (seen from the south)



2.2 Subgrouping relationship and language contact

The genetic subgrouping of Arta was studied by Reid (1989). He concludes that the language is an isolate within the Northern Luzon (NLZN) subgroup of Malayo-Polynesian (MP). Northern Luzon languages are widely distributed in the northern part of Luzon, surrounded by Bashiic languages spoken in Batanes islands in the north, and by Central Luzon languages in the south. Northern Luzon languages include Ilokano; Meso-Cordilleran languages such as Bontok, Ifugao and Kalinga; Cagayan Valley languages such as Yogad, Ibanag and Gaddang; and North-eastern Luzon languages including Dupanangan Agta, Pahanan Agta, Casiguran Agta, and Dinapigue Agta.⁵ One piece of linguistic evidence for the subgrouping as NLZN is the sporadic metathesis between *t ... s > /s ... t/, which is shared exclusively by NLzn languages, as in PMP *tañis > PNLZN *sanjit (Reid 2006). Although Reid

⁴ Nagtipunan Agta is another Negrito group occupying the area. Few ethnographic or linguistic studies are conducted on Nagtipunan Agta except Robinson and Lobel (2013), a comparative study on East coast Negrito languages. Nagtipunan Agta is mutually intelligible with Casiguran Agta (or Casiguran Dumagat; Headland and Headland 1974, Headland and Healey 1974). It is still unclear as to how long they have been occupying the areas in Disimungal, or how different it is from Casiguran Agta.

⁵ Northern and Southern Alta are also Negrito languages subgrouped within Northern Luzon (see Reid 1989, 1991).

(1989, 2013) does not provide such evidence for the subgrouping of Arta, the metathesis does exist, as shown in (3).

- (3) PMP *ditaʔas > *disat* ‘high (the sun)’
 PMP *təRas ‘hardwood, hard’ > *sarat* ‘narra wood’

Since PMP *ʔ was lost and PMP *R changed into /r/ in Arta, the items in (3) are in accordance with regular sound changes that occurred in the language. The cognates presented above seem to provide strong evidence for positing that Arta is subgrouped within other Northern Luzon languages.

Another important sound change that occurred in Arta is *R > /r/, which provides strong evidence for the subgrouping of NLzn languages. As Reid (1989) states, Arta and Ilokano reflect *R as /r/, Meso-Cordilleran languages /l/, and North-eastern and Cagayan Valley languages /g/. Arta has a different reflex from the latter two subgroups, except Ilokano, which is still difficult to subgroup with, partially because most of the forms are not uniquely shared. Even the cognate words that would be expected to have shared innovations (*R > /r/) are reflected differently, e.g. *bəRɲaw > Arta *biriɲaw*, Ilk. *bəɲaw* ‘fly (n.)’; *huRas > Arta *uras*, Ilk. *u:gas* ‘wash’; *kaRat > Arta *arat*, Ilk. *kagat* ‘bite’; *Raɲu > Arta *raɲu*, Ilk. *gaɲu* ‘wither’; *Rapu > Arta *rapu*, Ilk. *gapu* ‘be from’ (it seems that Arta has more coherent reflexes of *R as /r/ than Ilokano). In Arta, PMP *j consistently changed into /d/ and thus merged with *d, which exhibits a further difference from Ilokano, where *j is reflected as /g/ (Reid 1989).

This subgrouping, however, reflects only one aspect of the linguistic history of Arta; the speakers of Arta have undergone several major periods of language contact. The ancestors of Philippine Negrito groups, including Arta, are considered to have dispersed into the Philippine archipelago in the Palaeolithic age over 20,000 years ago. It is suggested by current genetic and archaeological studies that, after common ancestral populations of modern humans dispersed “out-of-Africa”, probably through south Asia along the coastal side, the populations settled in the Southeast Asia as a second dispersal no later than 25–38,000 years ago, after the first dispersal into New Guinea, Melanesia, and Australia ~62–75,000 years ago (Rasmussen et al. 2011, Reyes-Centeno et al. 2014). The populations that settled in the Philippines must have spoken non-Austronesian languages, although there seems to be no clear evidence for reconstructing the details, except probable non-Austronesian lexical residuals in some current Negrito languages (Reid 1994).

The early Austronesians residing in Taiwan migrated into the Philippines probably via the Batanes islands around 4,000 years ago (see Reid 2013, Ko et al. 2014, see further Blust 1999, 2013). The ancestors of Arta are considered to have switched their languages to an Austronesian language which may currently be called Proto-Northern Luzon. Furthermore, after the language shift, the Arta language was influenced by Ilokano and Cagayan Valley languages such as Yogad, Gaddang, and Ibanag. Yogad, in particular, was influential on Arta; in my count, Arta shares with Yogad at least 98 lexical items. This is probably because the Arta people resided in Alicia until four to five decades ago, and they were bilingual in Arta and Yogad until recently. Speakers of Arta claim that they are similar to Yogad as “relatives”.

Following a basic description of the sociolinguistic and historical profiles of the Arta language, the next section will provide the definition of the mora count and its application to the Arta synchronic phonology. It is argued that the mora is at work in various phonological aspects in Arta, which will in turn be a requisite for explaining diachronic change in vowel length.

3 The mora in synchronic phonology

3.1 Phonological status of vowel length in Arta

Arta has sixteen consonant phonemes: seven stops /p, b, t, d, k, g, ʔ/, three nasals /m, n, ŋ/, two fricatives /s, h/, two liquids /l, r/, and two glides /y, w/, and has six vowels /i/, /e/, /a/, /o/, /u/, and /ə/. In this paper, the symbol “y” refers to the glide /j/ in the phonemic and orthographic descriptions. Vowel length is phonemically distinctive, as in *bi:ləg* ‘bracelet’ vs. *biləg* ‘fastness’, *ana:* ‘child’ vs.

a:na: ‘children’. /i/, /a/, /u/, and /ə/ have long and short phonemic contrasts;⁶ however, /e/ and /o/ always appear as long vowels /e:/ and /o:/ unless they are reduced as short vowels within the CVC syllable template (as discussed in §3.2, long vowels cannot appear in CVC syllables).

How should long vowels be interpreted phonologically? This may be seen as vowel length as it is, or as the phonetic manifestation of the stress (or accent) system as in the case of other Philippine languages. Nevertheless, the present study demonstrates that it is phonologically the length of a vowel, rather than the phonetic manifestation of the stress system in two respects. First, if the long vowel is really a manifestation of lexical stress, the place of the stress may shift after affixation, as in English: *phótophraph* > *photógraphy* > *photográphic*. This is the case in Ilokano, as in *ba:sa* > *basa:-en* ‘read’ and *la:wa* > *ka-lawá:-en* ‘to widen’ and in Tagalog as in *gu:gol* > *gugu:l-in* ‘to spend’ and *tu:ro* > *turo:-an* ‘to teach’, but this kind of shift is not observed in Arta, as shown in (4):

- (4) *pe:nas* > *pe:nas-an*, *******pena:s-an* ‘wipe’⁷
di:muy ‘bathing’ > *pandi:muy-an*, *******pandimu:y-an* ‘bathing place’
ka:lig > *ka:lig-ən*, *******kali:g-ən* ‘carry, transfer something’

The second defining feature is that each word (not a *root*) has one stress nucleus. Even if two stresses are observed within a single word, they have a phonetic asymmetry to the extent that one stress is considered to be primary, and the other secondary (e.g. *dòcuméntary*, *téxt-bòok*). This is not the case in Arta. Several words in Arta do have more than one long vowel without such phonetic asymmetry. Consider the following forms in (5):

- (5) More than one long vowel can occur within one word
ta:me:ta ‘different’ > *ta:ta:me:ta* ‘quite different’
na:na:b (*na:na:bən*) ‘remember, recall’
no:no:t (*no:no:tən*) ‘think’
pe:be:bu:d=u ‘I am asking’

Since neither of these facts favor the long-vowel-as-stress analysis, this contrast will be treated more appropriately as the phonological long-short distinction of vowels.

3.2 *Mora and synchronic phonology*

In many of the world’s languages, prosodic and/or metric structures are sensitive to a different “weight” of syllables defined by the vowel length and other syllable features. The unit of the syllable weight is called *mora*, and a monomoraic syllable is defined as an open syllable with a short vowel. Languages may differentiate a monomoraic syllable (light syllable) with a short vowel (CV), a bimoraic syllable (heavy syllable) with a long vowel (CV:), and, in some languages, a trimoraic syllable (superheavy syllable) with a superlong vowel (CV::). Many languages such as Japanese and Latin count a coda consonant as having one mora. This means that there are two patterns for classifying syllables in terms of mora count, as shown in (6) and (7).

- (6) Pattern A (coda consonants are counted)
 i. monomoraic syllable: CV
 ii. bimoraic syllable: CV:, CVC
 iii. trimoraic syllable: CV::, CV:C

⁶ /ə/ does not appear as a long vowel except in one item *də:gi* ‘strip or string used for carrying a basket’.

⁷ The presence of double asterisks (**) before a form indicates that the form in question is “unacceptable” or “illegitimate”.

- (7) Pattern B (coda consonants are NOT counted)
 i. monomoraic syllable: CV, CVC
 ii. bimoraic syllable: CV:, CV:C
 iii. trimoraic syllable: CV::, CV::C

In the case of Arta, moraic classification of syllables can be shown in (8) and generalized in (9).

- (8) Arta (coda consonants are counted)
 i. monomoraic syllable: CV
 ii. bimoraic syllable: CV:, CVC
 iii. *trimoraic syllable: CV::, CV:C
- (9) Mora constraint on syllables: The syllable must not exceed 2 moras

First, as to be illustrated below, the syllables in Arta behave in a way similar to pattern A, that is, a syllable with a coda consonant and a syllable with a long vowel behave in the same way. The language is not allowed to have trimoraic syllables such as CV:: (superlong vowel) and CV:C (long vowel and coda consonant). If morphological conditions require such syllable structure, these syllables should be reduced to CV: and CVC types, respectively. This constraint can be paraphrased as (9): “The syllable must not exceed 2 moras.” Now let me introduce some mora-sensitive phenomena in Arta phonology, which support the statements in (8) and (9).

3.2.1 Allomorphs of enclitics

The mora count is necessary for generalizing some allomorphic conditions. Two enclitics, =*di* ‘already, just now’ and =*pa* ‘just, try -ing, do a little’ are sensitive to the mora count of the preceding syllable to which they attach. If the preceding word ends with a bimoraic syllable, then =*di* and =*pa* must appear, but if the preceding word ends with a monomoraic syllable, =*d* and =*p* must be used instead. The following example with =*di* illustrates the point.

- (10) CVC + di *awan*=*di* (NEG=already) ‘It does not exist already.’
 CV: + di *mebbuyu*:=*di* (bad-smelling=already) ‘It has already become bad-smelling.’
 CV + di *nappati*=*d* (died=already) ‘S/he has already died.’

The condition under which different allomorphs may occur cannot be explained solely in terms of vowel length nor only according to whether the preceding syllable is open or closed. Rather, this fact indicates that =*p* and =*d* may appear if the preceding word ends with a monomoraic syllable, whereas =*pa* and =*di* may appear if the preceding word ends with a bimoraic syllable, with coda consonants counted as one mora in the language.

3.2.2 Vowel length alternation

The moraic constraint on the maximal weight of the syllable structure, explicated in (8) and (9), predicts vowel length alternations. If a long vowel appears in a closed syllable, it is supposed to be realized as a reduced short vowel, whereas it is realized as a long vowel if it appears in an open syllable. This alternate realization is not applied only to the native lexemes (11a), but also to Spanish and English loan words (11b).

- (11) a. /latto:ŋ/ > *lattoŋ* vs. *latto:ŋ=i*⁸ ‘outside’
 /sa:y/ > *mas-say* vs. *sa:y-an* ‘ride on’
 /dage:t/ > *man-daget* vs. *dage:t-an* ‘sew’
 /du:t/ > *dut* vs. *du:t=i* ‘firewood’
 b. /hapo:n/ > *hapon* vs. *hapo:n=i* ‘Japan’
 /pantalo:n/ > *pantalon* vs. *pantalo:n=i* ‘pants’
 /hamba:g/ > *hambag* vs. *hamba:g=i* ‘handbag’
 /bolpe:n/ > *bolpen* vs. *bolpe:n=i* ‘ballpoint pen’

Each lexeme, which potentially has a long vowel, exhibits different realizations of the vowel according to the type of syllable in which it occurs. The first pattern blocks the realization of long vowels because of the presence of a coda consonant, whereas the second pattern, without a coda consonant, allows the vowel to be realized as long. This clearly comes from the moraic constraint on the syllable structure, in which any trimoraic syllable like CV:C is not allowed to occur.⁹

A similar case can be seen in nominalizing circumfixes *paC-* *-ən* and *paC-* *-an*, which cause the lengthening of the following vowel.¹⁰ This is illustrated in (12), where *tapik*, *bisag*, *lagip*, and *sirit* undergo the lengthening of the first syllable by the affixation of *paC-* *-ən* or *paC-* *-an*. However, the circumfixes do not trigger the lengthening of the vowel if the syllable in question has a coda consonant. The syllable structure again preempts vowel lengthening, because of the moraic-constraint on the syllable structure in which a trimoraic syllable such as CV:C is allowed to occur in Arta.

- (12) *pat-ta:pik-ən* ‘slapping something’
pab-bi:sag-ən ‘breaking something’
pal-la:gip-an ‘telling something, story’
pas-si:rit-an ‘defecating somewhere’
- (13) *pag-gimt-ən* ‘making/doing something’
pap-pissay-ən ‘tearing something’
pat-tuttud-an ‘sitting on, something to sit on’
pag-gusgus-an ‘scratching something’

All of the above data illustrate that the mora count is at work at least in the synchronic phonology in Arta. In addition, the mora count also explains the development of the short-long contrast in vowels as shown in §4.3.

4 The development of long vowels in Arta

As mentioned in the introductory section, all of the forms inherited from PPH lost their vowel length contrasts. A fuller list of items is shown in (14).

⁸ =*i* is a post-nominal specifier. See Table 5 in Appendix (and compare it with determiner sets shown in Table 4 in Appendix).

⁹ Another account for this vowel length alternation would be that the enclitic =*i* and the suffixes *-an* and *-ən* might trigger the lengthening of a preceding vowel. These formatives, however, do not have an ability to lengthen a preceding vowel; in fact, the following realizations do not include any lengthening: /asuk/ > *asuk* vs. *asuk=i*, /pabay/ > *pabay* vs. *pabay-an*, /idut/ > *idut-ən*.

¹⁰ *paC-* *-ən* and *paC-* *-an* are the progressive and nominalizing forms of *-ən* (patient-transitive) and *-an* (location-transitive), respectively (see Table 6 in Appendix).

- (14) PPH *ku:tu[h] > *utu* ‘lice’
 PPH *si:ku > *siku* ‘elbow’
 PPH *tu:bu > *tubu* ‘grow’
 PPH *su:su > *susu* ‘breast’
 PPH *qu:lu > *ulu* ‘head’
 PPH *bu:lan > *bulan* ‘moon’
 PPH *ŋa:jan > *ŋadin* ‘name’
 PPH *ka:yuh > *ayu* ‘tree’

A comparison with other Philippine languages which are considered to retain the old accentual system also suggests that Arta lost penultimate long vowels (the following data from Isnag, Bontok, Ifugao, Ilokano, Cebuano, and Bikol are collected from Blust and Trussel (In progress)).

- (15) Arta: *dudun* ‘locust’
 cf. Isnag, Bontok, Ifugao, Ilokano *du:dun*, Cebuano *du:lon*, Bikol *du:ron*
 Arta: *lutu* ‘cook’
 cf. Isnag, Bontok, Ifugao, Ilokano *lu:tu*, Bikol *lu:to?*
 Arta: *mula* ‘to plant’
 Isnag, Bontok, Ilokano *mu:la*
 Arta: *pusəd* ‘navel’
 cf. Isnag *pu:sag*, Bontok, Ilokano *pu:səg*, Ifugao *pu:hog*, Bikol *pu:sod*

These data strongly indicate that the long vowels in Arta are not the short-long contrast inherited from a proto-language common to languages retaining the old accent system. In what follows, it is shown that there are three types of lexemes with long vowels which should be treated separately: onomatopoeic words, loanwords, and inherited forms. After observing long vowels seen in onomatopoeic words and loanwords (the reason why onomatopoeic words should be treated separately will also be mentioned in §4.1), the most puzzling case, inherited forms containing long vowels, is discussed. It is shown that long vowels in inherited forms developed independently as a result of the loss of PMP consonants and subsequent compensatory lengthening or vowel fusion, both of which, arguably, involve the mora count conservation.

4.1 Long vowels in onomatopoeic words

Some ideophones, more specifically, expressions imitating sounds (*onomatopeia*), contain long vowels. The following five items with a long vowel are found in my corpus.

- (16) *tattara:kot* ‘cock-a-doodle-doo (a cry of roosters)’
ku:tak ‘cluck (a cry of hens)’
be:w ‘a cry of deer’
ku:rək ‘chicken’
pi:yək ‘chick’

The first three words are the imitations of an animal’s call. The rest of them are the names of animals probably via the semantic shift metonymically from the typical sound they emit. From a methodological perspective, these words should be treated separately from other non-onomatopoeic words. It is often the case that onomatopoeic words tend to develop differently in terms of sound change. This in fact occurred in the history of the Japanese language; Komatsu (1989) argues that some onomatopoeic words retained the sound /p/ as in /pitoku pitoku/ (a sound of a small bird’s crying) even after /p/ changed into /ϕ/ in Classical Japanese. The important point is that even if the above items had been inherited from a proto-language, these are not counterexamples. They might

have developed subsequently after the loss of the old accentual system, or might have retained a long vowel in it independently of the loss of accentual system.¹¹

4.2 Long vowels in borrowed items

A large number of items containing long vowels seem to be loanwords from Yogad and Ilokano (and Spanish and English via these languages).¹² The following items share the same surface forms with those in Yogad. The items containing /k/ are clearly loans because *k was lost in Arta. The formative *ta:* in *ta:gatut* ‘one hundred’, *ta:hulu* ‘ten’, and *ta:ribu* ‘thousand’, which probably underwent the change from *sa ‘one’ > /ta:/, exhibits *s > /t/, a sound change characteristic of Cagayan Valley languages, not of Arta. The sources of *ka:ya* and *la:ku* are unclear because Ilokano also has the same forms. Considering the large number of lexical items shared with, and thus possibly borrowed from, Yogad, it seems that there is a smaller number of loanwords with a long vowel than expected. This may reflect a historical change which occurred in Yogad, one of the languages which lost the old accentual system, with a large number of lexical items having a short penult.

(17) Possible loanwords from Yogad

<i>illa:yug</i> ‘long’	<i>li:nis</i> ‘clean’
<i>i:lug</i> ‘egg’	<i>no:not</i> ‘think’
<i>ka:ya</i> ‘can, be able to’	<i>ta:gatut</i> ‘one hundred’
<i>ki:gad</i> ‘until’	<i>ta:hulu</i> ‘ten’
<i>ki:bu</i> ‘mix’	<i>ta:ribu</i> ‘thousand’
<i>la:ku</i> ‘buy’	<i>tu:rak</i> ‘write’
<i>le:but</i> ‘walk around’	<i>u:bi</i> ‘violet/purple yam’

(18) shows a list of forms shared with Ilokano. These forms are likely to be loanwords because Ilokano is a language which retains the old accentual system, with a large number of items having a long penult, which would otherwise have been shortened if they should be inherited forms from PMP. Loanwords from Spanish and English are also found possibly via Ilokano or Yogad, as shown in (19) and (20).

(18) Possible loanwords from Ilokano

<i>badu:ya</i> ‘kind of cake’	<i>bu:ya</i> ‘watch TV’
<i>inda:yun</i> ‘hammock’	<i>tu:luy</i> ‘continue’
<i>ta:wa</i> ‘window’	<i>bi:lin</i> ‘order’
<i>ba:sa</i> ‘read letters’	<i>uga:li</i> ‘habit, custom’

(19) Loanwords from Spanish

<i>amerika:no:</i> ‘American’ (< <i>Americano</i>)	<i>kalsa:da</i> ‘paved road’ (< <i>calzada</i>)
<i>antipa:ra</i> ‘goggles’ (< <i>antiparras</i>)	<i>kande:la</i> ‘candle’ (< <i>candela</i>)
<i>ari:na</i> ‘flour’ (< <i>harina</i>)	<i>kasape:gu</i> ‘matches’ (< <i>casa de fuego</i>)
<i>binta:na</i> ‘window’ (< <i>ventana</i>)	<i>kla:se:</i> ‘class’ (< <i>clase</i>)
<i>bisi:ta</i> ‘guest’ (< <i>visita</i>)	<i>pantalon</i> (/pantalo:n/) ‘pants’ (< <i>pantalon</i>)
<i>hapon</i> (/hapo:n/) ¹³ ‘Japan’	<i>sakripi:syu</i> ‘sacrifice’ (< <i>sacrificio</i>)
<i>ka:da-</i> ‘each, every’ (< <i>cada</i>)	<i>taraba:hu</i> ‘work’ (< <i>trabajo</i>)

¹¹ See also Zorc (1990), which points out that Kalamianiac languages show the irregular retention of *k in sound-symbolic monosyllabic roots and their derivatives.

¹² The following sources were used for examining diachronic changes in phonology: Blust and Trussel (In progress) for Proto-Austronesian (PAN) and Proto-Malayo-Polynesian (PMP), Reid (1979) for Proto-Northern-Luzon pronominal forms, Rubino (2000) for Ilokano, Davis and Mesa (2000) for Yogad, Headland and Headland (1974) for Casiguran Agta. Unless other sources are referred to, the references for each data will not be explicitly mentioned.

¹³ See the description in (11b) for the explanation of the short realization of the potentially long vowel.

- (20) Loanwords from English
ba:bay ‘good-bye’
basket (/baske:t/) ‘basket’
bolpen (/bolpe:n/) ‘ballpoint pen’
hambag (/hamba:g/) ‘handbag’
me:kap ‘makeup’

As the above data indicate, Arta has borrowed a large number of items used in Yogad, Ilokano, Spanish, and English. Original stresses of the items in the source languages were re-interpreted as a long vowel when borrowed into Arta. And interestingly, two vowels *o* and *e* are always reinterpreted as long vowels, as in *kla:se:* ‘class’, /bo:lpe:n/ (e.g. *bolpe:n=i* ‘a specific ballpoint pen’), /hapo:n/ (e.g. *hapo:n=i* ‘a specific Japanese person’), regardless of whether the vowel has originally a stress or not. The possible reason for the reinterpretation will be considered in §4.3.2.

Having considered the effects of borrowings, we still find other lexical items with long vowels. These items lead us to consider the third case, in which inherited forms from PMP have a long vowel as a result of several kinds of sound changes.

4.3 Long vowels in inherited forms from PMP

Arta underwent some phonological changes from PMP. A list of PMP reconstructed phonemes and their reflexes in Arta is shown in Table 1 (see Kimoto 2017 for the fuller discussion of the sound changes in Arta).

Table 1: Reflexes of PMP phonemes

PMP ¹⁴	Arta	PMP	Arta	PMP	Arta
*p	/p/	*s	/s/	*a	/a/
*t	/t/	*R	/r/	*i	/i/
*k	∅ ~ /k/ ¹⁵	*l	/l/	*u	/u/
*q	∅	*m	/m/	*ə	/ə/
*b	/b/	*n	/n/	*a+i	/e:/
*d	/d/	*ŋ	/ŋ/	*a+u	/o:/
*j	/d/	*h	∅		
*z	/d/	*w	/w/		
*g	/g/	*y	/y/		

In Arta, *k, *q (glottal stop) and *h are reflected as zero, *R as /r/, *j as /d/, and, as to be discussed later, the two vowel sequences *a+i and *a+u are reflected as two new vowels /e:/ and /o:/, respectively. It is argued here that there are two types of process under which inherited forms underwent vowel lengthening: compensatory lengthening and vowel fusion. Both of which involve the conservation of the mora count.

4.3.1 Compensatory lengthening

The first type of process in which vowel lengthening occurs is COMPENSATORY LENGTHENING. Compensatory lengthening is a phonological process in which the loss of a coda consonant triggers the lengthening of an adjacent segment. A typical case is the vowel lengthening triggered by the loss

¹⁴ The representation of proto-phonemes follows a conventional style of Austronesian linguistics: PMP *R is considered to have been pronounced as [ɣ], and PMP *q as [ʔ].

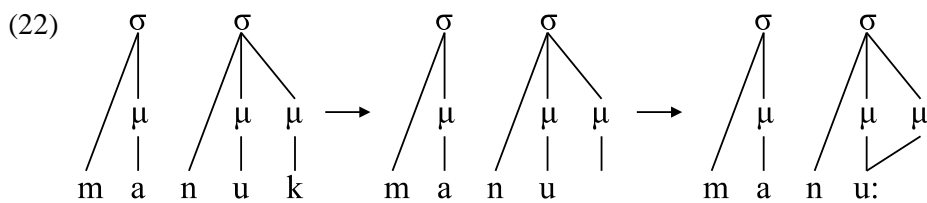
¹⁵ *k is sporadically reflected as /k/, such as *bəRək > bəRək, *anak > kanakannak ‘child’, *=ku > =ku ‘I’ (1SG.GEN).

of the following consonant, as in Latin **kasnus* > *ka:nus* ‘gray’ (Hayes 1989:260). The following items in Arta also illustrate the same point:¹⁶

- (21) **manuk* > *manu:* ‘bird’
**anak* > *ana:* ‘child’
**buyuk* > *buyu:* ‘bad-smelling’
**abak* > *abi:* ‘body’ cf. AltaS: *abek*
**buliq* > *buli:* ‘buttocks’

The development of long vowels seems to result from the loss of **k* in the case of *manu:*, *ana:*, *buyu:*, and *abi:*, and the loss of **q* in *buli:*. Since the language has the lowest percentage of retentions of reconstructed PMP vocabulary of Philippine languages, 27% (Reid 1989), not much evidence can be provided. However, in all the cases in which a coda consonant is lost, the preceding vowel is lengthened.

Compensatory lengthening can be explained in terms of the conservation of mora count, as argued by Hayes (1989, 1995). Remember that in Arta both CV: and CVC are treated as bimoraic. As depicted in (22), by dropping a coda consonant, /k/ in this case, a mora becomes empty, which is assigned to the preceding vowel by its lengthening (in the following illustration, σ represents a syllable, and μ the mora).



A more complex pattern of compensatory lengthening is found in person forms. The following enclitic pronominal forms underwent the loss of **k* (see Table 2 in Appendix for a full set of pronominal forms).

- (23) PNLzn (Reid 1979) > Arta
**=ka* > *=a* (2SG.ABS)
**=kamuyu* > *=am* (2PL.ABS)
**=kami* > *=ami* (1PL.ABS)
**=kita* > *=ita* (1+2SG.ABS)
**=kitam* > *=itam* (1+2PL.ABS)
**=ku* > *=ku* ~ *=u* (1SG.GEN)

The above forms are synchronically peculiar as well. They differ from other paradigmatically-related items in that they exhibit a complex morphophonemic alternation conditioned by the type of segments to which they attach. More specifically, if the preceding word ends with a consonant (except /n/) as in (24a), a vowel before the consonant is lengthened, and, in the case of *=ku*, /k/ is dropped. If the preceding word ends with /n/ as shown in (24b), /n/ changes to a velar, and the preceding vowel lengthened and, in the case of *=ku*, the deletion of /k/ occur. If the preceding word ends with a vowel as shown (24c), it is just followed by the enclitic in the case of *=ku*, *=ita* and *=itam*, and, in the case of *=am* and *=ami*, /y/ is inserted between the host word and enclitic.

¹⁶ Other forms which might involve compensatory lengthening are: *pura:* ‘white hair’, cf. Ilokano *purakrak* ‘(white things are) brilliant’ and *adu:yu* ‘far, distant’ (cf. Isneg *adayyu* with vowel harmony).

(24) Morphophonemic alternations

a. after a consonant:

babakat ‘old woman’ + =*ami* > *babaka:t=ami* ‘we are old women’

lusip ‘nail’ + =*ku* > *lusi:p=u* ‘my nails’

b. after /n/:

buka:gan ‘woman’ + =*ami* > *buka:ga:ŋ=ami* ‘we are women’

bunbun ‘house’ + =*ku* > *bunbu:ŋ=u* ‘my house’

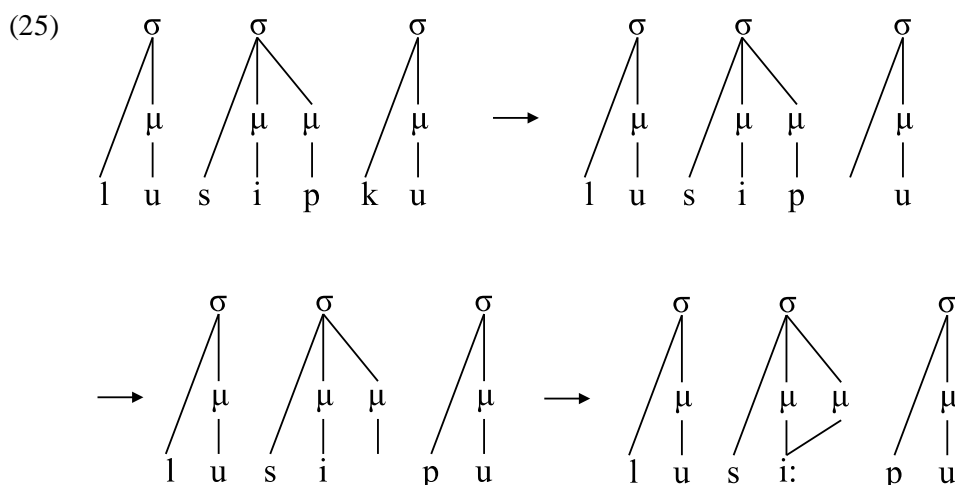
c. after a vowel (in case of =*ku*, =*ita*, =*itam*)

mata ‘eye(s)’ + =*ku* > *mata=ku* ‘my eyes’

after a vowel: insertion of /y/ (in case of =*am*/=*ami*)

me:na ‘go’ + =*am*/=*ami* > *me:nayam/yami* ‘You/we will go’

How should this complex set of morphophonemic alternations be interpreted in terms of a historical development? This synchronic fact seems to be subsumed in a “double-flop” (Hayes 1989:265), which is a subtype of compensatory lengthening. A double flop may occur when there is a cluster of consonants consisting of a syllable coda and a following onset consonant. After the loss of the onset consonant, the preceding coda consonant is re-interpreted as an onset consonant of the following syllable, and the empty mora carried by the original coda consonant is borne by the preceding vowel by lengthening. Consider the case of *lusip* ‘nail’ + =*ku* > *lusi:p=u* ‘my nails’, following the illustration in (25):



In this case, after the loss of the onset consonant /k/, the preceding segment /p/ is re-interpreted as the onset consonant probably by a universal principle of stable, unmarked syllable structure. This leads to the floating of the mora by the shift of /p/, thus being resolved by the lengthening of the preceding vowel /i/ to bear the mora. The change of /n/ to a velar nasal, in (24B), as in *buka:gan* ‘woman’ + =*ami* > *buka:ga:ŋ=ami* ‘we are women’, must be a result of assimilation to /k/; thus, it is more likely that this change occurred at the initial stage of (25), that is, before the loss of /k/.¹⁷

Positing the mora as an explanatory apparatus is further supported by the asymmetrical relation between the loss of coda and onset consonants. The moraic account predicts that the loss of onset consonants does not trigger a compensatory lengthening because the onset consonants do not bear a mora, unlike the coda consonant. Consider the following phonological changes involving the loss of onset consonants.

¹⁷ The item *la:səm* ‘sour’ (< *alsəm) involves another kind of compensatory lengthening. After the metathesis of /a/ and /l/ occurred, the bimoraic syllable seems to have been compensated for by lengthening the vowel, resulting in *la:səm* in Arta.

- (26) *kutu > *utu* ‘lice’
 *kua > *wa* ‘what-cha-ma-call-it’
 *kulit > *ulit* ‘bark, skin’
 *kaRat > *arat* ‘bite’
 *kan > *an* ‘eat’

The items shown in (26) demonstrate that a long vowel was not developed after the loss of an *onset* consonant; in fact, all the items with the loss of an onset do not exhibit vowel lengthening. This suggests that vowel lengthening in inherited forms involves the mechanism of the mora count as an explanatory apparatus.

4.3.2 Vowel fusion

The second process in which long vowels emerged is vowel fusions caused by the loss of intervocalic consonants. Vowel clusters *a+a, *u+u, *a+i, and *a+u lead to the emergence of four long vowels /a:/, /u:/, /e:/, and /o:/, respectively.¹⁸ Among them, /e:/ and /o:/ are interesting in that they are the phonemes that PMP did not have and that were developed by the very process.

The following five items exhibit vowel fusions *a+a > /a:/ and *u+u > /u:/.

- (27) *a+a > /a:/
 *sakay > /sa:y/ ‘ride on’: *mas-say* (INTR-ride), *sa:y-an* (ride-LV) ‘ride on’
 *ka-ama-ən > *ka:man* ‘big’
 *di *tahaw > /dita:w/ ‘outside’: *ditaw* ‘outside’, *dita:w=i* ‘the specific outside space’
- (28) *u+u > /u:/
 *dukut > /du:t/ ‘fire’ (cf. *dut* ‘fire’ vs. *du:t=i* ‘the fire’)
 *baqəRu > *bu:ru* ‘new’ (with a vowel harmony /a/ > /u/)

Note that, although *sakay, *tahaw and *dukut exhibit vowel fusions caused by the loss of intervocalic *k, given the mora constraint on syllable structure, a long vowel appears only if the word-final consonant is resyllabified as an onset consonant as in *sa:y-an* ‘ride on’ and *du:t=i* ‘the fire’.

The process of *a+a > /a:/ and *u+u > /u:/ is also observed in synchronic variations in some cases. A reduplication of a vowel-initial base, and the prefixation of *ma-* and *maka-* (potentive verb prefixes, see Appendix 6) followed by /a/-initial bases, trigger a vowel fusion (* represents a pre-Arta form in (29) and (30)).

- (29) *ma-ʔalap > *ma:lap* or *maʔalap* ‘can get, succeed’
 *maka-ʔaŋay > *maka:ŋay* or *makaʔaŋay* ‘can go into’
- (30) *ʔa~ʔana: (reduplication of *ana:*) > *a:na:* ‘children’
 *ma-ʔarawat-an > *ma:rawatan* ‘can grasp, understand’

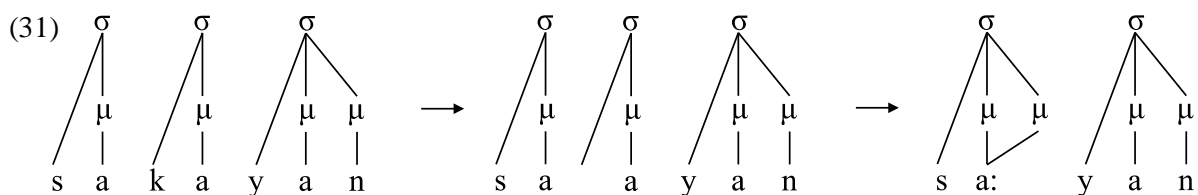
This involves the fact that an intervening glottal stop between two adjoining morphemes is synchronically on the verge of disappearing. Apart from PMP *q, which was lost at an early stage, Arta has another glottal stop, which is inserted before a vowel-initial base word regardless of whether the base word is realized by itself /ʔana:/ ‘child’, or undergoes further derivations /ʔa~ʔana:/ ‘children’.¹⁹ However, as shown above, many items have free variations in terms of the presence or

¹⁸ *i+i > /i:/ has not been attested so far, but this seems to come from the fact that Arta shows a low retention rate of PMP etyma.

¹⁹ There has been a discussion over the phonological status of the glottal stop in many Philippine and Formosan languages, which I will not get into in this paper. The author, however, consider vowel-initial glottal stop to be a phoneme rather than just a *phonetic* variant of /zero/. The glottal stop can be a target of geminates like other consonants (e.g. *meC- subəg* > *messubəg*, and *meC- a:du* > *meʔʔa:du*), thus it has a phonological status as a phonemic consonant.

absence of /ʔ/ as in (29), while some of the items do not allow the presence of /ʔ/, as in (30). And the loss of /ʔ/ triggers a vowel lengthening.

This kind of vowel lengthening by vowel fusions can be illustrated as follows. By the loss of an intervocalic consonant, two vowels adjoin each other. This leads to the re-interpretation of two homogeneous vowels as one long vowel. Note that this process again involves the conservation of mora count. After the resyllabification, the number of moras remains the same unless the structure violates the moraic constraint on the syllable structure. As mentioned above, the reason why *dukut and *sakay are realized with a short vowel *du*t and *sa*y, respectively, is that a trimoraic syllable such as ***du:t* and ***sa:y* is not allowed by that synchronic constraint.



Vowel fusions are also responsible for the development of two new phonemes /e:/ and /o:/. /e:/ emerged after the loss of *h and *k in the following items.

- (32) *ahi, *aki > /e:/
 *unahik > *une*:²⁰ ‘climb a mountain’
 *bahi > *beb~be*: ‘aunt’ (with a reduplication)
 *laki > *lel~le*: ‘uncle’ (with a reduplication)
 *maki- > *me:-*, *mi:-* (comitative)²¹

- (33) *a+u > /o:/
 *dahun > /do:n/ ‘leaf’ (e.g. *do:n=i*)
 *lahud > /di-lo:d/ ‘downstream’ (e.g. *dilo:d=i*) (with the fossilized prefix *di-*)

The changes *a+i* > /e:/ and *a+u* > /o:/ are again observed in synchronic morphophonemic alternations. If *ma-/maka-* or *pa-* (causative marker) is prefixed to a base beginning with /i/ or /u/, vowel fusions occur, realized as /me:/, /make:/, and /pe:/ on the one hand, and /mo:/, /mako:/, and /po:/ on the other. This again seems to involve the loss of the intervening glottal stop. This type of prefixation, however, hardly exhibits the variant with /ʔ/ except the case in (36).

- (34) *maka-/ma- (potentive), pa- + i > /make:/, /me:/, /pe:/
 *maka-ʔidəm > *make:dəm* ‘sleepy’
 *maka-inum > *make:nom* ‘drunk’
 *ma-ʔidəm > *me:dəm* ‘sleepy’
 *i-pa-ʔita (see) > *ipe:ta* ‘show’
- (35) *ma-ʔuras-an > *mo:rasan* ‘can be cleaned’
 *ma-ʔulit-an > *mo:litan* ‘can be peeled’
 *pa-ʔudiŋ-ən > *po:diŋən* ‘darken’

²⁰ The phonological change in *unahik > *une*: involves the reduction of mora; *nahik contains three moras, whereas *ne*: contains two moras. This reduction seems to come from the moraic constraint on the syllable, that is, a trimoraic syllable such as *ne::* is not allowed in the language.

²¹ See Table 6 in the Appendix for a full list of verbalizing affixes.

- (36) *ma-ʔune: > *maʔune:* ~ *mo:ne:* ‘can climb’

It should be noted that vowel fusions did not occur in the case of vowel + glide clusters, such as *ay and *aw. In Casiguran Agta, *ay was fused into /ɛ/ (Headland and Healey 1974), but that is not the case in Arta. With one exception in which *patay is reflected as /pati/, *ay is reflected as /ay/.

- (37) *ay > /ay/ (cf. *ay > /ɛ/ Casiguran Agta)
 *aNay > *anay* ‘go’ (cf. Casiguran Agta *aŋɛ*)
 *anay > *anay* ‘termite’ (cf. Casiguran Agta *ane*)
 *wasay > *wasay* ‘ax, hatchet’
 *patay > *pati* ‘die’

- (38) *aw > /aw/
 *ulitaw > *ulitaw* ‘unmarried man’
 *tahaw > *di-taw* ‘inside’ (with the fossilized prefix *di-*)
 *biRɲaw > *birɲaw* ‘fly (n.)’
 *lɔtaw > *lɔtaw* ‘float’

The traditional analysis in which /y/ and /w/ constitute part of a diphthong is not applied to Arta. /y/ and /w/ in the language are treated as consonants in synchrony as well. Pronominal enclitics such as =*ku* are realized differently depending on whether the preceding word ends with a consonant (realized as =*u*), or with a vowel (= *ku*). /y/ and /w/ exhibit the same pattern as other consonants (e.g. *tataw* > *tata:w=u*, not *tataw=ku* ‘I know’). There is no reason that /y/ and /w/ in the coda position should be analyzed separately from other consonants, and it is inappropriate to consider the vowel-glide sequences as diphthongs.

In §4.2, it is observed that /e/ and /o/ in borrowed items are always long vowels unless the length is overridden by the constraint on the syllable structure. It seems difficult to understand the reason why only these two vowels are always long. However, the above discussion on the historical development of /e:/ and /o:/ provides one reasonable account. That is, the feature of length which the two vowels acquired via the fusion of *a+i and *a+u must be applied to the vowels in loanwords. In the first stage, /e:/ and /o:/ were seen exclusively in the items which underwent vowel fusion. But in the process of interpreting loanwords containing [ɛ] and/or [ɔ] as the instances of the two existing vowels /e:/ and /o:/, the feature [+long] also penetrated borrowed items.

5 Competing motivations

The above section showed that long vowels in Arta are observed in inherited forms from PMP, and they are captured by compensatory lengthening and vowel fusion, both of which are subsumed under the principle of mora-count conservation. This suggests that the mora-based principle played a significant role in Arta, as well as the moraic constraint on the eligible syllable structure. However, this does not imply that the mora affects the phonetic and phonological organization in the language. Rather, a more naturalistic view would be that it is at work interacting, and sometimes competing, with other factors which may affect the organization (cf. “competing motivations” (DuBois 1985)). Some apparent exceptions to the mora principles tell us the interactions between the mora and other factors such as sociolinguistic and phonetic aspects.

The first case in which the regularity of the mora may be affected involves borrowings. The two lexical items *trabajo* in Spanish and *truck* in English were borrowed into Arta, meaning ‘work’ and ‘car’, respectively. The peculiar feature with the two items is that they have an onset cluster *tr*. This structure seems unstable in the phonology of Arta; in fact, they exhibit the following phonetic variations with/without vowel insertions.

- (39) Spanish *trabajo* > Arta [traba:hu] ~ [t̄araba:hu] ~ [taraba:hu] ‘work’
 English *truck* > Arta [trak] ~ [t̄arak] ~ [tarak] ‘car’

The onset cluster consisting of /t/ and /r/ may or may not have the insertion of [a] with different lengths. It is difficult to assume abstract phonological representations for each lexical item, and difficult to posit any specific syllable structure or mora count. Rather, this should be considered to be a case in which the application of mora count is unstable after adopting an unstable structure through borrowing.

A more puzzling case is found in inherited forms. The following two items are not explained by the borrowing of “unstable structure” from outside of the language, because they seem to be reflexes of PMP *diya and *tiyan, respectively. However, their reflexes again show instability in terms of the mora count and syllable structure.

- (40) *tiyan > [tʰi(j)an] ~ [tʰʌn] ‘belly’
 *diya > [dʰi(j)a] ~ [dʒa] ‘to him/her/it’ (3SG.OBL)

Each item may be pronounced like a bisyllabic word or with the palatalization of /t/ and like a monosyllabic word ([j] is parenthesized because whether a glide should be recognized is difficult to judge purely in terms of a phonetic observation). Furthermore, even when there is no palatalization of /t/, [i] is pronounced as shorter (hence the diacritic above [i]), and it is not obvious as to whether the words should be regarded as monosyllabic or bisyllabic words. This of course means the indeterminacy of the mora count as well.

This indeterminacy might be interpreted as the case in which the items are on the way of a gradual language change from bimoraic to monomoraic words. This, however, seems overgeneralized. It does not explain exactly why these particular items undergo the change in the mora count, and it is shown in §4.3 that the mora count was conserved, rather than reduced, in the process of phonological change. A more crucial factor applying to this particular case would be attributed to phonetic aspects. This seems to involve the fact that a sequence of two segments *iy shares the same phonetic features: front, close, and unrounded, thus it is easy to produce with a reduction. And more crucially, the articulatory movement from /t/ or /d/ to /a/ constitutes a single unidirectional movement from a complete closure to a fully opened position. The pressure by the ease of this particular articulatory movement may facilitate a reduction to monosyllabic words by kicking out the principle of the mora-count conservation.

As Blust (2013) discusses, mid-central vowel /ə/, or schwa has a systematic irregularity in terms of the mora count. The following four items in Arta show the irregularity, in which *ə is reflected with no mora value.²²

- (41) *tuqəlan > /tulan/ ‘bone’
 *pahəzam > /padam/ ‘lend’
 *bituqən > /bitun/ ‘star’
 *ka-ama-ən > /ka:man/ ‘big’

In the above cases, *tuqə is reflected as /tu/, *pahə as /pa/, *tuqən as /tun/, and *ma- -ən as /man/. The mora, which would be borne by other vowels, is not associated with the schwa. Furthermore, the following items containing schwa in the final syllable exhibit an irregular reflex. By the loss of final coda consonants, CVC syllables are reflected as monomoraic syllable with a short /a/.²³

²² Note that the reflex of PMP *baqəRu is *bu:ru* ‘new’ with a long vowel. This reflex exhibits an exceptional pattern. This might involve an intermediate stage *baqRu (by the loss of schwa with *q shifting to the coda position), which could be a target of compensatory lengthening after the loss of the coda *q.

²³ It is not obvious why the reflex of *CəC is *Ca*, rather than any other vowel. However, this reflex may be relevant to the fact that when schwa occurs in the position which is required to be lengthened, the segment is phonetically realized as a sound similar to [a].

- (42) *abək > *aba* ‘mat’
 *utək > *uta* ‘brain’
 *pədək > *pada* ‘step on’
 *mutəq > *muta* ‘gummy section of eyes’
 *luhəq > *luwa* ‘tears’
 *basəq > *bisa* ‘wet’ (with low vowel fronting (LVF): *a > /i/)

The relationship between the schwa and the mora count differs from the case in other vowels. Phonologically, /Cə/ is treated as zero mora and /CəC/ as one mora. Note that this generalization is established from a phonological or emic perspective. Seen from a phonetic or etic perspective, the length cannot be zero because it must have a physical length. Considering the internal logic of the language according to which syllables containing schwa are treated phonologically, it can be argued that the language treats the syllables with schwa as having one mora fewer than the syllables with other vowels. However, it is likely that this phonological treatment comes from the phonetic shortness of schwa. In explaining the relationship between stress and schwa, Blust (2013) claims that “[t]his behaviour derives from subphonemic differences of length in the vowels of PAN and their reflexes in many daughter languages, in which the schwa appears to be extra short” (ibid.:256). It seems also to be the case in Arta. The irregular reflex in the language may come from the phonetic peculiarity of schwa and historical reason which dates back to PAN.

6 Conclusions

This paper attempted to argue that the development of long vowels in Arta, which once lost the old accentual system, is largely explained by the notion of the mora. The first part of this paper discussed the validity of applying the mora to the synchronic phonology in Arta, in which it is argued that some allomorphic distribution is conditioned depending on whether the preceding syllable is monomoraic or bimoraic, and that vowel length alternation is explainable by the moraic constraint on the syllable structure. There are three cases in which a long vowel occurs in Arta: onomatopoeia, borrowings, and inherited forms from PMP. It is shown that vowel length in the inherited forms could be explained at least by assuming the principle of mora-count conservation as seen in compensatory lengthening and vowel fusion, as well as the moraic constraint on the syllable structure.

There are several items which seem not to be explained by compensatory lengthening or vowel fusion. The forms listed in (43) and (44) have a long vowel, which probably requires some other phonological and/or morphological mechanisms to explain.

- (43) *hadu > *a:du* ‘many, much’ (cf. Northern & Southern Alta *ádu*)
 *ikan > *i:yan* ‘fish’
 *zuRu > *di:ru* ‘soup’
 *ikəj ‘cough’ > *i:yər* ‘phlegm’²⁴
- (44) *u > /o:/ in doubled monosyllables
 *kutkut > /ko:tko:t/ ‘dig’
 *pukpuk > /po:kpo:k/ ‘beat (with a hammer)’
 *tuktuk > /to:kto:k/ ‘top, summit’
 *həyup > /yo:pyo:p/ ‘blow on’

²⁴ This item may not be an inherited form because *j is reflected as /r/, not /d/ (see Table 1).

The following items are the ones whose sources remain to be identified:

- | | |
|---|--|
| (45) <i>alilyo:gən</i> ‘kind of ghost’ | <i>di:muy</i> ‘bathing’ (cf. Southern Alta <i>di:muy</i>) |
| <i>atti:</i> ‘exist’ (cf. Ilokano <i>adda</i>) | <i>diso:no:</i> ‘inside’ |
| <i>aydi:</i> ‘and’ | <i>dupu:</i> ‘old man’ |
| <i>ba:kəw</i> ‘grain, corn’ | <i>o:gip</i> ‘lying’ |
| <i>baruwa:si</i> ‘clothes’ | <i>pulot/pulo:t/</i> ‘loincloth’ (cf. Maranao: <i>mampoyot</i>) |
| <i>be:kut</i> ‘ghost’ | <i>pu:nəd</i> ‘rain’ |
| <i>bi:lat</i> ‘python’ | <i>sa:ŋan</i> ‘how many’ |
| <i>bidi:yu</i> ‘buri palm’ | (cf. Central Cagayan Agta <i>haŋan</i>) |
| (cf. Casiguran Dumagat <i>bidi:yo</i>) | <i>si:paŋ</i> ‘one’ |
| <i>biyu:ŋət</i> ‘night’ | <i>ta:me:ta</i> ‘different’ |
| <i>de:kət</i> ‘stickiness’ | = <i>te:</i> ‘only’ |

In fact, the sources of many items in Arta are not identified yet because of its low retention rate of PMP etyma. Further studies need to be conducted to reveal how long vowels are developed in the items observed.

Appendix: Sets of grammatical forms in Arta

Table 2: Pronominal forms

	TOP	ABS	GEN/ERG	OBL
1SG	<i>tən</i>	= <i>tən</i>	= <i>ku</i>	<i>dən</i>
1PL	<i>tami</i>	= <i>ami</i>	= <i>mi</i>	<i>dami</i>
1+2SG	<i>tita</i>	= <i>ita</i>	= <i>ta</i>	<i>dita</i>
1+2PL	<i>titam</i>	= <i>itam</i>	= <i>tam</i>	<i>ditam</i>
2SG	<i>taw</i>	= <i>taw</i>	= <i>mu</i>	<i>daw</i>
2PL	<i>tam</i>	= <i>am</i>	= <i>muyu</i>	<i>dam</i>
3SG	<i>siya</i>	= <i>siya</i> , \emptyset	= <i>na</i>	<i>diya</i>
3PL	<i>tidi ~ tidu</i>	= <i>tid</i>	= <i>di</i>	<i>did, didu</i>

Table 3: Demonstratives

	TOP	ABS	GEN/ERG	OBL
PROXIMAL SG	<i>si:yəy</i>	<i>a:yi:</i> = <i>i</i>	<i>ni/na a:yi:/ayni</i> = <i>ni</i>	<i>ti/ta a:yi:</i> = <i>ti</i>
MEDIAL PL	<i>satidi:</i>	(<i>ay</i>) <i>tidi a:yi:</i>	(<i>ay</i>) <i>didi a:yi:</i>	(<i>ay</i>) <i>didi a:yi:</i>
MEDIAL SG	<i>sayna</i>	<i>a:yina,</i> = <i>ina</i>	<i>ni/na ayna</i> = <i>nina</i>	<i>ti/ta ayna</i> = <i>tina</i>
DISTAL PL	<i>satidi:na</i>	(<i>ay</i>) <i>tidi:na</i>	(<i>ay</i>) <i>didi:na</i>	(<i>ay</i>) <i>didi:na</i>
DISTAL SG	<i>saya</i>	<i>a:ya:, =ya:</i>	<i>ni/na a:ya:</i>	<i>ti/ta a:ya:</i> = <i>ta</i>
	PL	<i>satiddya:</i>	(<i>ay</i>) <i>tiddya</i>	(<i>ay</i>) <i>didi a:ya:</i>

Table 4: Determiners

		TOP/ABS	GEN/ERG	OBL
DEFINITE	SG	PERSONAL	<i>ti</i>	<i>ni</i>
		COMMON	<i>i</i>	<i>ni</i>
	PL		<i>tidi</i>	<i>didi</i>
INDEFINITE			\emptyset	<i>na</i>
				<i>ta</i>

Table 5: Specifiers

SPECIFIC ²⁵	= <i>i</i>
SPECIFIC (PAST)	= <i>ti</i>
SPECIFIC (PLURAL)	= <i>tidi</i>

Table 6: Verbs

VERBS	SIMPLE NONPAST	SIMPLE PAST	PROGRESSIVE/ NOMINALIZATION
DYNAMIC INTRANSITIVES (ACTOR-VOICE)			
PUNCTUAL (*um)	<um>	<in><um>	–
DISTRIBUTIVE ²⁶ (*maN-)	maN-	(mi)naN-	paC- ²⁷
DURATIVE (*maR-)	maC-	(mi)naC-	paC-
COMITATIVE (*maki-)	me:-, meC-	ne:-, neC-	pe:-
DYNAMIC TRANSITIVES (UNDERGOER VOICES)			
PATIENT-TRANSITIVE	-ən	<in>	paC- -ən
LOCATION-TRANSITIVE	-an	<in> -an	paC- -an
CONVEYANCE-TRANSITIVE	i-	(i)ni-	paC-
<i>ma-/maka-</i> POTENTIVES			
INTRANSITIVE	maka-	naka-	–
TRANSITIVE -ən	ma-	(mi)na-	–
-an	ma- -an	(mi)na- -an	–
i-	ma-, me:-	(mi)na-, (mi)ne:-	–
STATIVES			
AGENTIVE STATIVE	tiC-	–	–
PATIENTIVE STATIVE	maŋa:-	–	–
POSSESSIVE STATIVE	makaN-	–	–
CAUSATIVE DERIVATIONS			
INTRANSITIVE CAUSATIVE	mampa-, mama-	(mi)nampa-, (mi)nama-	pappa-
TRANSITIVE CAUSATIVE			
-ən	pa- -ən	pina-	pappa- -ən
-an	pa- -an	pina- -an	pappa- -an
i-	ipa-, pe:-	nipa-	pappa-

²⁵ Definiteness refers to a given-new distinction varying depending on the hearer's knowledge assumed by the speaker (Prince 1981), whereas specificity is another grammatical distinction relevant to a speaker's knowledge as to whether s/he can identify the referent or not, as is well-known in the two distinct readings of the following sentence: *I want a car*. See Lambrecht (1994).

²⁶ *maN-* prefixation does not trigger the deletion of stem-initial consonants (e.g. *taradtad* > *man-taradtad* 'escape, run away', *pili* > *mam-pili* 'choose').

²⁷ *maN-* and *maC-* show syncretism in progressive/nominalization form.

Table 7: Interrogatives

what	<i>a:nu</i>
who	<i>tatin</i>
when	<i>tanakan</i>
where	<i>adin</i> (past: <i>adinti</i> , present/future: <i>adi:ni</i>)
why	<i>ataʔay ~ adaʔay</i>
how (manner)	<i>kassandi</i>
how many/much	<i>sa:ŋan, saŋan</i>

Table 8: Negator, existentials and quantifiers

SENTENTIAL NEGATION	<i>awan</i>
EXISTENTIAL	<i>atti: (+ABS)</i>
EXISTENTIAL NEGATION	<i>awan (+OBL)</i>
many, much	<i>meʔa:du</i>
few, little	<i>killək</i>
one (of)	<i>gissa</i>
some (of)	<i>a:duwan</i>
all (of)	<i>attanan, atanan</i>

Table 9: Second-position enclitics

ASPECTUAL ENCLITICS	
again	<i>=mandi</i>
already, soon	<i>=di ~ =d</i>
still	<i>=tep</i>
just do, do a few	<i>=pa ~ =p</i>
EVALUATIONAL ENCLITICS	
only	<i>=te:</i>
PRAGMATIC ENCLITICS	
REQUEST	<i>=nen</i>
HYPOTHETICAL	<i>=mina</i>
MIRATIVE	<i>=sika, =mat</i>
HEARSAY	<i>=an</i>
META-COMMUNICATIVE ²⁸	<i>=hug, =ay</i>

References

- Blust Robert A. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Taipei: Academia Sinica.
- Blust, Robert. 2013. *The Austronesian languages*. Revised edition. ANU. URL: <http://hdl.handle.net/1885/10191>
- Blust, Robert, and Stephen Trussel. In progress. *Austronesian Comparative Dictionary*, Web Edition. www.trussel2.com/ACD.
- Davis, Philip W., and Angel D. Mesa. 2000. *A dictionary of Yogad*. München: Lincom.

²⁸ The formatives *=hug* and *=ay* which we call here “meta-communicative” are used to embed into the utterance the implicature that the utterance is conveying redundant information which should have been in the common ground because the information is already mentioned (*=hug*), or because it can be inferred from the linguistic/non-linguistic context (*=ay*).

- Du Bois, John W. 1985. Competing motivations. In *Iconicity in syntax*, ed. by John Haiman, 343-365. Philadelphia: John Benjamins Publishing.
- Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20:253-306.
- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and case studies*. Chicago: University of Chicago Press.
- Headland, Thomas N. 2003. Thirty endangered languages in the Philippines. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 47:1-12.
- Headland, Thomas N., and Janet D. Headland. 1974. *A Dumagat [Casiguran]-English dictionary*. Canberra: Pacific Linguistics.
- Headland, Thomas N., and Alan Healey. 1974. Grammatical sketch of Casiguran Dumagat. *Papers in Philippine Linguistics* No.5, 1-54. Pacific Linguistics Series A 43. Canberra: Pacific Linguistics.
- Kimoto, Yukinori. 2017. A grammar of Arta: A Philippine Negrito language. Ph.D. dissertation, Kyoto University.
- Ko, Albert Min-Shan, Chung-Yu Chen, Qiaomei Fu, Frederick Delfin, Mingkun Li, Hung-Lin Chiu, Mark Stoneking, and Ying-Chin Ko. 2014. Early Austronesians: Into and out of Taiwan. *The American Journal of Human Genetics* 94:426-436.
- Komatsu, Hideo. 1989. Nihongo no Rekishi: On-in [The History of Japanese: Phonology]. *The encyclopaedia of linguistics*, Vol.2, ed. by Takashi Kamei, Rokuro Kono, and Ei-ichi Chino, 1634-53. Tokyo: Sansei-do.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In *Radical pragmatics*, ed. by Peter Cole, 223-255. New York: Academic Press.
- Rasmussen, Morten, et al. 2011. An aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334(6052):94-98.
- Reid, Lawrence A. 1971. *Philippine minor languages: Word lists and phonologies*. Oceanic Linguistics Special Publication 8. Honolulu: University of Hawaii Press.
- Reid, Lawrence A. 1979. Towards a reconstruction of the pronominal systems of Proto-Cordilleran, Philippines. In *South-East Asian linguistic studies* 3, ed. by Nguyen Dang Liem, 259-275. Pacific Linguistics C-45. Canberra: Australian National University.
- Reid, Lawrence A. 1982. The demise of Proto-Philippines. In *Papers from the Third International Conference on Austronesian Linguistics, vol. 2: Tracking the travellers*, ed. by Amran Halim, Lois Carrington, and S. A. Wurm, 201-216. Canberra: Pacific Linguistics.
- Reid, Lawrence A. 1989. Arta, another Philippine Negrito language. *Oceanic Linguistics* 28(1):47-74.
- Reid, Lawrence A. 1991. The Alta languages of the Philippines. In *VICAL 2, Western Austronesian and contact languages: Papers from the Fifth International Conference on Austronesian linguistics*, ed. by Ray Harlow, 265-297. Auckland: Linguistic Society of New Zealand.
- Reid, Lawrence A. 1994. Possible non-Austronesian lexical elements in Philippine Negrito languages. *Oceanic Linguistics* 33(1):37-72.
- Reid, Lawrence A. 2006. On reconstructing the morphosyntax of Proto-Northern Luzon, Philippines. *Philippine Journal of Linguistics* 37:1-64.
- Reid, Lawrence A. 2009. The reconstruction of a dual pronoun to Proto Malayo-Polynesian. In *Discovering history through language: Papers in honour of Malcolm Ross*, ed. by Bethwyn Evans, 461-477. Canberra: Pacific Linguistics.
- Reid, Lawrence A. 2013. Who are the Philippine Negritos? Evidence from language. *Human Biology* 85:329-358.

- Reyes-Centeno, Hugo, Silvia Ghirotto, Florent D etroit, Dominique Grimaud-Herv , Guido Barbujani, and Katerina Harvati. 2014. Genomic and cranialphenotype data support multiple modern human dispersals from Africa and a Southern route into Asia. *PNAS* 111(20):7248–7253.
- Robinson, Laura C, and Jason William Lobel. 2013. The Northeastern Luzon subgroup of Philippine languages. *Oceanic Linguistics* 52(1):125–168.
- Ross, Malcolm. 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2):1–24.
- Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th Birthday*, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 521-563. Language and Linguistics Monograph Series Number W-5. Taipei: Institute of Linguistics, Academia Sinica.
- Rubino, Carl Ralph Galvez. 2000. *Ilocano dictionary and grammar*. Honolulu: University of Hawaii Press.
- Zorc, R. David. 1979. On the development of contrastive word accent: Pangasinan, a case in point. In *South-East Asian linguistic studies* 3, ed. by Nguyen Dang Liem, 241-258. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Zorc, R. David. 1990. The Austronesian monosyllabic root, radical or phonestheme. In *Linguistic change and reconstruction methodology*, ed. by Philip Baldi, 175–194. Berlin: Mouton de Gruyter.
- Zorc, R. David. 1993. Overview of Austronesian and Philippine accent patterns. In *Tonality in Austronesian languages*, ed. by Jerold A. Edmondson and Kenneth J Gregerson, 17–24. Honolulu: University of Hawaii Press.

RE-EVALUATING THE POSITION OF IRAYA AMONG PHILIPPINE LANGUAGES¹

Lawrence A. Reid
University of Hawai‘i
National Museum of the Philippines
reid@hawaii.edu

Abstract

Iraya (iry) of Mindoro has been grouped with Central Luzon languages primarily because of a shared sound change, but many questions remain because of the unique features of the language and because of the unusual phenotypic features of the people. This paper claims that Iraya people are descendants of Negrito groups that fully occupied Mindoro before the arrival of Austronesian-speaking peoples, and Iraya are the last remaining such group in Mindoro. The phenotypic features of Iraya are the result of inter-marriage with in-migrating groups from areas to the south who eventually forced their retrenchment into the most northerly mountains of the island. The unusual linguistic features of Iraya are considered to be a combination of language contact with other Philippine languages, and possibly also with languages from outside the Philippines.

Keywords: Negritos, language contact, Mangyan, Mindoro, Iraya
ISO 639-3 codes: iry

1 Introduction

Iraya is one of the more than 150 Malayo-Polynesian (MP) languages spoken in the Philippines. It is spoken by an estimated 5000 of the older Iraya population on the island of Mindoro, the 7th largest island in the Philippines.² There are at least seven mutually unintelligible languages spoken on the island, of which Iraya is the most northerly and is adjacent across the strait from the Batangas area of Luzon, south of Manila, where Tagalog is the main language (see Map 1). The indigenous languages of Mindoro and their cultures are often referred to as Mangyan. Various articles have appeared dealing with these languages. Tweddell (1958) primarily deals with the phonology and morphology of Iraya. Zorc (1974) is an extensive discussion of the relationships between the various languages of Mindoro, dealing with a wide selection of data (summarized below in sec. 3). Barbian (1977) provides additional data to those already available. Although language contact is a prominent explanation of some of the variability that occurs, this paper is the first that proposes a prehistoric scenario that

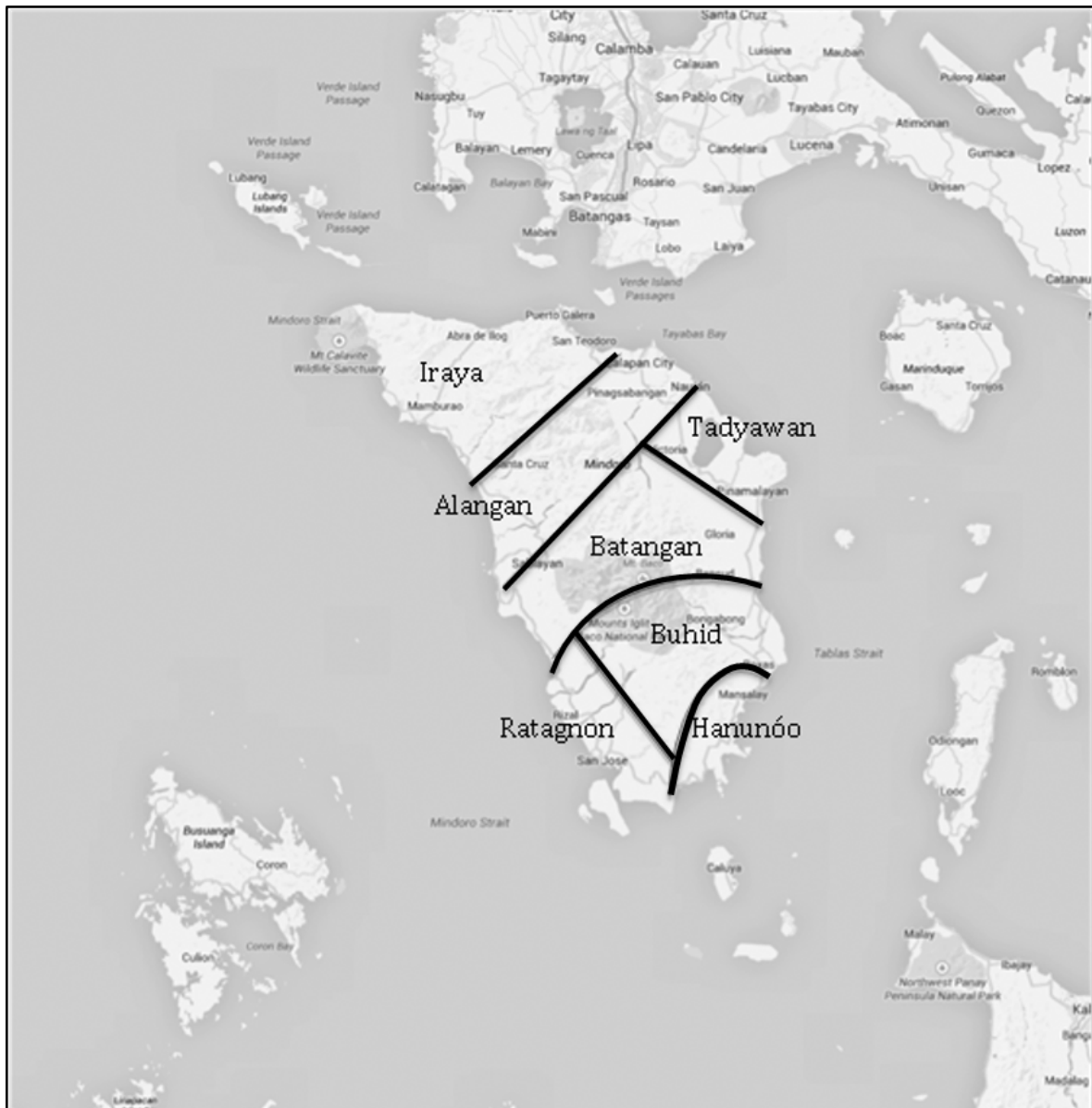
¹ This article is a revised version of presentations that were made in several venues, “*Baa, Baa, Black Sheep: Features distinguishing Iraya from most other Philippine languages,*” to Minpaku Linguistics Circle, Osaka, Japan, Sept. 7, 2014; and to Welcome Meeting, Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies, Sept. 10, 2014; ‘Identifying prehistoric population trajectories: Who influenced Iraya?’ to Migrations and Transfers in Prehistory: Asian and Oceanic Ethnolinguistic Phylogeography, University of Bern, Switzerland, July 28-30, 2014; and ‘Re-evaluating the position of Iraya among Philippine languages’ to the 13th International Conference on Austronesian Linguistics, Academia Sinica, Taiwan, July 18-23, 2015. I wish to thank all who commented on each of the presentations, and for comments subsequently received by email from Alexander Adelaar, Anthony Jukes, Hsiu-chuan Liao, and Richard McGinn, all of which have contributed valuable information, but not all of which is reflected in the present article. I am, as always, finally responsible for the present version.

² Lewis et al. (2016) give a population figure of 10,000, which is a rough estimate done by the Overseas Missionary Fellowship (OMF) in 1991. This figure apparently consists of all who consider themselves to be part of the Iraya cultural group, most of whom no longer speak the Iraya language. The estimate given here is possibly inflated.

accounts for some of the variability and the geographic location of some of these groups, particularly Iraya.

Iraya ('person' in the Iraya language) is derived from Proto-Malayo-Polynesian (PMP) *ʔi- 'person from' + *daya 'interior, upland' (Blust & Trussel Ongoing), and is distinctly different in a number of features from the other languages of Mindoro, and of the Philippines. These features include different pronominal forms and functions, several changes in the structure of noun phrases, changes in the patterns of verb structures, changes in word order and other sentential features not commonly found in other Philippine languages. Iraya itself is dialectally diverse, with populations in two provinces, Occidental Mindoro and Oriental Mindoro. The data given in this paper, unless otherwise noted, is from three periods of fieldwork in Oriental Mindoro. In November, 2013, with Avelino Pampilo (45?) and Mariano Garcia (60?) in Talipanan, White Beach, Puerto Galera, and in June-July and November, 2014, with Islas Malinaw (70?) and Elma Malinaw (35?) in Da Pirmida, Baclayan, Puerto Galera. Several hundred lexical items, and a corresponding list of sentences were elicited and recorded. Each assistant likewise recorded one or more narrative texts, which were transcribed and translated. All sound files and transcribed data have been deposited with The Mangyan Heritage Center, Calapan City, Oriental Mindoro, Philippines.

Map 1: *Mindoro Island and its languages (adapted from Barbian 1977:16)*



The question being asked in this paper is to what extent these features are retentions of earlier Philippine languages, the result of innovations that are unique to Iraya, or are developments that are the result of contact diffusion. One factor that needs to be considered is that Iraya people have phenotypical features that distinguish them from other Philippine groups. They typically have wavy to curly hair, a feature found to a more pronounced degree in Negrito populations of the Philippines, suggesting that these were also a Negrito group that has been heavily influenced by in-migration and intermarriage with non-Negrito groups. Tweddell (1958:2) noted the Negrito-like features of Iraya people. He also referred to Beyer (1921) who classified them as 'Sakai', his supposed second group of immigrants into the Philippines following 'Java man'. The Negrito connection is supported by HUGO (2009, Fig. 1) which reports on a genetic analysis of ancestral alleles of 75 populations. A maximum likelihood tree shows Iraya grouped with Mamanwa, Agta, Aeta (i.e., Ayta) and Ati populations, groups that self-identify as Negrito.

The general claim being made is that evidence suggests that like all surrounding areas of the Philippines, Mindoro was widely occupied by groups of Negrito people, before MP-speaking peoples arrived. The first contact in Mindoro with MP speakers was probably with people from the Batangas area where they had learned and were speaking an early version of what has now developed into the Central Luzon group of languages, so that all of Mindoro was initially occupied by speakers of a language that carried the features of Central Luzon languages. Subsequently migrants from the western Visayas to the south-east of Mindoro and from Palawan to the south-west intermarried with the local Negritos.

Over several thousand years, in-migration has resulted in a forced retrenchment of Negritos to the most northerly mountainous areas of the island. In-migration is still happening, with Tagalog being the language primarily spoken in lowland areas in the north of Mindoro, and by the younger generations of Iraya people. Ilokano is spoken in some communities in the coastal areas of Occidental Mindoro, while in the southeast, the major lowland language is a Central Bisayan language (Rombloanon) spoken in Tablas and other islands to the east. Ratagnon in the south of the island is one of the three dialects of Cuyunon, a West Bisayan language spoken directly to the south in the Semirara group of islands (Zorc 1974:561, see also Hammarström et al. 2016).

While features of their original MP language are still found in some of the southern group (see sec. 4), the southern Mangyan languages are now classified as part of the Greater Central Philippine subgroup (Blust 1991). Only the three languages in the north of the island, Iraya, Alangan and Tadyawan, are considered to be related to the Central Luzon subgroup of Philippine languages, and only Iraya people still appear to be phylogenetically distinct from other groups in the island.

This paper is organized as follows. Section 2 outlines some of the features which distinguish Iraya from most other languages of the Philippines. Section 3 provides information about the position of Iraya and the other northern languages of Mindoro based on lexical data and application of the comparative method. Section 4 considers various factors that suggest contact diffusion into Iraya and its related languages. Section 5 discusses the claim that the Iraya people are Negrito people that have been forced by in-migrating people into the most northerly mountains of Mindoro. The final section provides a summary of the issues discussed in the paper.

2 About the Iraya language

Iraya is unique in many ways among the languages of the Philippines. This section discusses first the pronouns, then features of noun phrases and verbal structures which are unusual for Philippine languages.

2.1 Personal pronouns

The features that distinguish Iraya personal pronouns from most other Philippine languages include the following: a) loss of case distinctions in non-singular forms; b) presence of dual pronouns for all non-singular forms, only one of which is inherited; c) the extension of earlier Set 2 forms for Set 1 functions in dual and plural forms; d) unique Set 2 forms; e) their position. All core personal pronouns are fronted before their verb or noun head.

Table 1 presents the two sets of Iraya personal pronouns used in texts and elicited materials (Reid 2013a). Case marking appears only in the three singular forms. These distinguish between Set 1, whose functions include Nominative (of actor voice and non-actor voice constructions), Topic and nominal Predicate, and Set 2, whose functions include Genitive (possessor in a noun phrase, as well agent of non-actor voice constructions) and Oblique, in which case the personal pronoun is preceded by the form *sa* ‘locative, oblique’. All non-singular personal pronouns have a single form for each of the functions that are distinguished by the singular forms. Set 1 forms are disyllabic (CV.CV), while Set 2 forms alternate between a disyllabic form, and a reduced monosyllabic (CVC) form. Only one of the plural Set 2 forms has a reduced CVC form.

Following the paradigmatic structure of most Austronesian languages, Iraya maintains a distinction between an inclusive 1st person plural form (‘we all’) and an exclusive 1st person plural form (‘we but not you’). However, Iraya is unique among Philippine languages in having dual forms for each of the corresponding plural forms. The only dual form which is inherited is *kita* ‘1st person inclusive dual, we-two’ (Reid 2009, 2016). Each of the other forms is uniquely formed (for Philippine languages) with an ending derived from the Iraya form *darawá* ~ *darwa* ‘two’. A similar paradigmatic structure is found in some languages in Borneo, such as Brunei Dusun (Table 2) with case distinctions maintained only in the singular forms (Lobel 2013:146). Brunei Dusun is distinct from Iraya, however, in maintaining distinct forms for genitive and oblique cases, while Iraya maintains the distinction only with the oblique marker *sa*. Brunei Dusun has reformed all its dual and plural forms so they are not directly comparable with the corresponding Iraya personal pronoun forms.

Comparison of the singular forms and the corresponding plural forms in 2nd and 3rd persons, suggests that Set 2 plural forms have taken over the function of Set 1 forms. Set 2 singular forms in 2nd and 3rd persons each has a first syllable *ku-* whose source is unknown. It is possible that these forms are a remnant of the *ku(n)-* initial oblique forms found among the Sambalic group of Central Luzon languages. The 1st person inclusive plural personal pronoun is a direct reflex of PMP *=*tamu*.

A full discussion of the development of Iraya pronouns is not possible in this paper because of space constraints, but it is clear that they are a combination of inherited forms, forms that have been borrowed from neighboring languages, and innovated forms. One of the innovations is unique among Philippine languages, *naʔay* ~ *nay* ‘I, my’, which is clearly a semantic shift from a demonstrative ‘this one’ that occurs (probably independently) in a number of Philippine languages, including Bontok, a Central Cordilleran language of Northern Luzon.

Finally, most Philippine languages have a set of enclitic genitive pronouns, although some languages can optionally replace them with Locative or Possessive forms before the head word. In Iraya (as in almost all Mangyan languages), there are no basic enclitic forms, although Hanunó’o has a full set of innovated forms built on the genitive personal marker *ni-* (Zorc 1974:571). The data show fixed positions for Set 2 (genitive) Iraya pronouns before a verb when actor, and when functioning as a nominal possessor, before a head noun. All Set 2 pronouns obligatorily occur before the form with which they are in construction. This is true also for all the example data in Tweddell (1958:48–49). When Set 2 pronouns function as Locatives preceded by Iry. *sa*, they may optionally occur after a verb. Tweddell (1958) provides examples of Set 1 pronouns, some of which precede the verb, ex. (1)–(2), and others which follow the verb, ex. (3), where *ʔiya* occurs at the end of the sentence. However, in none of the narrative texts that I recorded does a core pronoun follow the verb, ex. (4), nor in elicited data, given in response to Tagalog sentences, where pronouns followed the verb, ex. (5)–(6).

(1) Iry. *ʔaku* *ʔagtalima*.
 NOM.1SG remember
 ‘I remember’.

(2) Iry. *ʔiya* *ʔagtukawanan* *sa* *naʔay* *naʔapun*.
 NOM.3SG speaking LOC GEN.1SG yesterday
 ‘He was speaking to me yesterday’.

- (3) Iry. *kumu=ʔani tabuyun sa naʔay ʔiya.*
 GEN.2SG=now give LOC GEN.1SG NOM.3SG
 ‘You give him to me’.
- (4) Iry. *ʔaku nagmunaʔan, ʔaku tuwaʔ ʔagpamataw, ʔaku ba nay*
 NOM.1SG before NOM.1SG here live NOM.1SG BA GEN.1SG
kalkan sa tambuʔ kayu.
 sleep.place LOC top tree
 ‘Before, I used to live here, and my sleeping place was in the top of a tree’.
- (5) Iry. *kawu nay malyag.*
 NOM.2SG GEN.1SG like
 ‘I like you’.
- (6) Iry. *kawu tuwaʔ tumuŋkaʔ.*
 NOM.2SG here sit
 ‘You sit here’.

Table 1: Iraya personal pronouns (Reid 2016)

		Set 1 (Nom/Top/Prd)	Set 2 (Gen/(Loc/Obl))
1	Singular	<i>ʔaku</i>	<i>(sa) naʔay ~ nay</i>
2	Singular	<i>kawu</i>	<i>(sa) kumu ~ kum</i>
3	Singular	<i>ʔi:ya</i>	<i>(sa) kunin ~ kun</i>
1 (excl.)	Dual	<i>(sa) kidawa</i>	
1 (incl.)	Dual	<i>(sa) kita</i>	
2	Dual	<i>(sa) kandawa</i>	
3	Dual	<i>(sa) sidawa</i>	
1 (excl.)	Plural	<i>(sa) yamən ~ yam</i>	
1 (incl.)	Plural	<i>(sa) tamu</i>	
2	Plural	<i>(sa) kuyu</i>	
3	Plural	<i>(sa) kura</i>	

Table 2: Brunei Dusun personal pronouns (adapted from Lobel 2013:147)

		Nominative	Genitive	Oblique
1	Singular	<i>kuji?</i>	<i>ku, jai?</i>	<i>jai?</i>
2	Singular	<i>ikow, =kow</i>	<i>mu</i>	<i>ijun</i>
3	Singular	<i>iyo</i>	<i>yo, ?o, o</i>	<i>diso</i>
1 (excl.)	Dual	<i>indo?</i>		
1 (incl.)	Dual	<i>dodo?</i>		
2	Dual	<i>mundo?</i>		
3	Dual	<i>yodo?</i>		
1 (excl.)	Plural	<i>jami?</i>		
1 (incl.)	Plural	<i>jati?</i>		
2	Plural	<i>muyun</i>		
3	Plural	<i>soro</i>		

2.2 Noun phrases

The structure of Iraya noun phrases is likewise unique among Philippine languages. Iraya noun phrases are not marked for case, neither do they distinguish personal from non-personal nouns, discussed in sec. 2.2.1. A gender system has developed for personal nouns, to distinguish between masculine and feminine names, see sec. 2.2.2. A subordinating ligature has also generally been lost, examined in sec. 2.2.3, and optional unique forms for marking noun phrases have developed, discussed in sec. 2.2.4.

2.2.1 Loss of noun phrase marking for case and personal nouns

PMP is reconstructed as having a variety of forms that introduced noun phrases, referred to by a wide range of labels, commonly referred to as determiners, case markers, or nominal specifiers (Reid 2002:296–297). Their functions range from marking case, such as nominative/absolute, genitive/ergative, oblique, and locative to specifying features of the following head noun, whether or not it was definite, or specific; whether or not it was a common noun or a personal noun, and in the latter case, whether it included more than the person named, forming a comitative noun phrase. The loss of formal marking of noun phrases is found throughout the Mangyan languages of Mindoro to one degree or another. In Alangan, a language geographically adjacent to Iraya, and subgrouped with it, personal nouns are always unmarked if singular (7)–(8), but require a preceding *kura* ‘3rd person plural’ if the form is comitative, regardless of the case, see ex. (9)–(10). Nominative common nouns are preceded by the marker *in*, as in ex. (8) and (10) (with equivalent Tagalog sentences, provided by Dimaano 2005, for comparison).

- (7) Tag. *dumating si Gloria kagabi.*³
 Aln. *rumateng ∅ Gloria kapuni*
 arrived Gloria last.night
 ‘Gloria arrived last night’. (Dimaano 2005, ex. 64)

³ Literal and free translations here and elsewhere are modified from the sources given to conform to Leipzig glossing rules, and for consistency.

- (8) Tag. *ipinadala ni Vivian ang sulat.*
 Aln. *piyababa ∅ Vivian in surat.*
 sent Vivian SPCF letter
 ‘Vivian sent the letter’. (Dimaano 2005, ex. 65)
- (9) Tag. *umalis sina Aniway at Oliver sa paaralan.*
 Aln. *pumanaw kura Aniway usai Oliver sa iskul.*
 leave 3PL Aniway with Oliver LOC school
 ‘Aniway and Oliver left the school’. (Dimaano 2005, ex. 66)
- (10) Tag. *binasa nina Carlo at Martin ang sulat.*
 Aln. *biyasa kura Carlo at Martin in surat.*
 read 3PL Carlo and Martin SPCF letter
 ‘Carlo and Martin read the letter’. (Dimaano 2005, ex. 67)

In Iraya, as well as in Alangan (ex. 11) and other Mangyan languages, genitive and oblique common nouns are unmarked, as noted by Zorc (1974:577) and Barbian (1997:97).

- (11) Tag. *bumili ako ng bago=ng sasakyan.*
 Aln. *agbili aku ∅ bayu pag pagsakayan.*
 buy NOM.1SG OBL new LIG car
 ‘I bought a new car’. (Dimaano 2005, ex. 59)

As Zorc (1974:577) notes, this is only significant in contrast with all of the Central Philippine languages outside of Mindoro, where some kind of oblique marker is obligatory, such as Tag. *ng* /*naŋ*/, Ceb. *ʔug*, Kin. *kang*, Kuy. *iʔ*, Pal-Abr. *it*, Agy-Kal. *ta*, Bik. *nin*, Hil. *sing*, etc. Such markers are also obligatory in North Luzon and Central Luzon languages, such as Ilk. *ti*, Bon. *ʔas*, Kpm. *king*, *keng*, etc.

Iraya has none of these case-markers. The case of a noun phrase is identified only by its position relative to the predicate, and by pragmatic considerations. There is also no marking distinction in Iraya between common and personal nouns, such as is found for example in Tag. *ang/si* and *ng/ni*.

2.2.2 Development of gender distinctions for personal names

While there is no distinct marking for common vs. personal nouns in Iraya, the language distinguishes between masculine and feminine nouns, by introducing masculine names with *laki* ‘male’ (12), and feminine nouns with *baʔi* ‘female’. The same forms can be used for all core arguments; there is no distinctive nominative and genitive marking, whereas locative phrases can be optionally preceded by *sa*. Ex. (12)–(13) are both possessive constructions where the possessor is optionally marked as a locative NP. These gender specifiers can be replaced by the form *kuyay* ‘old person’ (14), or in the case of comitative nouns by *kura* ‘3rd person plural pronoun’ (13). In texts, where gender identification is already known, either by prior reference or general knowledge, a personal noun is optionally introduced by one of the nominal specifiers, typically *ʔag*, ex. (15).

- (12) Iry. *laki Pedro tiyaʔ.*
 male Pedro this.one
 ‘This is Pedro’s’.

- (13) Iry. *sa kura Pedro tiya?*
 LOC 3PL Pedro this.one
 ‘This belongs to Pedro and others’.
- (14) Iry. *tayuna, nay apu? kuyay mar?et ba maki kamutiyán.*
 before GEN.1SG grandfather old Mar-et BA exist camote.field
 ‘Before, my grandfather Mar-et had a sweet potato field’.
- (15) Iry. *?ag manhuŋ ba ?ilukub=ani.*
 SPCF Manhung BA fall.down=already
 ‘Manhung fell down’.

2.2.3 Loss of subordinating ligatures

A characteristic feature of Philippine languages, and one that is reconstructed to PMP, is the presence of a subordinating linker, commonly referred to as a ‘ligature’ between head nouns and their modifiers, whether nominal, adjectival, demonstrative, or full relative clauses. The same subordinating ligature is typically used also before verbal complements or other subordinate structures. In many Mangyan languages the form of the ligature to mark subordinate structures is *pag* (see ex. (11) above) and examples given in Zorc (1974:576). These also include examples from Iraya, but the text and elicited data I obtained from Iraya have no instances of *pag* as a subordinating ligature, ex. (16). While other Mangyan languages use *pag* between an adjectival form and a noun, Iraya has extended the ligature *ka* to link not only numeral constructions (as do many other Philippine languages), but uses it also to link an adjectival form and a noun, as in (17)–(18). But this is only when the noun follows the adjectival form. There is no ligature when the word order is reversed, with the noun first, as in ex. (19)–(21). The only remnant of *pag* is found as an enclitic =*g* (replacing a final glottal stop) between a demonstrative and a following noun, ex. (22).

- (16) Tag. *ang naghuhuni=ng alamid*
 Iry. *?ag magbə?ət barungi?*
 SPCF noisy=LIG civet.cat
 ‘the noisy civet cat’
- (17) Han. *mayad pag balay*
 Buh. *kafi?a?un fag balay⁴*
 Tdy. *maganda pag balay*
 Aln. *magalen pag balay*
 Iry. *piya ka balay*
beautiful LIG house
 ‘beautiful house’ (Zorc 1974, ex. 1–5).
- (18) Iry. *maraw [sa?i ka ?aldaw] ba batay mamahuy ?ag nay apu?*
 and.then one LIG day BA future visit SPCF GEN.1SG grandfather

sa kun kamutiyán.
 LOC GEN.3SG sweet.potato.field
 ‘Then one day my grandfather was going to visit his sweet potato field’.

⁴ Transcription modified to substitute letter *q* with *?.*

- (19) Iry. *lakul ka dalan ~ dalan lakul*
 big LIG trail trail big
 ‘a wide trail’
- (20) Iry. *piya ka daraga ~ daraga piya*
 beautiful LIG young.lady young.lady beautiful
 ‘a beautiful young lady’
- (21) Iry. *kum tayʔən [da kum galanit rangas].*
 GEN.2SG sew SPCF GEN.2SG clothes torn
 ‘Sew your torn clothes’.
- (22) Iry. *tiya=g ʔiruʔ tiyaʔ*
 this.one=LIG dog this.one
 ‘this dog’

2.2.4 Unique marking of noun phrases

Iraya utilizes two monosyllabic forms which optionally introduce noun phrases. One is *ʔag*, the other is *da*. Neither form marks case, as either can introduce both nominative, genitive and predicative NPs. They possibly mark specificity or definiteness in combination with other factors which are currently undetermined. Tweddell (1958:65) gives precisely the same definition for both, ‘a, the, the one who, that which’, noting that *ʔag* is more specific, while *da* is more general. The form *ʔag* commonly introduces a topicalized NP in which case the head noun is definite, whether a common noun or a personal noun, as in ex. (15) above, and it can also introduce a nominative NP at the end of a sentence, as in (23). Ex. (24) shows *ʔag* marking an indefinite oblique NP in an actor voice construction.

- (23) Iry. *maki ʔaŋani gulat sa kunin suʔut [ʔag nay kaka].*
 EXIST surely injured LOC GEN.3SG chest SPCF GEN.1SG brother
 ‘My brother had a bad injury in his chest’.
- (24) Iry. *yamən tanguna, ʔaku ba badyaʔ ʔuŋaʔ dapu, yamən*
 1PL.EX before NOM.1SG BA still child still 1PL.EX

nagpanawən [ʔag ʔuway, ʔuway lakul].
 AV.make SPCF rattan rattan big
 ‘What we did before, when I was still a young child, we were working with rattan, big rattan’.

The form *da* can introduce a nominative NP, as in (21) above, and commonly introduces NPs which follow a predicate demonstrative *naba* ‘that (near)’, referring to a story just told, as in (25)–(26), an environment where *ʔag* can also be found, (27). Ex. (28) illustrates repeated, explanatory nominative phrases, marked by *da*, following *naba*.

- (25) Iry. *naba da panultulun nay apuʔ.*
 that SPCF story GEN.1SG grandfather
 ‘That is the story of my grandfather’.

- (26) Iry. *naba da naʔay kadanásan*
 that SPCF GEN.1SG experience
 ‘That was my experience’.
- (27) Iry. *naba mana ʔag balay...*
 that surely SPCF house
 ‘That was certainly a house...’
- (28) Iry. *naba mana ba da kura pamaŋən, da kura pamaŋən,*
 that surely BA SPCF 3PL food SPCF 3PL food

da paskəd ba labəy.
 SPCF called BA yam
 ‘That must be their food, their food, which is called yam’.

Language assistants commonly translate *ʔag* as Tag. *ang*, however its probable source is *pag*, a form which as noted above, commonly occurs as a ligature in other Mangyan languages, but in Iraya that no longer has a *pag* ligature, the form has been reanalyzed as a nominal specifier with loss of the initial consonant.

One further function of *ʔag* is to mark a temporal phrase, ex. (29).

- (29) Iry. [*ʔag ʔaku ba ʔibun dapu*] *ba yamən ba sata? daku sa yam pamatawan.*
 SPCF NOM.1SG BA young still BA 1PL.EX BA there.far very LOC 1PL.EX home
 ‘When I was still young, our home was in a very far place’. (IM Text 1)

While common nouns in Philippine languages are generally not marked for plurality, a variety of pluralizing forms (apart from reduplication) are found. The common pluralizer in Tagalog and other Central Philippine languages is *mga* /maŋa/, and is borrowed widely into Mangyan languages, including Iraya. But in Iraya it is replacing an earlier common noun pluralizer *pad*, whose source is unknown, ex. (30)–(31).

- (30) Iry. *ʔag ʔanubliŋ kay ba magsələd sa pad lubut, lubut ʔoroŋ...*
 SPCF blowflies also BA entered LOC PL hole hole nose
 ‘The blowflies also entered into the holes, the holes of my nose..’.
- (31) Iry. *da? pad kayu lakul ba makaya nagpaŋatiglu?...*
 and PL tree big BA seem breaking
 ‘and the big trees seemed to be breaking...’

2.3 Other structural features

One of the unique features of Iraya is the frequency of occurrence of a monosyllabic form *ba*, discussed in detail in sec. 2.3.1. The other unusual feature of Iraya is the use of perfective forms of the verb to mark potential or future forms, explained in sec. 2.3.2.

2.3.1 BA

The form /ba/ is probably the most frequently occurring monosyllabic form in the language, and yet its functions still remain unclear. In one Iraya text, *ba* occurs 75 times in 116 sentences, sometimes as many as four times in a single sentence. Tweddell (1958:67) labels Iry. *ba* as a ‘copulative particle’. When asked what the equivalent form in Tagalog is, language assistants typically say Tag. *ay*, which is the form that marks an inverted construction in Tagalog (Schachter and Otnes 1972:485 et seq.), where some nominal form or adverb which occurs after the verb in unmarked constructions is fronted

and linked to the verb by the form *ay* (or =*y* following a vowel). An inverted construction in Tagalog primarily has the purpose of creating a formal structure, commonly used in writing and formal speeches, but less frequently in informal situations. While Iry. *ba* is commonly found (optionally) between an NP and a following verb (32), there is no indication that the construction is anything other than a normal construction, making it appear as a copula. However, it has a much wider function, as Tweddell (1958:67) noted, ‘it [*ba*] may also indicate syntactical juncture points’. These include the following: between a contrastive topicalized NP and a verb (33); between fronted locative and time phrases and a predicate (34)–(35); between a conditional clause, and its apodosis (36); between a quotation formula and a direct quote (37); between conjoined sentences (38); between a verb and its verbal complement clause (39); and between a noun and a following relative clause (40).

- (32) Iry. [*nay takər*] **ba** *tinapuñan=ani limatək*.
 GEN.1SG leg BA stuck=already leech
 ‘My legs were already stuck with leeches’.
- (33) Iry. [*?iya kay*] **ba**, *makita kun ?awak laki manhuñ kay*.
 TOP.3SG also BA painful GEN.3SG back male Manhung again
 ‘As for him, Manhung’s back was painful again’.
- (34) Iry. *da? [sa təbtəban kamutiyan]* **ba** *maki ?alug?ugan*.
 and LOC side camote.field BA exist stream
 ‘and beside the field there was a stream’.
- (35) Iry. *maraw [sa?i ka ?aldaw]* **ba** *batay mamahuy ?ag nay apu?*
 and.then one LIG day BA future visit SPCF GEN.1SG grandfather

sa kun kamutiyan.
 LOC GEN.3SG sweet.potato.field
 ‘Then one day my grandfather was going to visit his sweet potato field’.
- (36) Iry. [*nu binərya*] **ba** *batay pabali sa tamu*.
 if call BA future come.near LOC 1PL.IN
 ‘If you call it, it will come near to us’.
- (37) Iry. [*?amba kunu nay ?apu?*] **ba** “*?ayaw=ani kawu batay ?angət*
 said RPRT GEN.1SG grandfather BA do.not=now NOM.2SG future accompany

sa na?ay panawən ma?udan.
 LOC GEN.1SG because raining
 ‘My grandfather said, “Don’t come with me because it is raining.”’
- (38) Iry. [*yamən ba sata? magpamataw*] **ba** [*?iya ginhawa gidtə?*].
 1PL.EX BA there lived BA NOM.3SG comfortable really
 ‘We lived there and it was really comfortable’.

- (39) Iry. *maŋʔu ʔag nay kaka, [nay sinəɾəy]ba [maki ʔaŋani gulat*
 shout the GEN.1SG brother GEN.1SG notice BA exist surely injured

sa kunin suʔut ʔag nay kaka].

LOC GEN.3SG chest SPCF GEN.1SG brother

‘My brother shouted, and I noticed that my brother’s chest was badly injured’.

- (40) Iry. *nay sinəɾəyan [ʔag yamən mana balay] ba [rininas=ani]...*
 GEN.1SG saw SPCF 1PLEX also house BA destroyed=now

‘I saw our house that was destroyed’. or ‘I saw our house and it was already destroyed’.

Possible sources of Iry. *ba* will be considered in sec. 4.3.2 below.

2.3.2 Paradigmatic features of verbs

One of the most striking features of Iraya verbs is the use of originally perfective forms for future or potential activity, but preceded by an auxiliary verb, Iry. (*ba*)*tay* (noted also by Zorc 1974:574). Many Philippine languages generally reflect the PMP infix *<in> ‘perfective’, marking action that is completed, forming a past–non-past system in PMP. Central Philippine languages have combined the infix with a reflex of PMP *CVC- ‘imperfective’ reduplication to mark present, imperfective actions, forming a begun–non-begun system. Iraya retains PMP *<in> to mark completed actions, or past tense, but uses it also for future forms, forming a present–non-present system.

Elicited Iraya transitive (patient voice) constructions (41)–(43), with Tagalog equivalents for comparison (44)–(46), illustrate this paradigmatic shift (a morphological analysis of the verbs is given in line 2 of the examples). The same non-present form of the patient voice verb, Iry. *ʔininəm*, is used for both future and past events.

- (41) Iry. *nay ʔinəmən ʔag sapaʔ ɲuna.*
ʔinəm-ən

GEN.1SG drink-PV SPCF water now

‘I’m drinking the water now’.

- (42) Iry. *nay ʔininəm ʔag sapaʔ ʔaray ʔumaga.*
ʔ<in>inəm

GEN.1SG <NPRST>drink SPCF water today morning

‘I drank the water this morning’.

- (43) Iry. *nay batay ʔininəm ʔag sapaʔ girabas.*
ʔ<in>inəm

GEN.1SG FUT <NPRST>drink SPCF water tomorrow

‘I’ll drink the water tomorrow’.

- (44) Tag. *Iiinom ko ang tubig ngayon.*
C<in>V:-ʔinom

NPAST-drink GEN.1SG SPCF water now

‘I’m drinking the water now’.

- (45) Tag. *Ininom ko ang tubig kaninang umaga.*
ʔ<in>inom

<PAST>drink GEN.1SG SPCF water before morning

‘I drank the water this morning’.

- (46) Tag. *linomin ko ang tubig bukas.*
CV:-ʔinom-in
 NPAST-drink-PV GEN.1SG SPCF water tomorrow
 ‘I’ll drink the water tomorrow’.

The same system is found also with Iraya intransitive constructions. Exs (47)–(49) are extended intransitive (actor voice) construction with an indefinite oblique NP. The same non-present form of the actor voice verb, Iry. *minəm*, is used for both future as well as past events.

- (47) Iry. *ʔaku ʔagʔinəm sapaʔ ʔuna.*
ʔagʔinəm
 NOM.1SG AV-drink water now
 ‘I’m drinking water now’.

- (48) Iry. *ʔaku minəm sapaʔ ʔaray ʔumaga.*
m-inəm
 NOM.1SG NPRST-drink water before morning
 ‘I drank water this morning’.

- (49) Iry. *ʔaku batay minəm sapaʔ girabas.*
m-inəm
 NOM.1SG FUT NPRST-drink water tomorrow
 ‘I’ll drink water tomorrow’.

Text analysis shows that the Iraya infix <*in*> is no longer functioning as a perfective infix. Imperative forms of transitive verbs require the infix, as in (50). The same forms of the verb occur also with perfective forms (51).

- (50) Iry. *linəʔkəb ʔag balay, pad bintanaʔ daʔ pagsakbawan ba laʔkəb,*
 NPRST.close SPCF house PL window and door BA close

kuyu ba ʔayaw batay mamatpaʔ.
 2PL BA do.not FUT NPRST.watch
 ‘Close the house, the windows and doors and when closed, don’t look’.

- (51) Iry. *siniwan ʔag bintanaʔ daʔ mamatpaʔ.*
 NPRST.open SPCF window and NPRST.watch
 ‘She opened the window and watched’.

There are, however, multiple examples of verbs, that do not fit this pattern, and either reflect an earlier system, or conform to patterns of other languages, with which Iraya is in contact now or in the past. PMP (and earlier stages of Austronesian languages) is noted for verbal forms which are distinguished by what has been referred to as voice (Wouk & Ross 2002), by which the semantic role of the grammatical subject is referenced in the verb. Two major syntactic patterns are associated with the four or more voices. The different intransitive constructions are labelled actor voice, while transitive constructions are labeled undergoer or non-actor voice. Zorc (1974:578) notes a reduction of the three undergoer voices to one in the three northern languages of Mindoro, Iraya, Alangan, and

Tadyawan.⁵ All undergoer voice verbs in these three languages are affixed with a reflex of PMP *-en, although because of assimilation to the last vowel of the root, may appear as *-in*, *-ən*, *-an*, or *-un*. Zorc (ibid.) notes that, despite three examples in Tweddell’s Iraya grammar (1958:101) and the widespread use of such verbs in other languages of the Philippines, he was unable to elicit ‘portative’ verbs, such as *give*, *sell*, *throw*, *plant*, etc., with an *ʔi-* prefix in any of the three northern languages of Mindoro, Iraya, Alangan, and Tadyawan. In the text data I recorded, there is only one example of such a verb, and it is a frozen form, prefixed with an actor voice *ʔag-*.

- (52) Iry. *maki ʔiraya sataʔ ʔagʔilukub.*
 exist person there lie.down
 ‘There was a person lying down there’.

Another interesting fact about Iraya verb forms, is that reduplication is no longer productive. While some Iraya verbs retain a reflex of PMP *CVC- ‘imperfective action’, the forms are frozen. The form commonly used in Iraya to form imperfective or continuative action verbs, is a suffix *-an*, which can be doubled as *-anan* to form repetitive action verbs, as in the elicited data of two actor voice verbs *ʔinəm* ‘drink’ and *tinʔaraʔ* ‘look up’ in Table 3.

Table 3: Iraya verb forms

	Indicative	Continuative	Repetitive	Infinitive
Present	<i>magʔinəm</i> ‘drink’	<i>magʔinəman</i>	<i>magʔinəmanan</i>	<i>minəm</i>
Past	<i>minəm</i>	<i>minəman</i>	<i>minəmanan</i>	
Future	<i>batay minəm</i>	<i>batay minəman</i>	<i>batay minəmanan</i>	
Present	<i>magtinʔaraʔ</i> ‘look up’	<i>magtinʔaraʔan</i>	<i>magtinʔaraʔanan</i>	<i>tuminʔaraʔ</i>
Past	<i>tuminʔaraʔ</i>	<i>tuminʔaraʔan</i>	<i>tuminʔaraʔanan</i>	
Future	<i>batay tuminʔaraʔ</i>	<i>batay tuminʔaraʔan</i>	<i>batay tuminʔaraʔanan</i>	

3 Position of the Northern Mindoro languages

Zorc (1974) must be credited for first noting that the Northern Mindoro languages have a number of /y/ reflexes of PMP forms with *R, a development found in the Central Luzon languages as well as the Bashiic languages in the far north of the Philippines. Table 4 provides a short list of some of the forms provided by Zorc (1974) and Barbian (1977), which show such reflexes.

Table 4: Some Northern Mindoro forms showing /y/ reflex of PMP *R

	Gloss	PMP	Iraya	Alangan	Tadyawan
1	heavy	*beRqat	<i>ma-biyat</i>	<i>ma-biyat</i>	<i>ma-byat</i>
2	night	*Rabiqi	<i>yabiʔ</i>	<i>yabiʔ</i>	----
3	rib	*tageRaŋ	<i>tagyaŋ</i>	<i>tagyaŋ</i>	<i>tadyaŋ</i>
4	bite	*kaRat	<i>kayat</i>	<i>kayat</i>	<i>kayat</i>
5	earth	*daRaq	----	<i>diyaʔ</i>	<i>diyaʔ</i>
6	fire, embers	*baRaq	<i>bayaʔ</i>	<i>bayaʔ</i>	<i>baya</i>
9	hear	*deŋeR	<i>ka-riŋəy</i>	<i>ka-rəŋəy</i>	<i>-liŋy-an</i>
10	loincloth	*baqaR	<i>baʔay</i>	<i>baʔay</i>	<i>baʔay</i>
11	neck	*liqeR	<i>ləʔəy</i>	<i>ləʔuy</i>	----
12	new	*baqeRu	<i>bayu</i>	<i>bayuʔ</i>	----
13	satisfied	*besuR	<i>ʔa-bsuyan</i>	----	<i>ma-gsuy</i>
14	tail	*ʔikuR	<i>ʔikuy</i>	<i>ʔikuy</i>	<i>ʔikuy</i>
15	vein	*ʔuRat	<i>ʔəyat</i>	<i>ʔuyat</i>	<i>ʔiyat</i>
16	blood	*daRaq	<i>dayaʔ</i>	<i>dayaʔ</i>	----

⁵ The undergoer voices are labeled by Zorc (1974:577–578) direct passive; instrumental or associative passive, or ‘portative’; and local/referential passive, or ‘ablative’. In more recent publications, these three undergoer voices are labeled patient voice (with a reflex of PMP *-en /-ən/); locative voice (with a reflex of PMP *-an); and ‘instrumental’ or ‘conveyance voice’ (with a reflex of PMP *ʔi-).

There are several other possible innovations which link the Northern Mindoro languages with Central Luzon languages. Zorc (1974) cites two demonstrative innovations that are possibly shared between Iraya and the Central Luzon languages. He states:

The use of *ti as a base denoting nearness (rather than remoteness, as in Palawanic and SM [Southern Mindoro]) is, to my knowledge, only found in Pampango *iti* ‘this’, *keti* ‘here’, Sambal *bayti* ‘here’, and--if cognate--Iraya *tiya?* ‘this’. Likewise, the use of *ta as a base denoting remoteness is only found in Pampango *ita* ‘that’, *keta* ‘there’, and Iraya *nata?* ‘that’, *sata?* ‘there’. (Zorc 1974:589)

Zorc (1974) suggests certain pronominal innovations which are possibly shared between Iraya and the Central Luzon languages, e.g., *ʔi:ya* ‘3SG’ and *tamu* ‘1PL.IN’; he also notes a connection between Iraya and Bashiic languages with the pronoun *yamən* ‘1PL.EX’. Although these are interesting, they are probably retentions and cannot be considered evidence for subgrouping, since they occur in a number of other Philippine languages in different subgroups. The possible genetic relationship between the Northern Mangyan languages and the Central Luzon languages has also been carefully examined by Himes (2012:528-530). He concludes that the evidence for a Central Luzon–Northern Mangyan link is not overwhelming, but is probably sufficient to justify a closer relationship between these two microgroups than that enjoyed by either of them with other Philippine groups (Himes 2012:530).

But there is a pronominal innovation in Central Luzon languages not discussed by Zorc (1974), Barbian (1977), nor by Himes (2012) that appears to be shared by at least one of the Northern Mindoro languages. Some Central Luzon languages show an irregular development of PMP *kami ‘NOM.1PL.EX’, PMP *kamuyu ‘NOM.2PL’, and PCLuz *nāmen ‘GEN.1PL.EX’. These languages show the medial *m in these forms becoming a semivowel, either /y/ or /w/, depending on the vowel that follows, as in Table 5. Other Central Luzon languages, and alternate forms in the same languages maintain, or have restored the medial nasal, possibly as a result of the influence of languages such as Ilokano or Tagalog.

Table 5: Some Central Luzon reflexes of PMP pronouns

	Botolan Sambal	Kakilingan Sambal	Ayta Mag-anchi	Kapampangan
*kami ‘NOM.1PL.EX’	<i>kayi</i>	<i>kay</i>	<i>kay</i>	<i>ké</i>
*kamuyu ‘NOM.2PL’	<i>kawo</i>	<i>kaw</i>	<i>kaw</i>	<i>kó</i>
*nāmen ‘GEN.1PL.EX’	<i>na:wən</i>	---	---	---

This pronominal innovation is possibly shared with Tadyawan, which maintains a medial semivowel in *tawa* ‘2PL’, although the initial consonant and final vowel have changed, probably by analogy, as in the following possible sequence of events.

*kamu[yu] ‘NOM.2PL’ > **kawu (intervocalic *m > w)
 kawu > *kawa (analogy with -a final pronouns, e.g., *tama* ‘1PL.IN’, *ta* ‘1DL.IN’)
 ***kawa > *tawa* (analogy with t- initial pronouns, e.g., *tama* ‘1PL.IN’, *ta* ‘1DL.IN’, and to avoid homophony with *kawa* ‘2SG’ < PMP *kaʔu ‘2SG’)

Zorc (1974:592) likewise posits a number of lexical items which are shared between Kapampangan and Iraya, as shown in Table 6. Himes (2012:530) also suggests a number of other lexical and semantic innovations that are shared between Central Luzon languages and Northern Mangyan languages.

Table 6: Some shared lexical items between Kapampangan and Iraya

	Gloss	Kapampangan	Iraya
1	see	<i>akit, ikit</i>	<i>ʔakit-an</i>
2	wait	<i>panáy-an</i>	<i>panyaʔ-an</i>
3	cold	<i>ma-rimla</i>	<i>ma-dimla</i>
4	needle	<i>ka-rayum</i>	<i>ka-dayum</i>

4 Contact diffusion into Mindoro languages

This section presents evidence for long periods of contact diffusion into Mindoro languages, with particular reference to Iraya and the northern languages of the island. Sec. 4.1 repeats evidence from irregular sound change first reported by Zorc (1974) and discussed also by Barbian (1977), showing that even though the northern languages seem to subgroup with Central Luzon languages because of *R > /y/, the evidence is not strong. Sec. 4.2 presents evidence from verb morphology, that appears to link Iraya with Negrito languages of Luzon, and sec. 4.3 provides evidence from various syntactic features of Iraya and other Mindoro languages that suggest contact diffusion from various West Bisayan and Palawan languages.

4.1 Evidence from sound change

The main problem with most of the evidence given in sec. 3 that suggests a genetic connection between the Central Luzon and Northern Mindoro languages, is that various features are also shared with the Southern Mindoro languages, which supposedly group with other Greater Central Philippine languages. Zorc (1974:588) notes this, calling the evidence for the grouping of the Northern Mindoro languages with Central Luzon languages weak, but noting that there are more examples of a /y/ reflex of PMP *R in the northern languages than in the southern languages, and conversely there are more /g/ reflexes of PMP *R in the southern languages than the north. Examples of PMP *R > /g/ in northern Mindoro languages include forms translated as ‘coconut’, ‘lime’, ‘milled rice’, ‘northwest wind’, ‘molar’, ‘root’, etc. He similarly notes that most of the shared lexical items between Iraya and Kapampangan, are also found in various other languages in Mindoro, and there are some which are shared only with the southern languages.

Hanunó’o, the largest Mangyan language in the south shares many of its distinctive features with languages in Palawan, and there are forms in the northern Mangyan languages which suggest influence from Palawan. In Kalamianen, one of the languages in the north of Palawan, the reflex of PMP *R is /l/, so that the reflex of PMP *maRsi- ‘simultaneous/concomitant aspect’ (see Liao 2011) is /malsi-/. From this we can see that the Alangan form *malsiyatan* ‘hold one another’ probably originated from there, while the reflex of PMP *taR- ‘relationship between two people’ is /tal-/, with an apparent borrowing in Iraya *talyayaw* ‘husband and wife’ (see Blust 2003 for reconstructions of these prefixes). The expected /y/ reflex of PMP *R is found in Alangan *tayʔariʔan* and Tadyawan *tayʔaliʔan* ‘sibling’.)

What is very clear is that there has been considerable movement of people speaking a variety of languages into Mindoro. This is evident from the verbal systems of Mangyan languages. In sec. 4.2 below, data suggesting contact diffusion into the southern Mangyan languages, probably from a Palawan language is first presented, and then data suggesting contact diffusion into Iraya possibly from Northeast Luzon Negrito languages is discussed.

4.2 Evidence from verb morphology

Zorc (1974:591) discusses the relationship between the southern languages, Hanunó’o and Buhid, and the Palawan languages, noting that their verbal systems share an important innovation. He notes that in non-actor voice constructions (Zorc’s “passives”), progressive verbs are affixed with *pag--ən, a form, he claims, which does not occur in any other “Meso-Philippine” language. He suggests the source of the innovation, as follows.

[This] is apparently based on the analogy whereby the *pag- abstract prefix (used in temporal constructions) is generalized to a durative function, filling in the paradigm: *mag-* (future), *nag-* (past), *pag-* (progressive). Northern Tagbanwa has precisely this system in the active. This *pag-* (progressive) then was used independently in the passives of these Pal and SM languages, alongside the *-en (future). (Zorc 1974:591)

The linguistic connection between Hanunó’o and Buhid, and the Palawanic languages, is also clear from Zorc’s lexicostatistical analyses, which show consistently higher scores with languages in Palawan than with any of the northern Mangyan languages (Zorc 1974:585).

Iraya, although no longer considering themselves to be a Negrito group, has been shown to have been such a group before intermarriage with non-Negrito people (HUGO 2009, Fig. 1). It has been noted before that Negrito peoples of the Philippines not only maintained intermittent connections with non-Negrito peoples from whom they learned their Malayo-Polynesian languages, but also maintained on-going connections with other Negrito groups with whom they may have inter-married (Reid 1994b). It is therefore significant that the unusual verbal features of Iraya, in which apparently completive aspect (past tense) forms are also used for potential aspect (future tense), appear to be shared with other Negrito groups, such as Alabat Agta (AGT.AL), Manide (MND), Rinconada Agta (AGT.RN), and Umiray Dumaget (DGT.UM), although the overall verbal system of these languages is quite different from Iraya (see Lobel 2010 for a discussion of the morphology of these languages). Each of these Negrito languages has future forms which appear to have developed from a perfective form, with an infix *<in>, either PMP *m<in>aR- in actor voice verbs, or PMP *p<in>aR- in non-actor voice verbs, as in Table 6. In Alabat Agta, *nag-* shows the loss of the first two phonemes of the reconstructed actor voice, perfective aspect form, a common development in many Philippine languages, including Ilokano, Tagalog, Bikol and Cebuano. In the non-actor voice, the presence of an *i-* vowel in Alabat Agta and Manide signals the earlier presence of the infix <in>. Rinconada Agta with non-perfective (future tense) actor-voice *mig-* has the same form that is used in Kapampangan for some perfective actor voice verbs, as in (53).

Lobel (2010:496) in discussing these forms notes that the origin of the prefix *pig-* is unknown, but also occurs as a past and present prefix in a number of Bikol languages and dialects. He suggests the possibility that *mig-* and *pig-* are from earlier *magi- and *pigi- with hypothetical vowel metathesis or right-to-left raising, an explanation which ignores the fact that Kapampangan also uses *mig-*, and that perfective forms have apparently developed as future forms in these languages, as also in Iraya.

- (53) Kpm. *migsalúd ka palá king girípu.*
 bathed NOM.2SG surely LOC faucet
 ‘So, you took a bath at the faucet’. (Mirikitani 1972:103)

Table 6: Some future tense forms in Negrito languages (from Lobel 2010:496)

	Actor voice	Non-actor voice
AGT.AL	<i>nag-</i>	<i>pig-</i>
MND	<i>nig-</i>	<i>ig-/pig-</i>
AGT.RN	<i>mig</i>	---
DGT.UM	<i>nV-</i>	---

4.3 Evidence from syntax

4.3.1 Nominal specifiers

Another feature that connects Iraya and Negrito languages, such as Manide, Alabat Agta and Umiray Dumaget, is that all of these languages, apparently unique among Philippine languages, use the same nominal specifier (or ‘determiner’) for common and personal nouns, as discussed above in sec. 2.2.2, a fact that Lobel (2010) was apparently unaware of. He states, “... in fact, Umiray Dumaget is the

only other Philippine language known to use the same set of case markers for common nouns and personal names, yet the Umiray Dumaget forms are largely different from the Manide and Inagta.” (Lobel 2010:498).

While Manide, Alabat Agta and Umiray Dumaget maintain nominal specifiers that indicate the case of the noun phrase,⁶ and Iraya no longer uses nominal specifiers to mark case, it is striking that one of the Iraya specifiers, *da* (discussed above in sec. 2.2.4), is cognate with the oblique (locative) forms of the Luzon Negrito languages. These are cited by Lobel (2010:498), and are cognate with Iry. *da* given that the forms cited by Lobel all show regular low-vowel raising following a voiced obstruent, so that **da* > MND, DGT.UM *di*, AGT.AL *de*. This supports Liao (2015, 2016), that claims that Proto-Northern Luzon locative personal pronouns (which includes the Negrito languages of Northeast Luzon) were all marked by an initial **da-*, which was a locative nominal specifier in Proto-Northern Luzon that was reflected as *di* or *de* in some Northeastern Luzon Negrito languages.

Only Iraya among the Mindoro languages uses *da* as a nominal specifier, probably as a result of contact with Negrito languages of Northeast Luzon. However, the other nominal specifier in Iraya, *ag*, is evidence of contact with Central Philippine languages. As noted above, North Mindoro languages use /*pag-*/ as a nominal specifier in noun phrases (Iraya has reduced it to /*ʔag*/). Alangan marks agentive noun phrases with a reflex of the old agentive personal noun marker **ni* plus /*pag*/, as Aln. *ni pag N* (Iraya has lost a reflex of **ni*). This is an innovation probably based on the nominalizing function of /*pag-*/ in Central Philippine languages, such as Akl. /*pag-káʔon*/ ‘food’ (from /*káʔon*/ ‘eat’), /*pag-ʔabót*/ ‘arrival’ (from /*ʔabót*/ ‘arrive’). In addition, the use of *pag-* to introduce a temporal clause, as in (29) is clearly a borrowing of a common dependent clause morpheme in Central Philippine languages, such as Aklanon temporal verbs, e.g., Akl /*pag-ʔabót nána*/ ‘when he arrived’. Zorc (1974:591) suggests that this is also the source of the Mindoro languages that have replaced an inherited ligature between a head and its modifier with *pag-*, as shown above in (17).

4.3.2 Other syntactic features

In sec. 2.3.1 above, the ubiquitous Iraya morpheme *ba* was introduced. The source of this morpheme is still unclear. Blust and Trussel (Ongoing) reconstruct six different **ba* forms for PMP, noting that at least one (PMP **ba*₁ ‘conjunction: or, if, perhaps, because’) has a range of functions, “many of them introduc[ing] an element of doubt, qualification or negation.” Blust and Trussel give the source of the Tagalog interrogative marker *ba*, with its dialectal variant *baga*, as PMP **ba*₅ ‘post verbal interrogative particle’. However, its optional use in Tagalog sentences that are already marked as questions, either by intonation, or by the presence of interrogative words such as Tag. *saan* ‘where’, *sino* ‘who’, and *ilan* ‘how many’ (Schachter and Otanes 1972:424), implies that it is not an interrogative marker as such but a reflex of **baga*, a doublet of Blust’s PMP **bajaq* ‘tell, inform; ask, inquire’, as in (54)–(55).

(54) Tag. *Aalis ka na (ba)?*⁷
 leaving NOM.2SG now ASK
 ‘Are you leaving now?’ (Schachter and Otanes 1972:424)

(55) Tag. *Saan ka (ba) nakatira?*
 where NOM.2SG ASK live
 ‘Where do you live?’ (Schachter and Otanes 1972:424)

It is clear from the examples given in sec. 2.3.1 above, that Iraya does not use *ba* with any of the senses given for Blust and Trussel’s reconstructions of PMP **ba*, and is never an interrogative marker but rather, if indeed it is a reflex of a PMP form, is a shortened form of either PMP **bajaq* or **baga*

⁶ MND, AGT.AL *hu*, DGT.UM *i* ‘nominative noun’; MND *nu*, AGT.AL *nu*, DGT.UM *ni* ‘genitive noun’; MND *di*, AGT.AL *de*, DGT.UM *di* ‘oblique noun’.

⁷ Interlinear translations are not provided by Schachter and Otanes. They are provided here by me.

‘tell, inform’ with the sense bleached until it has become simply a pause marker at syntactic boundaries, including a conjunction, which links Blust and Trussel’s PMP *ba₁ and *ba₅ that must ultimately also be developments of PMP *bajaq ‘tell, inform; ask, inquire’.

- (56) Iry. [yamən ba sata? magpamataw] **ba** [ʔiya ginhawa gidtəʔ].
 1PL.EX BA there lived BA NOM.3SG comfortable really
 ‘We lived there and it was really comfortable’.

Iraya has also been affected by Tagalog in-migrants into Mindoro, with multiple Tagalog lexical items now commonly being used, and younger generations only speaking the language. One Tagalog form that appears as a common syntactic feature in Iraya is Tag. *batay* ‘based on or upon’ (Komisyon sa Wikang Filipino 2000). In Iraya, this form is an auxiliary verb which precedes future tense verbs, as in Table 2, and is part of the aspectual system of the language.

4.3.3 Possible influence of non-Philippine languages

Other speculations about possible sources of various features of the Mindoro languages exist. Iraya oral literature is full of stories about interaction with Chinese visitors and Muslim raiders. With reference to possible Chinese influence. Limahong, also known as Lim Ah Hong, or Lin Tao Kien was a Chinese pirate who invaded the northern islands of the Philippines and tried to seize the City of Manila from the Spanish occupiers in 1574. One of his safe harbors was Batangas with its deep waters. Lim Ah Hong is reported to have taken a Filipina wife but as a Chinese Warlord was allowed as many concubines as he could afford, thus, it is said, “he populated the province of Batangas with his wife and countless concubines; who gave forth progeny, of whom we are the direct descendants.”⁸ The fact that Batangas is within a short sailing distance of northern Mindoro, the area where Iraya is located, and the fact that local stories tell of the visits of Chinese ‘businessmen’ who required Iraya leaders to change their names to Manhong and Masahod (the first possibly in local imitation of Limahong), suggests that the Chinese had at least considerable social influence (Banaag 2014). Did the Chinese language that must have been spoken by some of the ‘businessmen’ influence Iraya word order, in which pronouns always precede their head nouns? And did the Mandarin Chinese /ba-/ (tone 3) construction which occurred between a subject and a fronted object (Sun 2008), re-inforce the use of *ba* in Iraya? These possible influences are speculative, but given Chinese social influence and probable intermarriage with Iraya women, they cannot be ignored.

Another possible external source is Muslim slave raiders, who over several hundred years devastated local communities (Warren 2007). Iraya oral literature is replete with such events that affected local people. The stories suggest that Muslim communities existed in Mindoro and were growing rice. One of the local heroes was killed by a Muslim wielding his rice-pounding pestle (Banaag 2014). At least two so-called ‘pirate’ Muslim communities existed in Mindoro, one at Pinamalayan on the east coast, and one at Mamburao on the west coast, from whence they raided to Luzon and other islands (Gardner und.). The question is what language were the raiders using, and did it affect in any way the Mangyan languages? Possibly hundreds of people were taken as slaves from the various language communities in Mindoro and transported to areas south, such as Sulu, Borneo and other areas. Did some of them ever return after being emancipated after having learned the language where they were taken?

One of the unique features of Iraya is the use of *laki* as a marker of male personal names, and *ba’i* as a marker of female personal names as described in sec. 2.2.2. To my knowledge, there are no MP languages that currently use such forms, but several South-East Sulawesi languages, including Buton and Bugis use *La-* and *Wa-* as prefixes for men and women respectively, e.g., *I La Galigo* is a character in a Bugis story cycle of the same name, and internet sources tell of folk heroes in South Sulawesi languages whose names carry such gender identifying forms, e.g., *Lakipadada*,⁹ a supposed

⁸ http://www.watawat.net/limahong_the_pirate_and_his_kingdom.html, accessed Aug. 31, 2016.

⁹ <http://dorogoblog.blogspot.jp/2008/06/adventures-of-lakipadada-as-told-by.html>, accessed Aug. 31, 2016.

ancestor for the major South Sulawesi kingdoms and *Lakilaponto* of Muna fame,¹⁰ whose wives and female children carry a *Wa-* prefix (Anthony Jukes pers. comm.). Although these prefixes are said to be of Arabic origin,¹¹ they are more likely to be shortened forms of *laki* and *ba'i*, with the latter undergoing a *b > /w/ shift, found in some forms in Javanese and Malay, as well as in Maranao of the Philippines (Blust 2009 [2013]:680). Significantly, several of the Muslim South Sulawesi groups, including Buton and Bugis are known to have gone on far-flung slaving raids, and are possibly the source of the prefixes which precede Iraya names today.

5 The retrenchment of Iraya

The foregoing sections have outlined the unusual features of Iraya and some of the other languages of Mindoro. This section is a speculative account of supposed events, long before the arrival of speakers of MP languages in the Philippines and subsequently, which attempts to provide an explanation for the current situation among the Mangyan peoples of Mindoro.

The presence of multiple bands of Negritos throughout the Philippines is well-documented (Reid 1994a, 2013b). Many of these peoples still retain their identity as Negritos, distinct from the MP populations that surround them, while others have lost their identity and consider themselves to be part of one of the MP groups with whom they have intermarried. There is only one remaining group in Mindanao, for example, that still considers themselves to be a Negrito group, that is the Mamanwa of north-east Mindanao. But the physical features of many Manobo groups in Mindanao suggest that they were also Negritos in the past, but have lost their identity. This is true not only of the Ata Manobo, who retain a form of the name that many other Negrito groups use, but also of other Manobo groups as well, as seen in HUGO (2009, Fig. 1), where a sample of 10 Manobo show clear Negrito alleles. The current distribution of Negritos in the Philippines shows bands of Negritos spread from the north of the Philippines down through the Sierra Madre along the east coast of Luzon, and into mountainous areas in the south of Luzon. In the west of Luzon there are multiple bands of Negritos in the Zambales Mountains. South of Luzon the islands of Negros and Panay have a number of bands, see Reid (2013b, Fig. 1). It is assumed that prior to the spread of MP people in the Philippines, Negrito bands occupied river valleys and lowland areas where food was plentiful, and their present locations in mountainous areas is the result of their being forced to move from their favored locations by the activities of the incoming MP people, who deforested the areas and farmed them. It is assumed that prior to the incoming MP population, Mindoro was just like Negros and Panay, widely occupied by Negrito bands who exploited the river valleys and coastal areas of the island. Palawan was probably also the home of numerous Negrito bands, of whom only the Batak still identify as Negrito.

We do not know what languages Negritos were using prior to the arrival of MP people. Given the extreme length of time that Negritos were present in the Philippines, possibly more than 50,000 years, we must assume that they were speaking a wide range of mutually unintelligible languages, although evidence suggests that Negrito groups interacted with one another, and may have exchanged wives. Negrito groups across wide areas of northern Luzon and associated with different MP subgroups share some lexical items not found to date in MP languages (Reid 1994b). What is clear, is that groups of Negritos interacted with the MP people, eventually giving up their languages for the one that was spoken by the MP group that was in their vicinity.

While at present Tagalog people occupy the area of Batangas and the provinces north, this area was originally settled by the ancestors of Kapampangan peoples who were forced to retreat into Central Luzon as Tagalog people moved north from the Bisayan area (Zorc 1993, Reid 2013b:347). So it is not surprising that Iraya people who live across the channel from Batangas show a sound change and other features that link them with Central Luzon languages. One must assume also that it was not simply the ancestors of the group known today as Iraya that was in contact with the residents of the Batangas area, but that the newly acquired language spread across the island among other groups of Negritos, that have long since been replaced or intermarried with other ancestral MP groups

¹⁰ <https://id.wikipedia.org/wiki/Lakilaponto>, accessed Aug. 31, 2016.

¹¹ <http://www.timur-angin.com/2010/06/punahnya-nama-la-dan-wa-di-masyarakat.html>, accessed Aug. 31, 2016.

who moved into the island. Given this scenario, the features of Central Luzon languages which are spread today across the island, such as PMP *R > y, are remnants, or substratal effects of the languages originally spoken in these areas.

All the southern languages of Mindoro are now considered to be part of Blust's (1991) Greater Central Philippine languages, along with Central Philippine languages (such as Tagalog, Bikol and Cebuano), Palawanic, Danao, Manobo, Subanen, and the Gorontalo-Mongondoic languages of north Sulawesi, all of which show the sound change PMP *R > g. But the probable movement of peoples from Palawan north into Mindoro is suggested, because of the linguistic features that the southern Mindoro languages share with languages in Palawan, as outlined above in sec. 4.2. It is also clear from shared cultural features between southern Mindoro languages and Palawan. It is well known that Hanunóo and the language to its immediate north, Buhid, have adopted a form of the traditional Indic script and use it today to write traditional poetry. Since this writing system was not found in any of the West Bisayan languages but was common in Palawanic languages, we must assume that it was brought into southern Mindoro by in-migrants from Palawan.

Just as Negrito groups in Luzon were forced from their preferred habitats in coastal areas and river valleys into less hospitable mountainous areas, it is assumed the same happened in Mindoro, with incoming peoples from Palawan and West Bisayan languages taking over areas originally occupied by Negrito peoples, who were gradually forced into mountainous areas. Has intermarriage between incoming MP people and Negrito people gradually erased the Negrito phenotype in the speakers of each of the southern Mindoro languages, or were the Negrito peoples gradually forced to move north, until only the Iraya are left, entrenched in the most northerly mountains of Mindoro? Even among the three northern languages, which supposedly share sound correspondences and other features with Central Luzon languages, it is only speakers of Iraya who still maintain some of the Negrito phenotypical features, while most speakers of the other two languages with which it apparently groups, Alangan and Tadyawan, no longer have Negrito features.

This scenario is primarily based on what is known about the distribution of Negrito groups and the fact that features of Central Luzon languages are found not only in the northern group of languages, but also in the southern group. It is also based on what seems to be a movement of peoples from Palawan into Mindoro in the far past, and in more recent times from some of the West Bisayan languages, with which Datagnon in the south of Mindoro is closely related.

The scenario outlined above is supported by genetic studies, not only the fact that Iraya carry Negrito genes (HUGO 2009), but also by Delfin et al. (2011). The latter paper, while lacking a balanced set of Philippine ethnolinguistic samples, does have samples from several of the Mangyan languages of Mindoro, specifically Iraya and Tadyawan of the northern group, and Hanunóo and Tawbuid (Buhid) of the southern group. The paper discusses two old Y-chromosome haplogroups which Negrito groups share. These are K-M9, which all Negrito groups that were sampled carry, and C-RPS4Y that is also carried by Agta (not specified), Ati and Mamanwa. Delfin et al. (2011:227) claim that haplogroup K-M9 is distributed among nine ethnolinguistic groups (including all Negrito groups sampled and three non-Negrito groups), and for the most part tend to involve groups that are geographically close, including Mamanwa and adjacent groups Surigaonon and Manobo (non-specific as to which of the many Manobo groups, but could include Ata Manobo whose name signals the possibility of Negrito origin), suggesting cross-group intermarriage. Of particular interest in this paper is that another clear grouping is between the Aeta (Ayta) of Zambal and Bataan (not specific which of the five Negrito groups were sampled), all of which are Central Luzon languages, and Hanunuo (Hanunó'o). These languages share Y-SNP frequency groupings of K-M9 as follows: Aeta Zambal 1.00, Aeta Bataan 0.87, and Hanunuo 0.67. Iraya has a K-M9 frequency of 0.25 (Delfin et al. 2011:226 Table 1). Tadyawan seems to be a different story in that samples from this group carried no specifically Negrito haplogroup, but had the haplogroup O-M110 at a frequency of 1.00. This is a widespread haplogroup among Philippine ethnolinguistic groups and "has a clear Taiwan-specific origin" (Delfin et al. 2011:229). This suggests that Tadyawan represents a non-Negrito group that moved into Mindoro and learned the local language without intermarriage with Negrito groups.

6 Conclusion

Zorc (1974:594) in his excellent analysis of the relationships of Mangyan languages, makes the following comment, “While some of the evidence discussed herein is suggestive, none of it is ineluctable. Most of the features discussed are spread in one way or another throughout the Palawan-Mindoro-Pampango area.” These are the facts that stimulated the present enquiry, and form the basis of the scenario outlined in this article.

This article primarily focuses upon Iraya, one of the three North Mindoro languages, a language spoken by people who identify themselves as one of the mountain peoples of the country, generally referred to as Mangyan, but whose phenotypic features are somewhat unique among Filipino people. Modern genetic testing of Iraya shows that they share a significant proportion of ancestral alleles with Negrito peoples of the Philippines. Their language as described above shows a number of unusual features, such as a pronominal system which is unique among Philippine languages, with a complete set of dual pronouns, including 1st person inclusive and exclusive dual pronouns. The pronominal system also retains case-marking distinctions only in the three singular forms, with all non-singular forms having only a single form, which appears to be originally based on an oblique or locative form. The pronominal system is also different from most other Philippine languages in that the singular genitive forms no longer reflect PMP enclitic forms. All pronouns precede their head nouns or verbs. Iraya is also distinct from other Philippine languages in that it uses historically perfective forms for future, creating a present–non-present tense, a feature reminiscent of similar morphology in some Negrito languages of Luzon. The language is also different from other Philippine languages in having lost case-marking on nominal specifiers, although at least one of the two forms currently used to introduce noun phrases appears to be a reflex of an old locative marker *da, found also among some Negrito languages of Luzon. Iraya is unique among Philippine languages in having names preceded by /laki/ ‘male’ or /ba?i/ ‘female’.

As examination of the features that distinguish Iraya and its related languages suggests multiple sources. Some are shared with Central Luzon languages. Others are probably the result of language contact. Some are not shared by any other Philippine language and are innovations. But as Zorc (*ibid*) noted, most of the features are shared throughout the Mindoro languages, and even into Palawan. Various historic and prehistoric events are responsible for this. Two historic events are considered as possibly resulting in language change. The possible influence of Chinese contacts, and the possible result of slave-raiding, with communities of individuals involved in slave-raiding establishing at least temporary communities in Mindoro.

The major influence however was prehistoric. Prior to the spread of MP people through the Philippines, it is assumed that Mindoro, like other parts of the Philippines was occupied by bands of Negritos deriving their livelihood from the ocean and rivers, and by exploiting easily accessible forest foods. Following the spread of MP through the Philippines, things changed. Contact with MP people was first through the ancestors of Central Luzon languages which at that time occupied Batangas and other areas of southern Luzon, prior to the move of the ancestors of Tagalog north from their homeland areas of northern Mindanao and Marinduque. Subsequently, as MP people expanded and needed new lands for farming, an influx of peoples from the western areas of the Bisayas and from Palawan moved into the country either intermarrying with Negritos and/or gradually forcing Negrito bands from the areas where they lived and into the mountains.

Over thousands of years, language contact has resulted in ancestral Central Luzon features being gradually lost and the languages in the south of the country becoming more like Palawan languages, and those in the north retaining more of their original features. This has matched the phenotypical features of Mangyan people, with those in the south more closely matching MP people, while only Iraya in the far north of the island retaining physical features that resemble those of Negritos.

Abbreviations

Abr	Aborlan Tagbanwa	Kpm	Kapampangan
Agt	Agtá	Kuy	Kuyunon
Agy	Agutaynen	LIG	Ligature
Al	Alabat	LOC	Locative
Aln	Alangan	Mnd	Manide
AV	Actor voice	NOM	Nominative
Bik	Bikol	NPAST	Non-past
Bon	Bontok	NPRST	Non-present
Ceb	Cebuano	OBL	Oblique
DET	Determiner	Pal	Palawan
Dgt	Dumaget	PL	Plural
EX	Exclusive	PMP	Proto-Malayo-Polynesian
FUT	Future	PRD	Predicate
GEN	Genitive	PV	Patient voice
Hil	Hiligaynon	Rn	Rinconada
Ilk	Ilokano	SG	Singular
Iry	Iraya	SPCF	Specifier
Kal	Kalamian	Tag	Tagalog
Kin	Kinaray-a	Um	Umiray

References

- Banaag, Leo. 2014. Ang kasaysayan ng tribong Mangyan Iraya [The history of the Mangyan Iraya tribe]. Unpublished document, with English translation by Lawrence A. Reid, 5pp.
- Barbian, Karl-Joseph. 1977. The Mangyan languages of Mindoro: A comparative study of vocabulary, phonology, and morphology. M.A. thesis, University of San Carlos, Philippines.
- Beyer, Otley H. 1921. The non-Christian people of the Philippines. In *Census of the Philippine Islands (Manila 1918) vol. 2 (separate from the Census)*, 907–957. Manila: Bureau of Printing.
- Blust, Robert. 1991. The Greater Central Philippine hypothesis. *Oceanic Linguistics* 30(2):73–129.
- Blust, Robert. 2003. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42:432–478.
- Blust, Robert. 2009 [2013]. *Austronesian languages*. Canberra: Pacific Linguistics. [2013 revised edition on-line at: <http://pacling.anu.edu.au/materials/Blust2013Austronesian.pdf>]
- Blust Robert & Stephen Trussel. Ongoing. Austronesian comparative dictionary, web edition www.trussel2.com/ACD.
- Delfin, Frederick, Jazelyn M. Salvador, Gayvelline C. Calacal, Henry B. Perdigon, Kristina A. Tabbada, Lilian P. Villamor, Saturnina C. Halos, Ellen Gunnarsdoóttir, Sean Myles, David A. Hughes, Shuhua Xu, Li Jin, Oscar Lao, Manfred Kayser, Matthew E. Hurles, Mark Stoneking & Maria Corazon A. De Ungria. 2011. The Y-chromosome landscape of the Philippines: Extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups. *European Journal of Human Genetics* 19:224–230.
- Dimaano, Gina M. 2005. A brief grammatical sketch of the Alangan language. LIN 626P term paper, De La Salle University, Manila, the Philippines.
- Gardner, Fletcher. und. Lingualization, an unusual mutation in Mangyan. Unpublished paper presented to the Faculty of Literature and Politics, Taihoku Imperial University, Taihoku, Formosa.

- Hammarström, Harald & Robert Forkel, & Martin Haspelmath, & Sebastian Bank. 2016. Glottolog 2.7. Jena: Max Planck Institute for the Science of Human History. (Available online at <http://glottolog.org>, Accessed on 2016-08-06.)
- Himes, Ronald. 2012. The Central Luzon group of languages. *Oceanic Linguistics* 51(2):490-537.
- HUGO Pan-Asian SNP Consortium. 2009. Mapping human diversity in Asia. *Science* 326, 1541. DOI: 10.1126/science.1177074.
- Komision sa Wikang Filipino. 2000. *Diksiyunaryo Filipino-English*. 2nd edition. Manila: Komisyon sa Wikang Filipino.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2016. *Ethnologue: Languages of the World, Nineteenth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Liao, Hsiu-chuan. 2011. On the reconstruction of Proto-Malayo-Polynesian *si- 'simultaneous/concomitant aspect', paper read at the 20th International Conference on Historical Linguistics (ICHL2011), National Museum of Ethnology (MINPAKU), Osaka, Japan, July 25-30, 2011.
- Liao, Hsiu-chuan. 2015. To raise or to front?: puzzles in the development of vowels in Ilongot, paper read at the 13th International Conference on Austronesian Linguistics (13-ICAL), Academia Sinica, Taiwan, July 18-23, 2015.
- Liao, Hsiu-chuan. 2016. Reconstructing Proto-Northern Luzon locative personal pronouns, paper read at the 26th Annual Meeting of the Southeast Asian Linguistics Society (SEALS 26), Century Park Hotel, Manila, the Philippines, May 26-28, 2016.
- Lobel, Jason W. 2010. Manide: An undescribed Philippine language. *Oceanic Linguistics* 49(2):478–510.
- Lobel, Jason W. 2013. Philippine and North Borneo languages: Issues in description, subgrouping, and reconstruction. Ph.D. Dissertation, University of Hawai'i.
- Mirikitani, Leatrice T. 1972. *Kapampangan syntax*. Oceanic linguistics Special Publication 10. Honolulu: University Press of Hawai'i.
- Reid, Lawrence A. 1994a. Unraveling the linguistic histories of Philippine Negritos. In *Language contact and change in the Austronesian world*, ed. by T. E. Dutton and D. T. Tryon, 443–475. Berlin: Mouton de Gruyter.
- Reid, Lawrence A. 1994b. Possible non-Austronesian lexical elements in Philippine Negrito languages. *Oceanic Linguistics* 33(1):37-72.
- Reid, Lawrence A. 2002. Determiners, nouns, or what? Problems in the analysis of some commonly occurring forms in Philippine languages. *Oceanic Linguistics* 41(2):295-309.
- Reid, Lawrence A. 2009. The reconstruction of a dual pronoun to Proto Malayo-Polynesian. In *Discovering History through Language: Papers in Honour of Malcolm Ross* (=Pacific Linguistics 605), ed. by Bethwyn Evans, 461–477. Canberra: Pacific Linguistics.
- Reid, Lawrence A. 2013a. Iraya field notes.
- Reid, Lawrence A. 2013b. Who are the Philippine Negritos? Evidence from language. In *Human Biology* 85 Issue 1, chapter 15 (on-line at <http://digitalcommons.wayne.edu/humbiol/vol85/iss1/>).
- Reid, Lawrence A. 2016. Accounting for variability in Malayo-Polynesian pronouns: Paradigmatic instability or drift? *Journal of Historical Linguistics* 6(2):130-164.
- Schachter, Paul & Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley, Los Angeles, and London: University of California Press.
- Sun, C. 2008. The theory of subjectification and the study on the Contemporary Chinese *ba*-construction. In *Contemporary linguistic theories and related studies on Chinese*, ed. by Y. Shen & S. Feng, 375–393. Beijing: Commercial Press.

- Tweddell, Colin E. 1958. The Iraya (Mangyan) language of Mindoro, Philippines: Phonology and morphology. Ph.D. dissertation, University of Washington, Seattle.
- Warren, James F. 2007. The Port of Jolo and the Sulu Zone slave trade: An 1845 report. *The Journal of Sophia Asian Studies* 25:303–322.
- Wouk, Fay & Malcolm Ross (eds). 2002. *The history and typology of Western Austronesian Voice Systems*. Canberra: Pacific Linguistics, Australian National University.
- Zorc, R. David. 1974. Internal and external relationships of the Mangyan languages. (Papers of the First International Conference on Comparative Austronesian Linguistics, vol. 1). *Oceanic Linguistics* 13 (1/2):561–600.
- Zorc, R. David. 1993. The prehistory and origin of the Tagalog people. In *Language—a doorway between human cultures: Tributes to Dr. Otto Chr. Dahl on his ninetieth birthday*, ed. by Ø. Dahl, 201–211. Oslo: Novus.

RECONSTRUCTING PROTO KENYAH PRONOUNS AND THE DEVELOPMENT OF A TRUE FIVE NUMBER SYSTEM

Alexander D. Smith
University of Hawai'i at Mānoa
smithad@hawaii.edu

Abstract

Typologists have in the past claimed that a 15 person system, with singular, dual, trial/paucal, plural, first person inclusive/exclusive, second person, and third person, is the maximum system allowed for systems of personal pronouns in the world's languages. These claims have been interpreted formally in feature systems that restrict the possible number distinctions in pronoun systems to four (singular, dual, trial or paucal, and plural). Within the past fifteen years more comprehensive works on number have published documented five number systems in widely available formats. Interestingly, all of the languages used to exemplify five number systems are from the Oceanic subgroup of Austronesian. However, complex number systems, including a five number pronoun system, have developed independently in central Borneo. Although Blust has made this data widely available, central Borneo remains untouched in general discussions of number. The Kenyah languages form a primary branch of the North Sarawak subgroup, located in central Borneo in both Sarawak (Malaysia) and Indonesia. A number of typologically rare features are present in Kenyah, including a pronoun system that includes 5 distinct numbers. Comparative data shows that a five number system can be reconstructed for Proto-Kenyah. This challenges both typological and formal descriptions of number.

Keywords: Kenyah, Pronouns, Feature Geometry
ISO 639-3 codes: xkl

1 Introduction¹

Number in pronoun systems is described formally as a product of the interaction of features.² At most, the formal literature allows for three features which interact to give a maximum of four numbers in the world's languages: singular, dual, trial/paucal, and plural (Harley and Ritter 2002). More restrictive feature based accounts of number allow for only two features, [singular, plural] (Adger 2003), or a single feature, [\pm plural] (Anderson 1992). The goal of such restricted feature sets is to constrain theory to represent only those languages which have attested attributes. From a typological perspective, it has been claimed that a 15 person system, with singular, dual, trial/paucal, plural, first person inclusive/exclusive, second person, and third person, is the maximum system allowed in the world's languages (Ingram 1978). Within the past fifteen years more comprehensive works on number such as Corbett (2000) and Cysouw (2003) have published documented five number systems in widely available formats, yet formal systems of features still insist on a four number maximum. Interestingly, all of the languages used to exemplify five number systems so far are from the Oceanic subgroup of Austronesian. However, complex number systems, including a five number pronoun system, have developed independently in central Borneo. Although Blust (2003, 2013 [2009]) has made this data widely available, central Borneo remains untouched in general discussions of number.

¹ Special thanks to Michael Laing, of Long San, for explaining in detail the difference between paucal and plural in his language. Many thanks to Lawrence Reid and Robert Blust who made comments on an earlier draft of this paper, greatly improving my draft. Any errors that remain are my own.

² In this paper, I use the term NUMBER to refer only to pronominal number. It is not meant to refer to grammatical number or agreement. None of the languages in this paper indicate number of objects (as in *book/books*).

In light of this, the aim of this paper is to review the theoretical predictions and constraints on number systems in the world's languages, while bringing special attention to the presence of a complex number system outside the Oceanic subgroup. To achieve this, I will describe in detail the pronominal system of Kenyah and its numerous dialects, and use the analysis to reconstruct a five number system to the level of Proto-Kenyah (PKEN). The reconstructed pronominal system makes at least 19 distinctions, including numeral distinctions of singular, dual, trial, quadral, and plural, as well as an inclusive/exclusive distinction. This reconstructed system challenges many of the more restrictive theoretical claims of number.

2 Feature representations of number

Theoretical accounts of number systems generally operate under two principles: (1) number systems are best represented as an interaction of features. Typically, these features incorporate [singular] and [plural],³ and in some cases additional features are added to allow for a greater number of contrasts; (2) theoretical accounts of number systems constrain number to an absolute maximum. That is, there is a limit to the number of distinctions that can be made in number systems, and this distinction is a result of the interaction of features. Different theoretical accounts of number constrain the possible distinctions to varying degrees. The remainder of this section examines three different feature proposals, a one feature system, a two feature system, and a three feature system.

2.1 The one feature system

A one feature number system is quite restrictive. In application, it can create no more than two numbers (singular and plural). Anderson (1992) provides just a system, with the number feature [\pm plural]. A theoretical framework with only one number feature fails almost immediately in accounting for empirical data. The data set used in Harley and Ritter (2002:497), for example, contains 91 languages. 18 of those languages (20%) have three numbers, a singular, dual, and plural. Three additional languages have four numbers. With Anderson's [\pm plural] feature, only three fourths of the languages in Harley and Ritter's data can be formally explained.

2.2 The two feature system

The two number system provides a fairly restrictive feature set, but one which allows for a dual number. Adger (2003) provides an argument for a three number system as a part of a much larger work on the Minimalist Program. Although this work is introductory, it presents a number of important assumptions about the nature of number systems including a claim that linguists "simply don't find languages which distinguish four varieties of number feature, and treat them all on an equal basis" (Adger 2003:22).

His system incorporates two features, singular and plural. These features have three possible arrangements, shown below:

Singular: [singular]
Dual: [singular, plural]
Plural: [plural]

The features are monovalent. That is, they only appear with a positive value. Thus, the singular is read as [+singular], not [+singular, -plural]. Monovalency rules out an otherwise predicted fourth arrangement, [-singular, -plural]. Because of this, a monovalent two feature system has only three possible arrangements. Adger thus makes the specific prediction that languages are not capable of forming higher number distinctions. The reason for this lack of ability is formal. Note also that this

³ All works on number include singular and plural features, though the names of these features varies, as do details of their function.

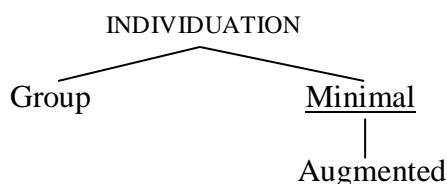
feature system lacks internal structure. Number in this view is composed of feature “bundles” rather than feature hierarchies.⁴

2.3 *The three feature system*

Harley and Ritter (2002) propose a three feature system, organized in a hierarchical structure which generates four numbers. Singular, dual, and plural are formed in the same manner as Adger’s system, albeit with different labels. A third feature, [augmented], is added to generate a fourth number. The three feature geometry is monovalent, like Adger’s feature bundles, so there are no negative feature values.

- Singular: [Minimal]
- Plural: [Group]
- Dual: [Minimal, Group]
- Paucal: [Minimal, Group, Augmented]

The features are implicational. Activation of the feature [augmented] implies activation of the feature [minimal]. Thus, the internal structure is hierarchical, and is represented with a featural geometry. The morphological feature geometry takes its inspiration from phonological feature geometries (see Clements 1985). In Harley and Ritter’s geometry, the INDIVIDUATION node is responsible for number. Harley and Ritter provide a tree representation of their feature geometry on page 486. A portion of that tree is re-printed below, focusing only on the section responsible for number.

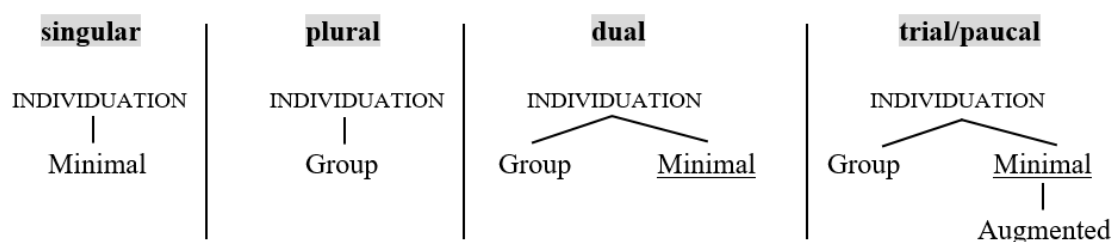


The dual number in this geometry predicts Greenberg’s (1963:94) Universal 34, that “no language has a dual unless it has a plural” because the dual relies on an “independently active” group node (Harley and Ritter 2002:493). The interpretation of dual from the features [group] and [minimal] arise from the logic that the smallest possible set [minimal] that is not singular [group] is in fact, a set of two.

The paucal number, in the feature geometry of Harley and Ritter, is also dependent and implicational. Paucal is achieved through activation of the feature [augmented] which is a daughter of [minimal]. [augmented] can only be activated if both [group] and [minimal] are active. This is a stated feature of the geometry, but is not explicitly represented in the formalism. Also, the same logic that derives dual from [minimal] and [group] is used to derive paucal. However, this logic does not distinguish trial from paucal, as implied in Harley and Ritter (2002:494) “...the paucal consists of the smallest possible group [the dual] plus one (trial) or a few (paucal).” It follows from this that if [minimal, group, augmented] is interpreted as a trial, then there can be no paucal, and if it is interpreted as a paucal, then there can be no trial. The structure, and its ambiguity on trial and paucal numbers, predicts the absence of any number system with five distinctions.

⁴ As pointed out in Harley and Ritter (2002) this is a common feature of earlier morphological analysis, e.g. Anderson (1992).

Figure 1: The four numbers of the feature geometry



We can thus interpret the structure of this system as specifically built to constrict the possible number distinctions to four, a decision that was influenced by the apparent lack of five number systems in the sources that they consulted. To quote Harley and Ritter (2002:496) “No language has more than four numbers, and so the geometry permits all and exactly the attested person and number distinctions in the world’s languages.”

3 Typological descriptions of number

Like theory, typological accounts of what is and is not attested in the world’s languages are constantly up for revision. New data often calls into question older ideas of what constitutes human speech. Thus, it is not surprising that a relatively recent work which focuses only on number (Corbett 2000) contradicts older typological works which include number (for example, Ingram 1978). The contradiction applies to exactly how many number distinctions are present in the world’s languages. Ingram (1978), using what data was available to him at the time, found a maximum of four number distinctions. Corbett (2000) updates the typological literature, noting the presence of five number systems in several languages in the Pacific.

3.1 Typological four number maximum

Ingram (1978:227), in his typology of personal pronouns, claims that there exist only three possible number systems. These three systems are a singular and plural system, a singular, dual, and plural system, and a singular, dual, trial, and plural system. The largest of his three systems makes four number distinctions. Thus, his typology of number matches the predictions of Harley and Ritter’s featural geometry; there can be at most four distinctions between number in a single language.

Ingram describes his typology with three conceptually different types of plural. A system with only singular and plural has a more-than-one plural. A system with a dual has a more-than-two plural. A system with a trial has a more-than-three plural. Ingram uses the term “feature” to describe these types of plural, but his features are specific, rather than general. There is, for example, a specific “dual” feature, a “more-than-two” feature, and so forth. Thus, his “features” are not comparable to the features outlined in section two, which are general, universal, and which rely on the interpretation of feature combinations.

3.2 Typological five number maximum

Modern works on number have recognized that a number of languages do in fact distinguish five numbers. Corbett (2000) and Cysouw (2003) are two works that attempt to describe the possible number systems of the world’s languages. Corbett is probably the most comprehensive account of number to date, while Cysouw’s main focus is pronominal systems and person marking, not specifically number. Nevertheless, the two works agree that the most distinctions made by any language are five, not four. These two works also discuss important issues with the trial/quadral/paucal distinction, and the geographic distribution of complex number systems (number systems with greater than three distinctions).

3.2.1 The trial/quadral/paucal problem

Both typological and theoretical accounts of number recognize not only the rarity of trials and quadral, but also issues with their description. Corbett (2000) provides a list of four Austronesian languages that have developed a five number system; Marshallese (Micronesian, from Bender 1969), Sursurunga (South New Ireland/West Solomonian, from Hutchisson 1986), Tangga (South New Ireland/West Solomonian, from Capell 1971 and Beaumont 1976), and Lihir (New Ireland, from unpublished field data collected by Malcolm Ross). Of these four languages, only the first three are said to have a quadral.

The best descriptive data available are for Sursurunga, and it is here that Corbett makes his argument for interpreting “quadral” as paucal. For Sursurunga, there are at least four reasons why paucal may be the better label for what has traditionally been called a quadral. 1) with kinship pairs such as my four/five/six uncles, the plural is never used. In these cases, the paucal is used to indicate a minimum of four, but not a maximum. 2) quadral are used in hortative statements that include the speaker. Here, the speaker is suggesting joint action, and the quadral is used even if the number of participants is more than four. 3) in Sursurunga the dual is used strictly for two people, regardless of the situation. According to Corbett, such a strict usage is also expected for anything that is to be considered a true quadral. 4) The trial is quite often used to refer to small groups that may consist of four participants, where one expects the quadral. Given these facts, Corbett re-analyzes the Sursurunga system from one with singular, dual, trial, quadral, and plural to one with singular, dual, lesser paucal, greater paucal, and plural. There remain five distinctions in his analysis, but the nature of those distinctions has changed considerably.

3.2.2 The geographic distribution of five number systems

Perhaps the most relevant claim in the typological literature, from the perspective of this paper, is the limited geographic distribution of five number systems. Both Corbett (2000:25–30) and Cysouw (2003:233) claim that five number systems are found *only* in the Oceanic subgroup of Austronesian languages. Cysouw’s work is more restrictive, as he claims that both four and five number systems are restricted to the “Pacific.” The data examined in the remainder of this paper come from languages of northern Sarawak and East Kalimantan on the island of Borneo, located in island Southeast Asia. The complex number systems found in Borneo, discussed in detail below, discredit geographic and linguistic restrictions that state that five number systems are found only in the Oceanic subgroup, or only in the Pacific.

4 Kenyah pronouns and number

Kenyah is a group of languages that are spoken in northern Sarawak and East Kalimantan on the island of Borneo that descend from a common ancestor, known as Proto-Kenyah (PKEN). Kenyah has developed a very complex pronoun system which contains not only a distinction between inclusive and exclusive, but also five number distinctions. A system of five numbers is reflected in separate primary branches of PKEN, which suggests that the modern system was inherited from the proto-language. Some work including lists of pronouns in various Kenyah dialects can be found in Blust (2013 [2009]), which contains data for the five number system of Lepo’ Sawa dialect of Long Anap, and Soriente (2013) which contains data for reduced four number systems in Lebu’ Kulit and Òma Lóngh dialects.

In this section I present data on the pronouns of four Kenyah languages. For one dialect, Lebo’ Vo’, I give a detailed description of the pronominal system, including how the higher number pronouns are used in speech. I give an additional description of the pronouns of Uma’ Pawe, a dialect which has reduced the PKEN five number system to four numbers. After presenting these data, I provide a reconstruction of the PKEN pronouns that show a five number system with 19 separate pronouns indicating clusivity, singular, dual, trial, quadral, and plural.

4.1 Lebo' Vo'

Lebo' Vo' is a Western Lowland dialect of Kenyah (Smith 2015) and is fairly divergent from the other Kenyah varieties spoken along the Baram river. The PKEN five number distinction is retained in Vo', with several important phonological innovations. In a process of grammaticalization (discussed in greater detail in §4.4) the PMP first person plural inclusive, *kita⁵, was paired with numerals to create dual, trial, and quadral forms in Pre-PKEN; e.g. Pre-PKEN *kita dua 'first person dual inclusive'. In PKEN, all available evidence points to a fused form, where *t from *kita replaced the first consonant of the numeral, thus, *kita dua > *tua. This fusion holds true for all of the numbers, in all modern Kenyah languages. Lebo' Vo', as spoken at Long San and Long Ikang, has expanded fusion of the pronominal and numeral elements to the second person. In PKEN, the second person plural *ikəm became associated with the numerals two through four. Using the dual again, this gave rise to the innovative form *ikəm dua 'second person dual'. Lebo' Vo' replaced the first consonant of the numeral with the *k from *ikəm, giving rise to the fused form *kui* 'second person dual'. The same pattern is found in the trial (see example 1 below). Lebo' Vo' is the only Kenyah language that has fused forms for the second person.

(1)	PKEN	Lebo' Vo'
	*ikəm dua	<i>kui</i>
	*ikəm təlu	<i>kəlu</i>

In the third person, Lebo' Vo' typically fuses the plural pronoun with the numeral. This fusion appears only in speech, and elicited forms still produce a two word third person. The fusion pattern replaces the first consonant of the numeral with *r* from *iri* 'third person plural'. This produces the forms in the example below.

(2)	PKEN	Lebo' Vo'
	*ida dua	<i>iri lui ~ rui</i>
	*ida təlu	<i>iri təlu ~ rəlu</i>

4.1.1 The dual in Lebo' Vo'

From examples in texts and language experience at Long San, the dual form in Lebo' Vo' refers to exactly two people, in all contexts. The dual keeps a distinction between inclusive and exclusive. The inclusive dual refers to the speaker and to the person being spoken to. The exclusive dual refers only to the speaker and one other person, but not the person being spoken to. In the second person, the dual is used by the speaker to address a group of exactly two people. The third person dual appears mostly in storytelling. There are several examples of the dual used in both conversation and in collected texts. Three of those examples are shown below.

⁵ The PMP reconstructions used in this paper can be found in Blust and Trussel (ongoing). Although the forms here are probably the most widely accepted, Reid (2009, 2016) presents a radically different reconstruction of PMP pronouns. In his papers, *kita is reconstructed as a dual, and is retained as such in several Philippine languages. Using only Philippine evidence, Reid (2016) reconstructs *=mu as a plural formative in PMP, which combined with *=ta (the genitive short form of *kita) to form the first person inclusive plural *=tamu. This paper, however, has neither the scope nor space to address these claims, other than to state that because Reid uses only Philippine evidence to form his reconstructions, I have chosen to continue to use Blust's reconstructions. The Philippine languages likely represent only a single subgroup in Malayo-Polynesian, and thus evidence from the Philippines alone cannot provide evidence for a PMP reconstruction.

- (3) *ame? lui ləpəh tai tə uma? Michael*
 1dl.ex already go to house Michael⁶
 ‘The two of us have already gone to Michael’s house’. (Written on a note)

- (4) *laya? wi kuman?*
 good 1dl.in eat
 ‘Shall we eat?’

- (5) *mafo? rui pə-pəsoy*
 a while 3dl recp-talk
 ‘The two of them talked to each other for a while’.

4.1.2 The trial in Lebo’ Vo’

In texts, there are very few examples of trials being used. It is rare that one finds a need for the trial in stories about individuals. On one occasion however, while doing research at Long Palai on the upper Baram river, my consultant turned to me and one other man and uttered the following sentence,

- (6) *təlu tai kuman*
 1tr.in go eat
 ‘The three of us will go to eat’.

From this interpretation alone, it seems that the trial is best interpreted as an actual trial. That is, a pronominal number that refers to precisely three participants. In elicitation, this is certainly true. When asked directly, speakers of Lebo’ Vo’ say that the trial means three, and refers to only three people. Groups of four or more speakers are consistently referred to with a different pronominal number, either the quadral or the plural.

4.1.3 The quadral in Lebo’ Vo’

During elicitation sessions, pronouns formed with *pat* ‘four’ were given when plurals were requested.⁷ When eliciting the first person pronouns, I typically asked for a word used for ‘we, as in the four of us’. The word given is always *təpat*. When I moved on to the next number, ‘we, as in the five of us’, *təpat* was again given as the appropriate word. Following this, I then asked, ‘we, for any amount of people’ to which *təpat* was again given as the most appropriate word. Most groups of people with four or more members are referred to with a pronoun formed with *pat*. However, it is not necessarily the case that *təpat* is used as a plural. Another set of pronouns, which refer to large groups of people, was also elicited. This argues against a strict reading of ‘quadral’ or ‘plural’ in the modern languages. In Lebo’ Vo’ the “quadral” is actually a paucal.

An important distinction between paucal and plural is made when referring to a group of people performing the same task. For example, during elicitation the following situation was constructed: a group of people are traveling from one city to another. When the paucal is used in this situation, the sentence is interpreted as meaning that the group will be traveling in the same car, or the same boat. When the plural is used, however, the sentence is interpreted as meaning that the group is traveling in several cars or boats. Two sentences were given to show this distinction:

⁶ The following abbreviations are found in glosses throughout this paper: 1 = first person, 2 = second person, 3 = third person, DL = dual, TR = trial, PC = paucal, PL = plural, EX = exclusive, IN = inclusive, RECP = reciprocal.

⁷ First person exclusive *ame? pat*, second person *ikəm pat*, and third person *iri pat* were elicited, but most speakers indicated that they were no longer used.

(7) *daw vam* *təpat* *tai* *site*
 tomorrow 1PC.IN go there
 ‘Tomorrow we will go there (using the same car)’.

(8) *daw vam* *ilu* *tai* *site*
 tomorrow 1PL.IN go there
 ‘Tomorrow we will go there (using more than one car)’.

4.1.4 The plural in Lebo’ Vo’

If plurality is a designation of greater than one, then there arise numerous situations where plural pronouns may be used. In Kenyah, however, the plural is not simply ‘more than one’. Because there is a dual to refer to two people, a trial to refer to three people, and a paucal that can refer to a “small group,” the plural is quite restricted. It can only appear in stories to refer to a large group of people. In story-telling and in conversation, most groups are small and consist of two or three participants, and in Lebo’ Vo’ these are indicated by the dual and trial. One example of a true plural being used is in reference to an entire village. Kenyah villages traditionally consist of a single longhouse with as many as 200–400 individuals. From a numerical point of view, the plural is the most appropriate form for referring to an entire village. One such example is given below.

(9) *itu* *uniŋ wap* *ubən* *adət* *ame?* *adət* *ame?* *ləbo?*
 this Uning Wap because law 1PL.EX law 1PL.EX village
 ‘This, Uning Wap, is because of our traditional laws of the village’.

4.2 Uma’ Pawe

The pronominal system of Uma’ Pawe has been reduced from the five number system found in PKEN to a four number system, with distinctions between a singular, dual, paucal, and plural. Reduction of the pronominal system targeted the quadral series; Uma’ Pawe is the only dialect used in this analysis without reflexes of PKEN *təpat, *ami? pat, *ikəm pat, and *ida pat (but see the appendix for a full list of languages that have lost the quadral). The trial has been reanalyzed as a general paucal, while the dual remained unchanged.

In reflexes of both the dual and trial forms, Uma’ Pawe reduced the two syllable pronominal and two syllable numeral to monosyllables. For example, PKEN *ami? dua ‘first person exclusive dual’, was reduced to *mi? we*.⁸ Data used for this paper however, does not show a similar reduction in the third person, which remains disyllabic, *ira*. Example 10 below demonstrates how these pronouns were reduced in Uma’ Pawe.

⁸ The phonetic form was [duwa] with a phonetic transition glide from *u to *a. Loss of the initial syllable left only [wa]. Final *-a was fronted to -e in Uma’ Pawe, giving the modern form *we*. An anonymous reviewer pointed out that Uma’ Pawe *tua* ‘we two inclusive’ and *ira* ‘they’ did not in fact front *a to e. Numerous examples outside of the pronouns, however, show the change *a > e. Some of these examples are *əj^ha > *se* ‘one’, *dua > *lue* ‘two’, *usah > *use* ‘body’, *mata > *mate* ‘eye’. Although Blust (2000) describes low-vowel fronting in many of the languages of northern Sarawak, those instances were all conditioned by the presence of a voiced obstruent before *a in the final syllable, and without an intervening blocking consonant. In Uma’ Pawe however, *usah > *use* ‘body’ and *mata > *mate* ‘eye’ do not seem to be exhibiting the same phenomenon as that described by Blust. Moreover, fronting in Uma’ Pawe is restricted to *a in final position or where *a came to be in final position after the deletion of PKEN *-h. *a never fronts in a closed final syllable. Thus, this fronting seems to be a word final phenomenon, much like raising of *-a to *i* in Lebo’ Vo’ and is not conditioned by voiced obstruents earlier in the word. What is odd about the Uma’ Pawe data with regard to this paper is that the only exceptions that I know of are in fact *tua* ‘we dual inclusive’ and *ira* ‘they’. It is not immediately apparent why the only two exceptions are found in the pronouns, but regular sound correspondences in other Kenyah languages, as well as in Uma’ Pawe *kəm *we* ‘you two’, support the reconstructions in this paper.

(10)	PKEN	Uma' Pawe
	*ami? dua	mi? we
	*ami? təlu	mi? təw
	*ami?	ami?
	*ikəm dua	kəm we
	*ikəm təlu	kəm təw
	*ikəm	kəm

Uma' Pawe has made another reduction to its pronouns. In the first person paucal, but not in the dual or the plural, there is no recorded distinction between inclusive and exclusive. A reflex of the PKEN first person trial exclusive, *ami? təlu, is found for both inclusive and exclusive. The correct glossing for Uma' Pawe *mi? təw* is thus 'first person paucal', with no inclusive or exclusive distinction.

4.3 Lepo' Tau and Lepo' Gah

Both Lepo' Tau and Gah are Highland Kenyah languages, and fairly conservative. There are no major innovations that distinguish the Highland pronouns from PKEN.

4.4 The Proto-Kenyah pronouns

Reconstructing a five number system is supported by regular correspondences between the pronominal systems of languages in at least two primary subgroups, Highland and Lowland Kenyah. The languages used in this paper for reconstruction are Lebo' Vo' and Uma' Pawe (two lowland languages), and Lepo' Gah and Lepo' Tau (two highland languages). Subgrouping assumptions are based on Smith (2015). A table with reconstructed forms and evidence from modern languages is given below, followed by etymologies for each word.

Table 1: Proto-Kenyah, Lebo' Vo', Uma' Pawe, Lepo' Gah, and Lepo' Tau pronouns

English	PKEN	Vo'	Pawe	Gah	Tau
1SG	*aki?	ake?	aki?	ake?	ake?
2SG	*iku?	iko?	iku?	iko?	iko?
3SG	*ia	yi	ye	ya	ia
1DL.IN	*tua	wi	tua	tua	tua
1DL.EX	*ami? dua	ame? lui	mi? we	ame? dua	ame? dua
2DL	*ikəm dua	kui	kəm we	kəm dua	ikəm dua
3DL	*ida dua	iri lui	ira we	ida dua	ida dua
1TR.IN	*təlu	təlu	mi? təw	təlu	təlu
1TR.EX	*ami? təlu	ame? təlu	mi? təw	ame? təlu	ame? təlu
2TR	*ikəm təlu	kəlu	kəm təw	kəm təlu	ikəm təlu
3TR	*ida təlu	iri təlu	ira təw	ida təlu	ida təlu
1QD.IN	*təpat	təpat		təpat	təpat
1QD.EX	*ami? pat	(ame? pat)		ame? pat	ame? pat
2QD	*ikəm pat	(ikəm pat)		kəm pat	ikəm pat
3QD	*ida pat	(iri pat)		ida pat	ida pat
1PL.IN	*ilu	ilu	iləw	ilu	ilu
1PL.EX	*ami?	ame?	ami?	ame?	ame?
2PL	*ikəm	ikəm	kəm	kəm	ikəm
3PL	*ida	iri	ira	ida	ida

***aki?** 'first person singular'. Lebo' Vo' *ake?*, Uma' Pawe *aki?*, Lepo' Gah *ake?*, Lepo' Tau *ake?*. Support for reconstructing *aki? is widespread. All Kenyah languages reflect *aki?, which is an

innovation that replaced PMP *aku. Most Kenyah languages have lowered the ultimate vowel from *i to *e*, but evidence from Uma' Pawe *aki?* supports reconstructing a high vowel. Reflexes of *aki? are found only in Kenyah languages. This is thus a highly distinctive innovation attributable to PKEN and has been used as a diagnostic in delineating the Kenyah subgroup (Smith 2015).

***iku? 'second person singular'**. Lebo' Vo' *iko?*, Uma' Pawe *iku?*, Lepo' Gah *iko?*, Lepo' Tau *iko?*. Like *aki?, PKEN *iku? is both innovative and supported by reflexes in all Kenyah languages for which data are available. PKEN *iku? replaced PMP *i-kahu. Once again, the majority of Kenyah languages have lowered the ultimate vowel, but evidence from Uma' Pawe *iku?* supports the reconstruction proposed here.

***ia 'third person singular'**. Lebo' Vo' *yi*, Uma' Pawe *ye*, Lepo' Gah *ya*, Lepo' Tau *ia*. This is a retention from PMP *si-ia, and requires no special attention.

***tua 'first person dual inclusive'**. Lebo' Vo' *wi*, Uma' Pawe *tua*, Lepo' Gah *tua*, Lepo' Tau *tua*. This reconstruction is supported by Uma' Pawe *tua* as well as Lepo' Gah and Lepo' Tau *tua*. Lebo' Vo' has simplified the pronoun by dropping the consonant, innovating the form *wi*. The first person dual inclusive is best analyzed as a fusion. In Pre-PKEN, it appears that the PMP first person inclusive plural *kita formed higher number pronouns by combination with the numerals two through four. This would have given rise to the forms *kita dua, *kita təlu, and *kita əpat. The onset of the final syllable of *kita then replaced the onset of the numeral. For the first person dual inclusive, the history can be visualized as follows: *kita dua > t+ua > *tua. This fused form was then inherited in all Kenyah languages. Note that the regular sound change PKEN *-a > Uma' Pawe *-e did not occur in *tua*. This exception remains unexplained.

***ami? dua 'first person dual exclusive'**. Lebo' Vo' *ame? lui*, Uma' Pawe *mi? we*, Lepo' Gah *ame? dua*, Lepo' Tau *ame? dua*. It is clear that PKEN had not fused the inherited pronominal with the numeral, as is the case with the inclusive series. All Kenyah languages reflect two separate words.

***ikəm dua 'second person dual'**. Lebo' Vo' *kui*, Uma' Pawe *kəm we*, Lepo' Gah *kəm dua*, Lepo' Tau *ikəm dua*. The second person dual was also formed by associating an inherited pronominal, PMP *i-kamu, with the numerals two through four.

***ida dua 'third person dual'**. Lebo' Vo' *iri lui*, Uma' Pawe *ira we*, Lepo' Gah *ida dua*, Lepo' Tau *ida dua*. This is a straightforward association of PMP *si-ida 'third person plural' with the numeral 'two'.

***təlu 'first person trial inclusive'**. Lebo' Vo' *təlu*, Lepo' Gah *təlu*, Lepo' Tau *təlu*. Although the modern form is homophonous with the numeral three in all Kenyah languages that reflect *təlu, comparative evidence suggests that the first person trial inclusive formed through a fusion of PMP *kita with the numeral *təlu 'three'. If the formula for generating the first person dual inclusive is applied to the trial, the result is PKEN *təlu. Start with the pronominal plus the numeral, *kita təlu. Take the onset of the final syllable of the pronominal, *t, and use it to replace the onset of the numeral. Thus *kita təlu > *t+əlu > *təlu. Note that in the Uma' Pawe dialect, no clusivity distinction was recorded in the first person trial. The historical exclusive *ami? təlu is reflected as *mi? təw* with no clusivity (see table 1 above).

***ami? təlu 'first person trial exclusive'**. Lebo? Vo? *ame? təlu*, Uma' Pawe *mi? təw* 'first person paucal', Lepo' Gah *ame? təlu*, Lepo' Tau *ame? təlu*.

***ikəm təlu 'second person trial'**. Lebo' Vo' *kəlu*, Uma' Pawe *kəm təw* 'second person paucal', Lepo' Gah *kəm təlu*, Lepo' Tau *ikəm təlu*.

***ida təlu ‘third person trial’.** Lebo’ Vo’ *iri təlu*, Uma’ Pawe *ira təw* ‘third person paucal’, Lepo’ Gah *ida təlu*, Lepo’ Tau *ida təlu*.

***təpat ‘first person inclusive quadral’.** Lebo’ Vo’ *təpat*, Lepo’ Gah *təpat*, Lepo’ Tau *təpat*. If we apply the same formula used to form the PKEN first person dual and trial inclusive pronouns to the quadral, we can derive the form *təpat from Pre-PKEN *kita əpat. Take the onset of the final syllable of the pronominal, *t, and use it to replace the onset of the numeral (or if no onset is present, simply prefix *t to the numeral). Thus, *kita əpat > *t+əpat > *təpat. A reflex of *təpat is found in all Kenyah languages with the exception of Uma’ Pawe, which has lost the quadral completely.

***ami? pat ‘first person exclusive quadral’.** Lebo’ Vo’ *ame? pat*, Lepo’ Gah *ame? pat*, Lepo’ Tau *ame? pat*. Although Uma’ Pawe seems to have lost reflexes of the quadral, *ami? pat (with a high vowel) is reconstructed here based on Uma’ Pawe *ami?* ‘first person exclusive plural’.

***ikəm pat ‘second person quadral’.** Lebo’ Vo’ *ikəm pat*, Lepo’ Gah *kəm pat*, Lepo’ Tau *ikəm pat*.

***ida pat ‘third person quadral’.** Lebo’ Vo’ *iri pat*, Lepo’ Gah *ida pat*, Lepo’ Tau *ida pat*.

***ilu ‘first person plural inclusive’.** Lebo’ Vo’ *ilu*, Uma’ Pawe *iləw*, Lepo’ Gah *ilu*, Lepo’ Tau *ilu*. This is an innovative form, and is found in several Kenyah dialects including all dialects used for this paper. It does not seem to be derived from a pronoun + numeral, and its history remains obscure. Nevertheless, it must be reconstructed to PKEN.

***ami? ‘first person plural exclusive’.** Lebo’ Vo’ *ame?*, Uma’ Pawe *ami?*, Lepo’ Gah *ame?*, Lepo’ Tau *ame?*. The pronominal *ame? appears in all first person exclusive pronouns. When alone, it indicates the plural, but when paired with numerals, it has either a dual, trial, or quadral interpretation. If *ame? reflects PMP i-(k)ami, then one must explain the innovation of a word final glottal stop. Sporadic word final glottal stop insertion is an areal feature of much of the island of Borneo. Iban, Brunei Malay, Kayan, Modang, Segai, Tunjung, Berawan, Penan, Sebop, Ngorek, and Kenyah all have varying degrees of word final glottal stop insertion or deletion. Because of this, the appearance of final glottal stop in this pronoun is not as troubling as one might think.

***ikəm ‘second person plural’.** Lebo’ Vo’ *ikəm*, Uma’ Pawe *kəm*, Lepo’ Gah *kəm*, Lepo’ Tau *ikəm*. Although PKEN *ikəm is quite similar to PMP *i-kamu, there are two inexplicable exceptions in the sound correspondences. Both PNS *a and *u were inherited unchanged in PKEN. Loss of final *-u and raising of *a are unexpected. It is likely that when the prefix *i- was fossilized to the base, giving Pre-PKEN *ikamu, that the final vowel was deleted due to internal pressures which favor a two syllable canonical word. Raising of *a to *ə however, must remain an unexplained irregular sound change.

***ida ‘third person plural’.** Lebo’ Vo’ *iri*, Uma’ Pawe *ira*, Lepo’ Gah *ida*, Lepo’ Tau *ida*. This is a retention of the PMP third person plural, *si-ida and requires no special attention. Note, however, that Uma’ Pawe again failed to raise word final -a to -e.

4.5 Reconstructing quadral

In the data used for this paper, there is not enough descriptive material to make a definitive choice between quadral or paucal as the most appropriate gloss for PKEN pronouns formed with *pat. In Lebo’ Vo’ it seems that paucal is the correct choice. Blust (2013 [2009]:318) offers pronominal data from the Kenyah variety spoken at Long Anap where the label ‘quadral’ was used for pronouns formed with *pat*. Because of that, the gloss ‘quadral’ was chosen for reconstruction in the above section. However, Corbett (2000:26–30) notes that where detailed descriptions are available, ‘quadrals’ are best analyzed as ‘paucals’. The Kenyah data is no exception. Only word lists and a few sentences are available for most dialects of Kenyah. However, because I was able to spend several

weeks at Long San, I had the opportunity to perform a more in-depth analysis and decided on paucal for Lebo' Vo'.

5 Number systems in languages of Borneo

Complex number systems have developed in several languages of Borneo. This seems to be part of a linguistic area that covers most of northern Sarawak and East Kalimantan. Some of those languages are described below, and include Kayan, Kiput, Kelabit, Sa'ban, and Punan Bah. Kelabit and Kiput are found in separate branches of North Sarawak. Kenyah languages also belong in this subgroup, but Kayanic languages apparently do not (Blust 2010, Smith 2015).

5.1 Kayan

Kayanic languages show four numbers in their pronoun systems. The Kayan spoken at Long Naah is typical of Kayan languages in general (see table 2). The Kayan system is notable for its large number of fused forms. In many cases, duals and trials are not easily analyzable as a pronoun and a grammaticalized numeral. Dual forms, with the exception of *ito?* 'first person dual inclusive' reflect the final syllable of *dua as *wa?* plus the first consonant and final vowel of the plural pronoun. Thus *kita dua > k+a+wa? > *kawa?* 'first person dual exclusive'.

One might propose that this formula affected all pronouns, but it only holds true for first person exclusive dual and trial, second person dual, and third person dual and trial. The formula fails for second person trial, where **kulo? is predicted, and for all first person inclusive forms.

Table 2: Kayan pronouns

English	Kayan, Long Naah			
	singular	dual	trial	plural
1IN	akoy	ito?	təlo?	itam lim
1EX		kawa?	kalo?	kame?
2	ika?	kuwa?	kəlo?	ikam
3	iha?	dawa?	dalo?	daha?

5.2 Kiput

Kiput is a Lower-Baram language, and exhibits a four number system. Data here is from Blust (2003). The pronoun section in this work is limited. It was never intended to be a thorough description of the language. It does, however, provide clear evidence that Kiput has at least a singular, dual, trial, and plural. Whether or not the trial may be analyzed as a paucal remains to be seen. The full pronoun set is shown below in table three.

Table 3: Kiput pronouns

English	Kiput			
	singular	dual	trial	plural
1IN	kaw	kifih	killaw	kiteh
1EX		kafih	kallaw	kamay
2	naw	ifih	illaw	uñew
3	ñih	difih	lew	idih

5.3 Kelabit

Data below is from the Bario dialect of Kelabit as presented in Blust (2013 [2009]). Again, the work where these data were published was not a description of Kelabit, so information on how these forms are used in sentences is limited. Like Kiput and Kayan, however, Kelabit (in table 4) shows a clear four number system, regardless of how those numbers are interpreted.

Table 4: Kelabit pronouns

English	Kelabit			
	singular	dual	trial	plural
1IN	uih	kitəh	təluh	tauh
1EX		kədiwəh	kətəluh	kamih
2SG	iko	məduəh	mətəluh	muyuh
3SG	iəh	diwəh	dətəluh	idəh

5.4 Sa'ban

Sa'ban is a highly divergent member of the Kelabit-Lun Dayeh group of North Sarawak. Rapid sound change has left many of the pronouns unrecognizable (compare the Sa'ban forms in table 5 to the Kelabit forms in table 4). Its pronominal system is complex, and needs to be sorted out with care. Of special interest is the trial and paucal series. There are two trials, formed from the number three, *məlaw* 'first person trial exclusive' and *kalaw* 'second person trial'. Blust (n.d.) explicitly points out that there is no trial for the first person inclusive. There is, however, a paucal. Interestingly, the paucal in Sa'ban is formed with the number 'four', which indicates that Sa'ban had, at one time, a fully functional five number system like that found in Kenyah. The third person also has a paucal rather than a trial, but the third person paucal is formed with the numeral three, not four. Blust (n.d.) also has a second person paucal listed as identical to the third person paucal. If this is accurate, then Sa'ban also holds the distinction of being one of a small number of languages that have a second-person/third-person pronoun. The full set is given below in Table 5.

Table 5: Sa'ban pronouns

English	Sa'ban				
	singular	dual	trial	paucal	plural
1IN	ek	tah	tah pat		tam
1EX		aməw	məlaw		aməy
2SG	cəh	keəh	kalaw	dalaw	ciəm
3SG	yəh	deəh	dalaw		dəh

It seems clear that Sa'ban had at one time a five number system. There was a trial series formed with 'three', and a quadral/paucal series formed with 'four'. Sa'ban reduced this system in an asymmetrical fashion, and replaced the first person trial inclusive with the paucal, but lost the paucal everywhere else. This change parallels Lebo' Vo' Kenyah, where the first person inclusive paucal is the only paucal in regular use.

5.5 Punan Bah

Punan Bah is an under-described language spoken along the Rejang river, in central Sarawak. Smith (2017) contains the only full account of Punan Bah pronouns and shows a fully developed 5 number system similar to the one described earlier in Lebo' Vo'. The paucal series in Punan Bah is formed with a reflex of *əpat 'four', which also resembles the quadral in Lebo' Vo'. Punan Bah does not subgroup immediately with Kenyah. Rather, it is part of the Punan subgroup as proposed in Smith (2017), which is itself part of the larger Central Sarawak group. It is thus not closely related to any North Sarawak language. Table 6 below shows the complete pronoun system.

Table 6: Punan Bah pronouns

English	Punan Bah				
	singular	dual	trial	paucal	plural
1IN	oəʔ	tou	tolu	topat	to
1EX		kuo	kotolu	kipat	kai
2SG	kou	komo	komo tolu	kopat	kom
3SG	en	duo	do tolu	dopat	do

The evidence for a northern Sarawak East Kalimantan linguistic area that has affected number systems in several languages is quite strong. Evidence from several North Sarawak languages and from Kayan show independent innovations of complex number systems in Borneo. Even more languages are listed in the appendix, showing the number of pronominal distinctions and their type (singular, dual, trial, quadral, paucal, or plural). Because there are distinct trials and paucals in Blust (n.d.), Sa'ban is listed in the appendix as having four *or* five numbers, pending further investigation.

6 Development and evolution of complex number

Cysouw (2003:236) examines the case of Bolaang Mongondow, spoken on Sulawesi. In this language, a set of “determined” pronouns can be paired with a numeral, giving a reading somewhat similar to English “the three of us” or “the four of you.” Here, the higher number “pronouns” are not pronouns at all, but simply reflexes of the PMP plural and a number. It is not a restricted set, and forms like “the fourteen of us” or “the thirty of you” are theoretically possible. One must then ask if the higher number pronouns in Kenyah are similarly formed by adding a numeral to the pronouns.

To this question, the answer is straightforwardly no. There are two reasons why this reading is inappropriate for the Kenyah data. First, the higher number pronouns in Kenyah are a closed set. Numbers higher than four cannot be added to the pronouns. While eliciting data on Kenyah languages, I naturally asked questions like the following, “If I can say *ida dua* for two people, *ida talu* for three people, and *ida pat* for four people, can I say *ida lama* for five people?” The response to this question was an emphatic no, in all dialects. The pronouns in Kenyah are not formed by the synchronic adding of numerals to pronominal bases.

Second, while many of the higher number pronouns are formed with two clear historical components, the first person inclusive set can be reconstructed to PKEN as a single morpheme without any analyzable morpheme boundary. Thus, even if one were to analyze *ami? pat ‘first person exclusive quadral’ as a pronoun plus number, the same cannot be done for *təpat ‘first person inclusive quadral’, which is a single word with no morpheme boundary. The Lebo’ Vo’ dialect has innovated single morpheme forms in the second person as well, and in speech the third person is also showing signs of fusing into a single morpheme (§4.1). This dialect in particular strongly resists any analysis that would seek to write off higher number pronouns as simple pronouns plus a specifying number.

Another question which arises from Cysouw’s observation is whether the PKEN higher number pronouns might have arisen through grammaticalization. In the Bolaang Mongondow example, Cysouw hypothesized that the pronoun + numeral strategy will eventually lead to grammaticalized forms with elements of the number three. What Cysouw proposed for Bolaang Mongondow may have already happened in Kenyah. It is fairly obvious from the PKEN reconstructions that the higher number pronouns did in fact arise through a pronoun + number strategy that became grammaticalized in forms like *tua, *təlu, and *təpat. Forms that did not fuse into a single morpheme, like *ami? dua ‘second person dual exclusive’, nevertheless became semantically and lexically “grammaticalized” as single pronouns.

We can make a further inference on the evolution of complex number systems from observing how the five number PKEN pronouns were reduced in Uma’ Pawe and other dialects. First, although the PKEN ‘quadral’ was formed from the numeral *pat ‘four’, the only Kenyah language with a thorough description of number utilizes this series as a paucal. It’s already been noted that the existence of quadral numbers have been called into question. If PKEN truly had a quadral, it has not

survived as such in Lebo' Vo'. Thus, quadrals are highly marked and prone to simplification, or shift to paucal. Second, in the Uma' Pawe dialect, not only was the 'quadral' lost, the entire set of pronouns formed with *pat were lost. The result was a reduction in number distinctions from five to four. This allows us to infer that five number systems are marked, and are likely to reduce in number, by deletion of the higher numbers. In Uma' Pawe, the quadral was deleted, but in other languages, the inherited plural is sometimes deleted, and the next highest number shifts to take its place. The Badeng dialect of Kenyah spoken at Lio' Matoh has lost the inherited plural and the quadral has shifted to a plural (see example 11 below). In Badeng, the first person plural inclusive is *təpat*. Data collected at Lio' Matoh is limited however, and needs elaboration. Blust (n.d.) lists *təpat* as a replacement plural in at least Long Atun and Long Jeeh Kenyah as well.

(11)	1pl.in	<i>təpat</i>
	1pl.ex	<i>ameʔ pat</i>
	2pl	<i>ikəm pat</i>
	3pl	<i>eda pat</i>

Blust (2013 [2009]:318) gives another example of this phenomenon in Melanau (Mukah), a language of Borneo with a three number system (singular, dual, plural) that is not closely related to Kenyah. In Melanau, the dual series was formed by a fusion of PMP plural forms with the numeral *dua, much like the inclusive dual *tua in PKEN. The plural series in the Mukah dialect (but not in all dialects of Melanau), is formed from a past trial, and bears a fused reflex of PMP *təlu 'three' (see example 12 below).

(12)	1pl.in	<i>tələw</i>
	1pl.ex	<i>mələw</i>
	2pl	<i>kələw</i>
	3pl	<i>(də)ləw</i>

It appears that Melanau had at one point a four number system, with a trial or paucal formed from the number three. Over time, the marked four number system reduced to a three number system. This reduction, as predicted, targeted the higher numerals. Interestingly, however, the inherited plurals were lost, and replaced with the innovated trials.⁹

One must eventually ask why inherited plural pronouns, with unambiguous plural meanings, were replaced by pronouns formed with 'four' or 'three'. A possible answer to this question has been alluded to in previous sections. In a two number system, with singular and plural only, plural will refer to any number of people greater than one. Most conversations take place between a very small group of people, and in many cases, between only two people. Because of this, plural pronouns in a five number system are rarely used. In a conversation between two people, the dual will be used. In a conversation between a small group, the trial or paucal (or quadral) will be used. Because the trial or paucal pronouns are used in the majority of cases where more than two people are being addressed, they will eventually come to replace the underutilized plural. This is the most likely reason why one finds plural pronouns formed with the numbers four and three in Borneo and in the Pacific.

Complex number systems can be viewed as arising through grammaticalization of numerals with pronouns. This grammaticalization can sometimes produce strikingly large pronoun systems. These overly-complex systems are unstable, and prone to reduction. Reduction targets the semantics, by shifting trial and quadral forms to paucal, then it targets the numbers themselves, reducing five number systems to four number systems, and further reducing four number systems to three number systems. There is no evidence, however, that three number systems were ever further reduced to two number systems in Borneo. When complex number systems are reduced, it is the higher numbers

⁹ A similar change happened in many Polynesian languages. Hawaiian, for example, has *kā-kou* 'first person inclusive plural', which is formed from the numeral three (in the pronouns, Proto Polynesian *tolu 'three' > *kolu > *kou*).

(paucal and plural) that are lost. It is likely, then, that complex number systems have come and gone in many of the world's languages. They have only appeared in Borneo in the last one or two millennia, and are already being reduced and simplified.

7 Conclusion

The sections above support a strong argument that typological studies of number, which have recognized five number systems only in languages of the Oceanic subgroup, are in need of revision. Five number systems have developed independently in Borneo, separated both geographically and linguistically from Oceanic languages. A five number system is found in several closely related languages and can be reconstructed to PKEN. It remains true, however, that five number systems are found exclusively in the Austronesian family.¹⁰ There is no evidence that large number distinctions are inherited, rather, innovations of three, four, and five number systems has taken place independently in a number of languages.

Furthermore, the area of Borneo occupied by North Sarawak, Melanau, Punan, and Kayanic languages represents a linguistic area where the innovation of complex number systems has flourished in recent history. A number of languages in this area have four number systems, while most of the four number examples in Corbett (2000) are again from the Pacific.

Innovation of higher number pronouns can be fairly readily explained for the dual. It likely arose as a product of one on one conversations. This explains why in many Philippine languages, dual forms were innovated in the first person only (Liao 2008). In Borneo and in the Pacific, different events led to innovation of complex number systems in very different environments. In order to understand how such number systems are innovated, we must begin with a more complete picture of exactly where these systems are found.

Theoretical accounts of number are by their nature restrictive. The goal of theory is to propose a system that generates exactly what is found in the world's languages and nothing more. As more data is gathered, it has become clear that theories of number, and specifically the feature geometry of Harley and Ritter, are in need of revision. The feature geometry is too restrictive and does not allow for attested number systems. In the past, researchers were able to claim that five number systems occurred in only a handful of languages, in a geographically restricted area, and with poor documentation. This no longer seems to be the case. There are numerous examples, from the Kenyah group of languages in Borneo, that five number systems are not geographically restricted, nor are they poorly documented.

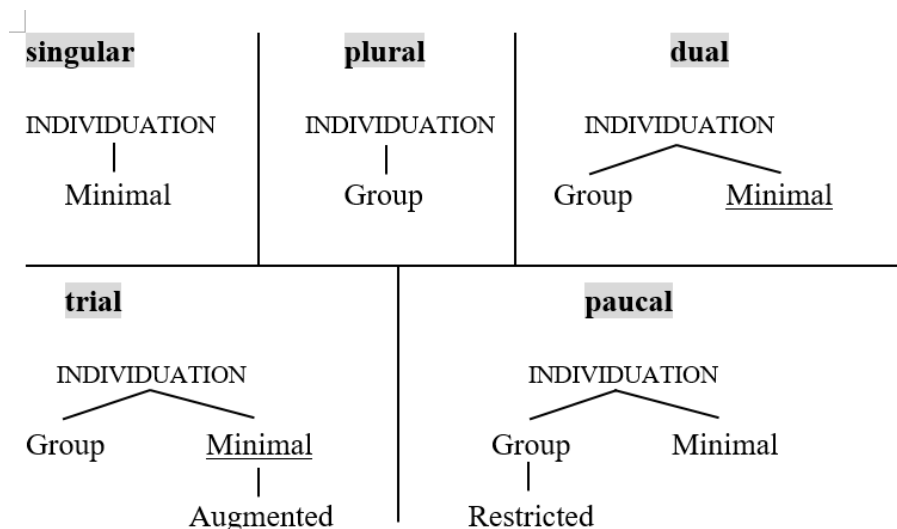
In order to reconcile the feature geometry of Harley and Ritter with new evidence of five number systems, it is necessary to adjust the proposed structure. Harley and Ritter (2002:494) left open this possibility in a footnote where they state (referring to a fifth number) "The existence of such systems could be accommodated in our framework by the addition of a node, probably as a dependent of Group." It is possible to add such a node, and to theorize how a restricted group may better represent a true paucal than [group, minimal, augmented].

An inadequacy of the Harley and Ritter model is that it does not distinguish between trial and paucal. The addition of features that are dependent on minimal will always contain the restriction of "smallest possible set." This inference arises from their treatment of the dual, discussed above as the smallest possible set [minimal] that is not singular [group]. Thus, having [augmented] be a dependent of [minimal] suggests that [group, minimal, augmented] should be interpreted as a trial, which is the smallest possible set larger than two, and not as a paucal. In order to generate a paucal, the geometry needs a restricted group. That is, a group of people that has an approximate range, more than the trial but not plural. This can be achieved by the feature [restricted], under the node [group]. Thus, paucal is an interaction of the features [group, restricted] while trial is an interaction of [group, minimal,

¹⁰ The significance of this restriction is debatable. There are around 1,200 Austronesian languages, comprising one fifth to one sixth of the world's languages. The restriction of five number systems to Austronesian is likely due to the fact that there are so few languages with five number systems, compared with so many Austronesian languages. The restriction of five number systems to Austronesian is best explained as due to chance, rather than as some abstract property specific to the family.

augmented]. A new feature geometry of number is given below, with five possible configurations including the new [restricted] feature needed for deriving the paucal.

Figure 2: A five number feature geometry



Appendix

Central Borneo is home to a linguistic area where complex pronominal number systems have developed independently in a number of languages. The following list indicates which languages in this area have developed number systems of three or more. Note also that all of the languages in this list have a distinction between inclusive and exclusive. Several subgroups are present below, including Kayan-Murik (KAY-M), Kelabit-Lun Dayeh (KEL-LD), Berawan-Lower Baram (B-LB), Bintulu (BIN), Kenyah (KEN), Punan (PUN), Melanau (MEL), and Dusunic (D). Of these groups, KEL-LD, B-LB, BIN, and KEN form a still larger group, North Sarawak. There is no evidence, however, that higher number systems in these languages are inherited from Proto-North Sarawak. Some languages show evidence of past number distinctions, now lost. For example, the plurals in Badeng Kenyah are from a past quadral. These are marked in the list below. Kenyah, Kayan, Punan Bah, and Ngorek number information is from Smith (2017). Ōma Lóngh and Lebu' Kulit Kenyah are from Soriente (2013). All others are from Blust (n.d.).

Kenyah, Lebo' Vo' (KEN)	Five numbers. singular, dual, trial, paucal (from a quadral), plural.
Kenyah, Lepo' Sawa (KEN)	Five numbers. singular, dual, trial, quadral(?), plural.
Kenyah, Lepo' Tau (KEN)	Five numbers. singular, dual, trial, paucal(?), plural.
Kenyah, Lepo' Gah (KEN)	Five numbers. singular, dual, trial, paucal(?), plural.
Punan Bah (PUN)	Five numbers. singular, dual, trial, paucal (from a quadral), plural.
Sa'ban (KEL-LD)	Four/five numbers. singular, dual, trial, paucal (from a quadral), plural
Berawan (B-LB)	Four numbers. singular, dual, trial, plural.
Bisaya (D)	Four numbers. singular, dual, trial, plural.
Kayan (KAY-M)	Four numbers. singular, dual, trial, plural.

Kelabit, Bario (KEL-LD)	Four numbers. singular, dual, trial, plural.
Kenyah, Badeng (KEN)	Four numbers. singular, dual, trial, plural (from a quadral)
Kenyah, Lebu' Kulit (KEN)	Four numbers. singular, dual, paucal (from a trial), plural.
Kenyah, Òma Lóngh (KEN)	Four numbers. singular, dual, paucal (from a trial), plural.
Kenyah, Uma' Pawe (KEN)	Four numbers. singular, dual, paucal (from a trial), plural.
Kiput (B-LB)	Four numbers. singular, dual, trial, plural
Ngorek (KAY-M)	Four numbers. singular, dual, paucal (from a trial), plural.
Penan, Long Labid (KEN)	Four numbers. singular, dual, trial, plural.
Sarikei (MEL)	Four numbers. singular, dual, trial, plural.
Tring (KEL-LD)	Four numbers. singular, dual, trial, plural.
Bintulu (BIN)	Three numbers. singular, dual, plural (from a trial).
Dalat (M-K)	Three numbers. singular, dual, plural (from a trial)
Matu (M-K)	Three numbers. singular, dual, plural (from a trial)
Melanau (M-K)	Three numbers. singular, dual, plural (from a trial)
Narum (B-LB)	Three numbers. singular, dual (incomplete dual), plural.
Penan, Long Merigam (KEN)	Three numbers. singular, dual, plural (from a trial).
Sebop (KEN)	Three numbers. singular, dual, plural (second person from a trial).

References

- Adger, David. 2003. *Core syntax: a Minimalist approach*. Oxford: Oxford University Press.
- Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- Beaumont, Clive H. 1976. Austronesian languages: New Ireland. In *Austronesian languages: New Guinea area languages and language study II*, ed. by Stephen A. Wurm, 387–397. Canberra: Pacific Linguistics.
- Bender, Byron W. 1969. *Spoken Marshallese: An intensive language course with grammatical notes and glossary*. Honolulu: University of Hawai'i Press.
- Blust, Robert. 2000. Low-vowel fronting in northern Sarawak. *Oceanic Linguistics* 39(2):285-319
- Blust, Robert. 2003. *A short morphology, phonology and vocabulary of Kiput, Sarawak*. Canberra: Pacific Linguistics.
- Blust, Robert. 2010. The Greater North Borneo hypothesis. *Oceanic Linguistics* 49:44–118.
- Blust, Robert. 2013 [2009]. *The Austronesian languages*. Canberra: Asia-Pacific Linguistics Research School of Pacific and Asian Studies, The Australian National University.
- Blust, Robert. n.d. Fieldnotes on 41 languages of northern and central Sarawak.
- Blust, Robert and Stephen Trussel. Ongoing. Austronesian comparative dictionary. Online: <http://www.trussel2.com/ACD/>

- Capell, Arthur. 1971. The Austronesian languages of Australian New Guinea. In *Current trends in linguistics VIII: Linguistics in Oceania*, ed. by Thomas A. Sebeok, 240–340. The Hague: Mouton.
- Clements, George N. 1985. The geometry of phonological features. *Phonology Yearbook* 2:225–252.
- Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.
- Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of human language*, ed. by Joseph Greenberg, 73–113. Cambridge, Mass: MIT Press.
- Harley, Heidi and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3):482–526.
- Hutchisson, Don. 1986. Sursurunga pronouns and the special use of quadral number. In *Pronominal systems*, ed. by Ursula Wiesemann, 217–255. Tübingen: Narr.
- Ingram, David. 1978. Typology and universals of personal pronouns. In *Universals of human language*, Vol. 3, ed. by Joseph H. Greenberg, 213–247. Stanford: Stanford University Press.
- Liao, Hsiu-chuan. 2008. A typology of first person dual pronouns and their reconstructibility in Philippine languages. *Oceanic Linguistics* 47(1):1–29.
- Reid, Lawrence. 2009. The reconstruction of a dual pronoun to Proto Malayo-Polynesian. In, *Discovering history through language: Papers in honour of Malcolm Ross*, ed. by Bethwyn Evans, 461–77. Pacific Linguistics 605. Canberra: Pacific Linguistics.
- Reid, Lawrence. 2016. Accounting for variability in Malayo-Polynesian pronouns: Paradigmatic instability or drift? *Journal of Historical Linguistics* 6(2):130–164.
- Smith, Alexander D. 2015. On the classification of Kenyah and Kayanic languages. *Oceanic Linguistics* 54(2):333–357.
- Smith, Alexander D. 2017. *The languages of Borneo: a comprehensive classification*. Unpublished doctoral dissertation. Department of Linguistics, University of Hawai‘i.
- Soriente, Antonia. 2013. Undergoer voice in Borneo Penan, Punan, Kenyah, and Kayan languages. In *Voice variation in Austronesian languages: Linguistic studies of Indonesian and other languages in Indonesia*, ed. by K. Alexander Adelaar, NUSA 54:175–203.

LINGUISTIC EVIDENCE FOR PREHISTORY: OCEANIC EXAMPLES

Malcolm Ross

The Australian National University

Abstract

Historical linguistic analysis of a language family can provide evidence about speakers' prehistory in at least three respects. First, linguistic geography and the reconstruction of a linguistic phylogeny by the comparative method can tell us something about the movements of its speakers from place to place. Second, evidence of contact-induced change can tell us about the interactions of its speakers with groups speaking other languages and occasionally about instances of language shift. Third, reconstructed lexicon may tell us about the culture of speakers of that interstage language, revealing probable features of that culture that are less accessible to archaeology. This article provides examples of these three kinds of evidence, drawn from the study of languages of the large Oceanic subgroup of Austronesian.

Keywords: Oceanic languages, prehistory, linguistic evidence

ISO 639-3 codes: tbo, meu, ksd, nlg, kwd, ton, tah, rar, mri, haw

1 Introduction

This article is a lightly edited transcript of a plenary delivered at the Thirteenth International Conference on Austronesian Linguistics (13-ICAL), which took place at Academia Sinica, Taipei, from 18 to 23 July 2015.¹

The title is 'Linguistic evidence for prehistory: Oceanic examples'. By 'prehistory' I mean human history before detailed written records were kept. I am going to talk about the Oceanic languages of the Pacific, where, in parts, prehistory stretched forward into the late nineteenth century and even into the twentieth.

The foundation of historical linguistics is the comparative method, developed by nineteenth-century Danish (Rask 1818) and German (e.g. Grimm 1848, Schleicher 1861) scholars on foundations laid by others in the preceding two centuries. Recently the statistical approach known as Bayesian phylogenetics, inspired by evolutionary biology, has been applied to data yielded by the comparative method in order to answer questions like, 'What is the most likely phylogeny ('family tree') implied by the linguistic data?' (Greenhill & Gray 2009). I mention it here because the evidence indicates that it has a promising future, but as it lies outside my area of expertise, I will not discuss it further.

I have chosen to use Oceanic examples for three reasons. First, all my fieldwork has been on Oceanic languages in the New Guinea region and I have spent much my career studying Oceanic languages. Secondly, more than 500 of the perhaps 1200 Austronesian languages, or around 40%, belong to the Oceanic subgroup (Figure 1).² Thirdly, Oceanic languages provide a good example of what historical linguistics can contribute to human prehistory, as enough is known to put together information from different disciplines and to start making a coherent story—and an interesting story at that, starting with interaction between Oceanic and Papuan speakers in New Guinea, the sudden appearance in the Bismarck

¹ My thanks go to the organisers of 13-ICAL for inviting me to give this plenary, and to Andrew Pawley for his insightful comments on an earlier version of this transcript. I am also very grateful to two anonymous reviewers whose comments enabled me to correct a number of errors and infelicities.

² All figures are found at the end of the paper.

Archipelago of the Lapita culture, and then the conquest by outrigger canoe of the planet's last great area of habitable lands, the islands of the Pacific.

The word 'Lapita' occurs quite often in this article. Lapita is an archaeological culture that appeared in the Bismarck archipelago of Northwest Melanesia around 1400 BC (Kirch 1997) (Figure 2). It is striking for its decorated, presumably ceremonial pottery with repeated patterns that often include what seem to be human faces. The patterns were produced with stamping tools, which themselves had to be carefully constructed.

We live at a time when interdisciplinary studies are greatly increasing in significance, and historical linguists have important contributions to make to the study of prehistory, alongside archaeology, archaeogenetics, physical anthropology and human ecology. It is for this reason that I would like to discuss the kinds of evidence that historical linguistics can offer to the study of prehistory. The article falls into three parts, each dealing with one of the ways in which the application of the comparative method to a language family can cast light on the histories of its speakers. First, a phylogeny can tell us something about the movements of its speakers from place to place if we read it carefully. Second, evidence of contact-induced change, carefully analysed, can tell us about the interactions of a language's speakers with groups speaking other languages. Third, reconstructed lexicon can tell us various things about the culture of speakers of that interstage language, revealing probable features of that culture that are not easily accessible to archaeology.

2 Reading a phylogeny

Before much is said about how reading a phylogeny can contribute to our understanding of prehistory, we need to look at where Oceanic languages came from.

2.1 The phylogeny of Austronesian

The higher-order phylogeny of Austronesian that most Austronesianist historical linguists accept is largely due to the work of Robert Blust, starting with a 1977 paper, and looks something like Figure 3. The conventions of this diagram are explained in §2.1.3. To read a phylogeny effectively, it needs to be related to a map, and this is provided as Figure 4.

Austronesian divides first into several Formosan subgroups, confined to Taiwan, and the Malayo-Polynesian subgroup that includes all Austronesian languages spoken outside Taiwan proper. The Malayo-Polynesian subgroup divides in turn into western Malayo-Polynesian languages and the Central-Eastern Malayo-Polynesian (CEMP) subgroup. Next the CEMP subgroup divides into the central Malayo-Polynesian languages and the Eastern Malayo-Polynesian subgroup. And finally the Eastern Malayo-Polynesian subgroup divides into the South-Halmahera West New Guinea and Oceanic subgroups. Oceanic is geographically widespread but it is a very well defined subgroup.

2.1.1 Subgrouping and the comparative method

What do I mean by 'a very well defined subgroup'? Before that question is answered, something needs to be said about what a subgroup is and about how a protolanguage like Proto Oceanic (POc) is reconstructed. Many readers will know this very well, but as historical linguistics has been taught in rather few universities worldwide in the last half-century, a brief explanation is needed.

A subgroup is a group of languages within a larger group. In particular, it is a group of languages that is descended from a single language. So Oceanic is a subgroup within the Austronesian family, and all its members are descended from a single Austronesian language that we call 'Proto Oceanic'. In a publication in 1937 Otto Dempwolff, one of the founders of Austronesian historical linguistics, demonstrated the existence of the Oceanic subgroup within Austronesian. How did he achieve this? He had published a reconstruction of Proto Malayo-Polynesian (PMP) three years earlier (Dempwolff 1934), and he recognised that a group of languages stretching from New Guinea eastwards into the Pacific all

showed the same set of sound changes (phonological innovations) relative to PMP. He inferred that this set of phonological innovations had occurred in a single language, which we now call ‘Proto Oceanic’, and which was the ancestor of all the languages that display these innovations.

The technique Dempwolff used was the one that historical linguists generally use when they work out phonological histories. It takes advantage of the fact that sound change is generally regular. In principle, all words with the same sound in the same environment undergo the same change in that word. First, words in different languages of the subgroup are tabulated (Table 1).³ Here, I have cheated a bit, because Oceanic languages with simpler phonological histories have been chosen. The items in blue are irregular, but to avoid descending into detail they are not discussed here. Figure 5 shows their locations.

Table 1: Cognate sets across selected Oceanic languages

	‘eye’	‘back’	‘father’	‘hand, five’	‘mosquito’	‘ear’
Tawala	<i>mata-</i>	<i>muli-</i>	<i>ama</i>	<i>nima-</i>	...	<i>taniga-</i>
Motu	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>ima-</i>	<i>namo</i>	<i>taia-</i>
Tolai	<i>mata-</i>	<i>mur-</i>	<i>tama-</i>	<i>lima-</i>	...	<i>taliŋa-</i>
Gela	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>lima-</i>	<i>namu</i>	<i>taliŋa-</i>
Kwaio	<i>mā-</i>	<i>huri-</i>	<i>maʔa</i>	<i>nima-</i>	<i>namu</i>	<i>aliŋa-</i>
Bauan	<i>mata-</i>	<i>muri</i>	<i>tama-</i>	<i>liŋa-</i>	<i>namu</i>	<i>daliŋa-</i>
Tongan	<i>mata</i>	<i>mui</i>	<i>tamai</i> ¹	<i>nima</i>	<i>namu</i>	<i>teliŋa</i>
Tahitian	<i>mata</i>	<i>muri</i>	<i>tama</i> ²	<i>rima</i>	<i>namu</i>	<i>tariʔa</i>
Rarotongan	<i>mata</i>	<i>muri</i>	<i>tama</i> ²	<i>rima</i>	<i>namu</i>	<i>tariŋa</i>
Maori	<i>mata</i>	<i>muri</i>	<i>tama</i> ²	<i>rima</i>	<i>namu</i>	<i>tariŋa</i>
Hawaiian	<i>maka</i>	<i>muli</i>	<i>kama</i> ²	<i>lima</i>

¹ also *tama* ‘child’. ² ‘child’.

The phoneme /m/ springs out immediately, and we can infer that *m must have been present in these words in POc (Table 2). Bauan Fijian displays an irregular form because it reflects the variant *limʷa- rather than POc *lima-.

Table 2: POc *m in cognate sets across selected Oceanic languages

	‘eye’	‘back’	‘father’	‘hand, five’	‘mosquito’	‘ear’
Tawala	<i>mata-</i>	<i>muli-</i>	<i>ama-</i>	<i>nima-</i>	...	<i>taniga-</i>
Motu	<i>mata-</i>	<i>muri</i>	<i>tama-</i>	<i>ima-</i>	<i>namo</i>	<i>taia-</i>
Tolai	<i>mata-</i>	<i>mur-</i>	<i>tama-</i>	<i>lima-</i>	...	<i>taliŋa-</i>
Gela	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>lima-</i>	<i>namu</i>	<i>taliŋa-</i>
Kwaio	<i>mā-</i>	<i>huri-</i>	<i>maʔa</i>	<i>nima-</i>	<i>namu</i>	<i>aliŋa-</i>
Bauan	<i>mata-</i>	<i>muri</i>	<i>tama-</i>	<i>liŋa-</i>	<i>namu</i>	<i>daliŋa-</i>
Tongan	<i>mata</i>	<i>mui</i>	<i>tamai</i>	<i>nima</i>	<i>namu</i>	<i>teliŋa</i>
Tahitian	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariʔa</i>
Rarotongan	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariŋa</i>
Maori	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariŋa</i>
Hawaiian	<i>mata</i>	<i>muli-</i>	<i>kama</i>	<i>lima</i>

³ A final hyphen on a table entry indicates that the word takes a pronominal suffix.

The phoneme /t/ is also clearly reflected, and we can also infer that *t must have been present in these words in POC. Notice that *t regularly becomes zero in Kwaio (Table 3). The irregularity in Bauan Fijian in Table 3 is only apparent. Geraghty (1983) recognised that the initial voiceless obstruent of nouns that were normally preceded by the determiner /na/ regularly became voiced, so *t- became *d-.

Table 3: POC *t in cognate sets across selected Oceanic languages

	‘eye’	‘back’	‘father’	‘hand, five’	‘mosquito’	‘ear’
Tawala	<i>mata-</i>	<i>muli-</i>	<i>ama-</i>	<i>nima-</i>	...	<i>taniga-</i>
Motu	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>ima-</i>	<i>namo</i>	<i>taia-</i>
Tolai	<i>mata-</i>	<i>mur-</i>	<i>tama-</i>	<i>lima-</i>	...	<i>taliŋa-</i>
Gela	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>lima-</i>	<i>namu</i>	<i>taliŋa-</i>
Kwaio	<i>mā-</i>	<i>huri-</i>	<i>maʔa</i>	<i>nima-</i>	<i>namu</i>	<i>aliŋa-</i>
Bauan	<i>mata-</i>	<i>muri</i>	<i>tama-</i>	<i>liŋa-</i>	<i>namu</i>	<i>daliŋa-</i>
Tongan	<i>mata</i>	<i>mui</i>	<i>tamai</i>	<i>nima</i>	<i>namu</i>	<i>teŋiŋa</i>
Tahitian	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariʔa</i>
Rarotongan	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariŋa-</i>
Maori	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariŋa-</i>
Hawaiian	<i>maka</i>	<i>muli</i>	<i>kama</i>	<i>lima</i>

When we tabulate the liquids, we see that there are two sets of correspondences and therefore two POC liquids that we reconstruct as *r and *l (Table 4). The Kwaio and Tongan words for ‘hand, five’ have initial /n-/ where we expect the same consonant as in the words for ‘ear’ (/n-/ in Tawala is regular). They reflect variant forms whose history stretches back to before POC.

Table 4: POC *l and *r in cognate sets across selected Oceanic languages

	‘eye’	‘back’	‘father’	‘hand, five’	‘mosquito’	‘ear’
Tawala	<i>mata-</i>	<i>muli-</i>	<i>ama-</i>	<i>nima-</i>	...	<i>taniga-</i>
Motu	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>ima-</i>	<i>namo</i>	<i>taia-</i>
Tolai	<i>mata-</i>	<i>mur-</i>	<i>tama-</i>	<i>lima-</i>	...	<i>taliŋa-</i>
Gela	<i>mata-</i>	<i>muri-</i>	<i>tama-</i>	<i>lima-</i>	<i>namu</i>	<i>taliŋa-</i>
Kwaio	<i>mā-</i>	<i>huri-</i>	<i>maʔa</i>	<i>nima-</i>	<i>namu</i>	<i>aliŋa-</i>
Bauan	<i>mata-</i>	<i>muri</i>	<i>tama-</i>	<i>liŋa-</i>	<i>namu</i>	<i>daliŋa-</i>
Tongan	<i>mata</i>	<i>mui</i>	<i>tamai</i>	<i>nima</i>	<i>namu</i>	<i>teŋiŋa</i>
Tahitian	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>tariʔa</i>
Rarotongan	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>taliŋa</i>
Maori	<i>mata</i>	<i>muri</i>	<i>tama</i>	<i>rima</i>	<i>namu</i>	<i>taliŋa</i>
Hawaiian	<i>maka</i>	<i>muli</i>	<i>kama</i>	<i>lima</i>

Filling in a few more gaps, we have a set of reconstructions:

POC	*mata-	*muri-	*tama-	*lima-	*ñamuk	*taliŋa-
	‘eye’	‘back’	‘father’	‘hand, five’	‘mosquito’	‘ear’

The initial *ñ- and final *-k of *ñamuk ‘mosquito’ are attested in Oceanic languages that are not represented in the tables. The reconstructions also give us a set of sound correspondences, just four of which are shown in Table 5.

Table 5: Sound correspondences attested in Tables 1–4

Proto Oceanic	*m	*t	*l	*r
Tawala	<i>m</i>	<i>t</i>	<i>n</i>	<i>l</i>
Motu	<i>m</i>	<i>t</i>	∅	<i>r</i>
Tolai	<i>m</i>	<i>t</i>	<i>l</i>	<i>r</i>
Gela	<i>m</i>	<i>t</i>	<i>l</i>	<i>r</i>
Kwaio	<i>m</i>	∅	-l-	<i>r</i>
Bauan	<i>m</i>	<i>t</i>	<i>l</i>	<i>r</i>
Tongan	<i>m</i>	<i>t</i>	-l-	∅
Tahitian	<i>m</i>	<i>t</i>	<i>r</i>	<i>r</i>
Rarotongan	<i>m</i>	<i>t</i>	<i>r</i>	<i>r</i>
Maori	<i>m</i>	<i>t</i>	<i>r</i>	<i>r</i>
Hawaiian	<i>m</i>	<i>k</i>	<i>l</i>	<i>l</i>

Notice that in the bottom four languages of Table 5, the POc liquids *r and *l have merged. This suggests that these languages form a subgroup within Oceanic. This subgroup is known as ‘Nuclear Polynesian’, and the merger took place in their shared ancestor, Proto Nuclear Polynesian.

In the same way, Dempwolff identified a number of innovations that had taken place in Proto Oceanic. So when I say that Oceanic is a ‘very well defined subgroup’, I mean that it is defined by a number of innovations that it does not share with other languages. For example, the PMP pairs *p, *b and *k, *g first merged and then split, giving the kind of crossover shown in the labial and velar examples in Table 6.

Table 6: Mergers and splits in Proto Oceanic labial and velar stops

PMP *panas	‘hot, warm’	POc *panas
PMP *punay	‘wild pigeon’	POc *bune
PMP *baqeRu	‘new’	POc *paqoRu
PMP *beRek	‘pig’	POc *boRok ‘domestic pig’
PMP *kuden	‘cooking pot’	POc *kuron
PMP *kabut	‘mist’	POc *gabu
PMP *gapgap	‘stammer’	POc *kaka(p)
PMP *gemgem	‘make a fist’	POc *gogo(m)

The examples in Table 7 show vowel and diphthong changes.

Table 7: Vowel and diphthong changes in Proto Oceanic

PMP *e [ə]	> POc *o	PMP *be(R)say ‘paddle’	POc *pose
PMP *-ahi, -ay	> POc *-e	PMP *babinahi ‘woman’	POc *papine
		PMP *qatay ‘liver’	POc *qate
PMP *-aw	> POc *-o	PMP *kasaw ‘rafter’	POc *kaso
PMP *-uy	> POc *-i	PMP *apuy ‘fire’	POc *api

These are just some of the phonological innovations that characterise POc. Not all shared innovations are phonological, however. We also find morphological and lexical innovations.

If a set of shared innovations points back to a common ancestor, like Proto Nuclear Polynesian or Proto Oceanic, then that common ancestor is represented as a node in the tree of that language family. So we get a tree in which innovations ABCD define the languages under one node, innovations EFGH another, and innovations JKLM a third. Each set of innovations defines a subgroup.

2.1.2 Trees and linkages

But what if the innovation sets overlap, and one group of languages is defined by innovations ABCD, another by innovations CDEF, and a third by ABEF (Figure 6)? We can say that the tree sometimes—and perhaps often—fails to model historical events. Related languages may form **linkages** rather than trees (Ross 1988, François 2014). There is nothing new about this idea. It has its roots in 19th-century dialectology, and has also been exploited by Pawley (1999), who writes about the probable role of linkages in the rapid spread of Malayo-Polynesian languages, and in Oceanic historical linguistics by scholars researching the languages of Fiji (Geraghty 1983), of the Caroline islands (Jackson 1983), of NW Melanesia (Ross 1988), of the SE Solomons (Lichtenberk 1988, 1994; Pawley 2011a) and of Vanuatu (Tryon 1976, Clark 1985, François 2011, 2014).

The tree model presupposes that there is a sharp divide between (a) shared innovations that define a subgroup because they reflect changes that occurred in the subgroup's shared ancestor, and (b) later shared innovations due to contact or to independent but parallel change. But there is plenty of work in European dialectology⁴ and in the variationist sociolinguistics championed by William Labov (Labov 2001, Kerswill 2003) to suggest that the division is unrealistic. Every innovation begins somewhere in a speech community and spreads across its speakers. In the tree model, innovations stop at the boundary of the speech community. If they pass beyond it, they are labelled as borrowing or contact-induced change. This is in part an artificial distinction.

If we abandon this distinction, we can instead say that overlapping subgroups form a linkage; that is, they arise as languages progressively differentiate out of a dialect network, so that innovations spread in various directions and at various times across the boundaries among emergent languages, resulting in configurations of innovations like that on the right of Figure 6 but often far more complex.

The distinction between a tree and a linkage is obviously not a sharp one. Some phylogenies, or parts of a phylogeny, are more tree-like, others more linkage-like, but this is a gradation that can't be captured in diagrams like those used here.

It is because relationships among languages can be quite complicated that I prefer the term 'phylogeny' over 'tree'. 'Tree' implies that language relationships always form branching nodes. 'Phylogeny', at least as it is used in genetics, also accommodates more complex relationships like those in a linkage.

2.1.3 More on the phylogeny of Austronesian

We need to take these complexities into account in reading a phylogeny, otherwise we will read the wrong history off it. In the Austronesian phylogeny in Figure 3, the Formosan and western Malayo-Polynesian languages are shown as grey blocks. This is intended to say that these are **not** subgroups—they are simply collections of languages—and that we do not know how many subgroups the block includes. One sometimes still reads in the literature that a certain language 'belongs to the Western Malayo-Polynesian subgroup', but it has been recognised for at least 20 years that western Malayo-Polynesian languages do not form a subgroup. Scholars disagree about how many Formosan subgroups there are,⁵ and at least the higher-order subgrouping of western Malayo-Polynesian languages remains unknown. Each language that is shown in roman script in Figure 3, like Proto Oceanic, has undergone a set of shared innovations that is inherited by its daughter-languages, in other words, its daughter-languages form a subgroup.

Proto CEMP is shown with a dashed vertical line. This says that it was a language that broke up into a dialect network, and the horizontal dashed line indicates that the central Malayo-Polynesian languages emerged from a part of that dialect network. The little fan at the right-hand end of the line says that the Central Malayo-Polynesian languages are a linkage. I mentioned above that POC is defined by a large set

⁴ A recent example is Stiles (2013) on Germanic.

⁵ Ross (2012) provides a survey. So fraught are the issues that I would no longer defend the position I took there.

of innovations. Proto CEMP is defined by far fewer innovations. There has been argument about this. Blust (1982, 1983-84, 1993) argued for the unity of CEMP. Donohue & Grimes (2008) argued that the CEMP node should be removed because some of the innovations on which it was based also occur in some western Malayo-Polynesian languages, but Blust (2009) responded that innovations unique to CEMP remain. CEMP lies outside my specialist area, but reading these publications carefully, I think there **is** evidence for Proto CEMP, but it is much less than the evidence for POc. Again, the diagrammatic conventions used here cannot capture this.

The point here, however, is that one needs to know what innovations define each node in order to read a phylogeny as an account of history. We can say that the Proto CEMP speech community lasted for a shorter time than the POc speech community. Ideally, this should be confirmed by correlation with archaeological evidence. The POc speech community is usually equated with the earliest communities of the Lapita archaeological culture in the Bismarck Archipelago (Figure 7), and this implies perhaps a 300-year period for the development of POc before its break-up. Unfortunately, we have no archaeological evidence that would allow us to figure out how long the PCEMP community lasted, but one may guess at a century or less.

The phylogeny thus implies that after the break-up of Proto Malayo-Polynesian, Austronesian speakers moved through the Philippines and Indonesia as far as the Bismarck archipelago rather quickly with no lengthy pause. The first major pause was in the Bismarcks, and led to the emergence of POc and of the Lapita culture. The archaeology roughly confirms this. It suggests that the first settlers from Taiwan arrived in the northern Philippines around 2000 BC. We don't know how long it was before they moved further south, but if we infer that they spoke PMP, which is quite well defined, it was perhaps sometime around 1800 BC. Austronesian speakers had reached the Bismarcks by, and perhaps before, 1400 BC.

2.1.4 The phylogeny of the Oceanic subgroup

Now, what lies beyond the Oceanic node at the bottom right of Figure 3? Figure 8 shows the Oceanic phylogeny as it appears in vols 3 & 4 of *The lexicon of Proto Oceanic* (Ross, Pawley & Osmond 2008, 2011). The first striking feature of this phylogeny is how different it appears in overall shape from the Austronesian phylogeny in Figure 3. The latter displays recursive branching, that is, branches off branches, matching its progress through the Philippines and Indonesia. The Oceanic phylogeny looks more like a rake, with nine first-order branches, two of which, Yapese and Mussau & Tench, consist of just one or two modern languages. Andrew Pawley (pers. comm.) points out that the grey blocks in Figure 3 perhaps also represent partial rakes: Proto Austronesian itself broke up into the Formosan subgroups plus PMP, and PMP broke up into the numerous western MP groups plus PCEMP.

The rake-like structure of Oceanic is nonetheless unexpected. One would expect the Oceanic phylogeny to display the same kind of recursive pattern as the phylogeny in Figure 3, and indeed a recursive tree was used in the first two volumes of *The lexicon of Proto Oceanic*. The editors changed their minds in volume 3 because they realised that this phylogeny was based on recursion that they **expected**, not on what the **evidence** was telling them. Oceanist linguists have tried again and again to find more recursions and larger subgroups within Oceanic, but none of the evidence for such groupings stands up. The most promising is Blust's (1998) proposal that all Oceanic languages outside the Admiralties form a single group because they have merged POc *s and *c, but a subgroup based on a single innovation is rather shaky if the innovation is likely to have happened independently in different languages. In any case, this proposal would bifurcate Oceanic into Admiralty and non-Admiralty primary subgroups, and the latter would still form an eight-branch rake.

As for the unexpected rake, its very unexpectedness must be telling us something. It suggests that early Oceanic must have dispersed very quickly, leaving no time for interstage languages to undergo change as they had to some extent done in earlier stages of Austronesian. It happens that this inference matches the archaeology. Lapita emerged in the Bismarcks with or after the arrival of speakers of the language that was to become POc, and archaeologists have focussed on dating the spread of this culture

into the Pacific. The fact that, from the Solomons eastward, Lapita people were entering empty territory apparently makes the interpretation of the earliest archaeology fairly straightforward. Figure 9 corresponds to the rake. As we might expect, there is more of a patchwork in the west, where Oceanic originated. As Figure 10 shows, Lapita appeared at about 1400 BC. By 1000 BC bearers of Lapita were in Santa Cruz and Vanuatu, and maybe 50 years later in New Caledonia and Fiji. They reached Tonga by 900 BC, Samoa by 750 BC.

The presence of obsidian (volcanic glass) from the Bismarcks in places as far away as Fiji and New Caledonia in the early settlement sites and the similarity of pottery styles suggests that people remained in contact in the early years, and it was after that that the innovations specific to each branch of Oceanic occurred. The distance in a straight line between two of the earliest settlements, Talasea on New Britain and Santa Cruz Island is 1800 km. The distance from Talasea to the south of New Caledonia is about 2700 km, from Talasea to Fiji 3300 km. A real sea journey would be several hundred kilometres longer. These are huge distances for even the most well built ocean-going outrigger canoe, and it is not surprising that languages quickly diverged as contact between the old and the new communities was reduced.

Another striking feature of the Oceanic phylogeny in Figure 8 is the number of linkages in it. In fact almost everything that looks like possible recursion entails linkages. The Southern Oceanic and Central Pacific linkages are clearly the outcomes of migrations, but their nature is odd. We cannot find a convincing set of innovations relative to POc that defines either a Proto Southern Oceanic or a Proto Central Pacific. Geraghty (1983, 1996) wrestled with the latter, and Lynch (2000, 2004) with the former. In both cases the conclusion is that there are patterns of overlapping innovations that link the languages within each grouping together, but no innovations that mark them off clearly in relation to POc. This suggests the possibility that each was actually the outcome of more than one Oceanic arrival, perhaps at different times, speaking different early Oceanic dialects. This is certainly suggested by genetic studies, particularly in the case of Central Pacific, which find multiple populations entering the region. In both regions the case can be made that the first arrivals were Polynesian in appearance, and later arrivals had a considerable admixture acquired in the New Guinea region (Blust 2008, Xu et al. 2012, Forster & Renfrew 2014, Valentin et al. 2016).

Returning to Western Oceanic, its heartland is located within the area where we suppose POc itself was spoken (Pawley 2003, 2008), and it is a reasonable inference that it reflects part of the POc dialect network that simply stayed in place. Western Oceanic falls into the three large groupings shown in Figure 8, namely the North New Guinea and Meso-Melanesian linkages and the Papuan Tip subgroup and indicated by the shading on Figure 13. The phylogeny in Figure 12 shows the internal subgrouping of Western Oceanic. One of the three groupings, Papuan Tip (in the middle), is evidently descended from a single language, Proto Papuan Tip, since its languages reflect a set of shared innovations. North New Guinea and Meso-Melanesian are clearly linkages. In each, the languages are linked by a complex web of innovations, but, unlike Papuan Tip, no innovations characterise the grouping as a whole. At the same time, Figure 12 shows that North New Guinea and Meso-Melanesian each include defined subgroups. Each of these subgroups had as its parent a single language (shown in red) that broke away from the rest of the linkage.

If we plot on a map these subgroups and the rumps of the North New Guinea and Meso-Melanesian linkages after the subgroups are subtracted, as in Figure 13, the map tells an interesting story. The triangles show early Lapita sites. The sites in Mussau predate Western Oceanic and reflect the earliest spread of Lapita, as the phylogeny shows.⁶ The linkages together form a single area in New Britain and New Ireland, reflecting the early Western Oceanic linkage (in orange), and the subgroups (in purple) are distributed around it. Each subgroup is descended from a single language that has put some distance between itself and the ancestral Western Oceanic linkage, and we see a pattern that corresponds quite

⁶ One would expect Lapita in the Admiralties also to reflect the earliest spread, but relevant evidence is limited, perhaps because much of it is under metres of volcanic ash.

nicely with the archaeology of Lapita settlement in the region. The early Western Oceanic linkage corresponds with the area of early Lapita sites, and the subgroups reflect departures from that area. So we see that a careful construction of a phylogeny by the comparative method and a correlation of the phylogeny with geography can provide insights into the prehistoric movements of people.

3 Evidence from contact-induced change

I commented above that later arrivals in the Central Pacific and Southern Oceanic areas had a considerable admixture of genes from the New Guinea region. This brings us to our second topic, evidence of contact-induced change, which can tell us about the interactions of speakers with other groups and occasionally about instances of language shift.

We are confronted with the fact that most of the Austronesian-speaking region that was inhabited before the Austronesian spread today speaks only Austronesian languages. New Guinea, however, is dominated by languages of various Papuan families (Figure 1). What does this tell us about the population histories of these regions?

To the west of New Guinea there is disagreement about whether the Austronesian diaspora was largely powered by agriculture, as the archaeologist Peter Bellwood (1984, 1995, 2002) envisages, or by trade (Bulbeck 2008, Donohue & Denham 2010). It obviously involved both, probably in varying degrees at different times and places. But the linguistic effects would be different. A search for land implies the establishment of new Austronesian speaking communities. Trade doesn't, but it may imply the presence of a trade language. Can linguistics cast some light here? I think it can.

The Austronesian languages fall into the three broad typological regions shown in Figure 14. They coincide only partially with the nodes in the phylogeny of Figure 3. The eastern region corresponds with the CEMP node, but the northern and western regions have no exact equivalents in the phylogeny.

The northern region is typologically conservative and reflects the voice and applicative system of reconstructed early Austronesian. In the Philippines, early MP speakers encountered low-density populations of hunter-gatherers, the so-called 'Negritos'. Today's 30 or so Negrito languages are Austronesian and some are mostly closely related to the languages of their immediate agricultural neighbours.⁷ An obvious hypothesis is that Negritos long ago entered into symbiotic relationships with their neighbours and, after a period of bilingualism, lost their earlier languages but retained some specialised cultural vocabulary. There is no obvious evidence of contact-induced change in Philippine Austronesian languages. Linguistically the Philippine situation points clearly to the establishment of Austronesian speaking agricultural communities in territory previously occupied by hunter-gatherers.

By contrast, the Austronesian languages of the western and eastern regions display substantial typological change. In the western region, we find various transformations of the early Austronesian grammatical type. We can often recognise chunks of the earlier system, but they have been rearranged in various ways (Ross 2002a, 2002b). The situation in the eastern region is similar: typological change has taken place, but here less of the early Austronesian grammatical system survives. Nonetheless, much early Austronesian morphology is preserved in the restructured systems. This kind of restructuring of a grammatical system is usually attributed to contact, so we may reasonably infer that the western and eastern regions already had substantial populations that interacted with incoming Austronesian speakers, and 'substantial populations' means agriculture. Archaeology increasingly supports this inference.

The morphosyntactic **differences** between western and eastern Austronesian languages suggest that the pre-Austronesian languages of the two regions were different, possibly because the regions had been affected by separate Neolithic agricultural revolutions. The pre-Austronesian agriculture of the western region evidently came from mainland SE Asia (Bellwood 2004, Donohue & Denham 2010), the agriculture of the eastern region from New Guinea (Wright et al. 2013).

⁷ See Reid (2013) for a survey and further references.

Two areas of the eastern region, north Halmahera and Timor, Alor & Pantar (hardly visible on the map), remain Papuan speaking, and various scholars have argued that Wallacea (the part of the eastern region that lies west of New Guinea) is a linguistic area—a hotbed of contact (Klamer et al. 2008; Schapper 2015).

It is tempting at this point to say, OK, the typological changes in Austronesian languages took place as a result of contact, and to leave things there. But this tells us little about the social history of speakers.

There are three or four kinds of contact-induced change that we can diagnose from their outcomes (Ross 2013, 2014):

- 1) **Bilingually induced change**: children grow up speaking two languages. One is their heritage language, the other the language of a neighbouring, perhaps numerically dominant, population. Over generations they gradually adapt the structures of their heritage language to those of the other language, and metatypy or typological change occurs (Ross 2007).
- 2) **Language shift**: a community abandons their heritage language for another language. This seems to happen in two ways. In **normal language shift**, children first grow up speaking two languages, then over generations lose their heritage language. Because the language to which they are shifting has become one of their native languages, the shift leaves few clues except cultural lexicon retained from the old language (Ross 2014). On the other hand, in **catastrophic language shift** a group of adults is forced to suddenly adopt a new language. They learn it incompletely, and pass on this incompletely learned language to their children, who, for whatever reason, have no access to the ‘complete’ version of the new language. Obviously, this requires a rather rare set of circumstances, and usually results in simplification and creolisation.
- 3) **Language mixing** is rare, and also occurs in two varieties. In cases like Media Lengua, spoken in Ecuador with a Quechua grammatical system and Spanish vocabulary, one language emerges out of two. Media Lengua seems to have emerged out of its speakers’ perceptions of their own identity (Muysken 1997). In other cases, the verbal system and its morphology and lexicon are drawn from one language, the nominal from another. The most famous cases are probably Copper Island Aleut (Golovko & Vakhtin 1990, Golovko 1994) and Michif (Bakker 1994, 1997). Such languages have also emerged recently in remote Aboriginal communities in Australia, and are apparently the result of parents and carers using a mixed code when they speak to their children (McConvell & Meakins 2005, O’Shannessy 2005).

On this classification, the Negrito languages of the Philippines represent normal language shift. They are ordinary Philippine languages, but Reid (1994) shows that they have retained cultural lexicon.

On the other hand, it is pretty clear that many languages in the western region and the vast majority in the eastern regions reflect bilingually induced change. They display structural change. Most languages are too complex to reflect creolisation and they don’t display evidence of language mixing. In other words, in the western and eastern regions, structural change took place because Austronesian speakers were strongly bilingual in a non-Austronesian language and their children were growing up speaking both languages and adapting the structure of their Austronesian language to that of their non-Austronesian language.

For this to have happened, two things have to be true. First, there were Austronesian speaking communities. In other words, the incoming Austronesian speakers comprised more than a few traders, an inference supported in recent years by genetic evidence for a human lineage stretching from Taiwan across Island SE Asia and New Guinea and eastward to Polynesia (Delfin et al. 2012, Xu et al. 2012, Cox 2013, Forster & Renfrew 2014). Second, there were substantial pre-existent populations who interacted with arriving Austronesians. Unlike in the Philippines, they must have been agricultural communities.

But if this is the case, why have apparently so many pre-Austronesian languages been replaced by Austronesian? We must infer, I think, that Austronesian speakers established agricultural village communities and a cultural framework into which they were able to recruit their non-Austronesian speaking neighbours. Reconstructing the Slavic expansion in Europe, Nichols’ (1998:240–241)

hypothesises that ‘a strong and explicit sense of ethnic identity as manifested in language helps the language of one agricultural society spread to other agricultural societies of the same cultural level’. This was achieved partly by intermarriage and partly perhaps by involving one’s neighbours in ritual obligations, as Thurston (1996:200–201) describes for the Bariai of NW New Britain, where Oceanic languages are in the final stages of expanding at the expense of Papuan languages. Nichols suggests that intermarriage with neighbouring ethnic groups is a natural outcome when a society based on clan exogamy finds itself stranded where some clans are no longer represented. These factors would favour shift from pre-Austronesian languages to Austronesian.

4 Lexical reconstruction

Since the mid-1990s a team at the Australian National University has been working on the reconstruction of Proto Oceanic lexicon. Four volumes of *The lexicon of Proto Oceanic* have been published, a fifth will appear soon (Ross, Pawley & Osmond 1998, 2003, 2008, 2011, 2016), and a sixth and seventh are planned.

The brainchild of Andrew Pawley, we call our approach ‘terminological reconstruction’. We begin with the assumption that the culture of present-day Oceanic speakers who have retained their village lifestyle is sufficiently similar to the culture of speakers of POc for us to use their terminologies as a guide for reconstructing POc. We realise of course, that this approach would not work, for example, in IE, where so much cultural change has occurred that scholars still argue about what the culture of PIE speakers was like (Heggarty 2013).

The approach is two-pronged. First, we use the terminologies of present-day speakers of Oceanic languages as the basis for constructing a hypothesis about the semantic structure of a corresponding POc terminology, taking account of descriptions of the lifestyles of Oceanic communities and the geographical and physical resources of particular regions of Oceania. Second, we search for cognate sets from which forms can be reconstructed to match each meaning in our hypothesised terminology. The search is not restricted to Oceanic languages; if a term found in an Oceanic language proves to have non-Oceanic Austronesian cognates, this confirms its POc antiquity and often provides additional evidence about its meaning.⁸

For example, by comparing terms across languages for outrigger canoes and parts of an outrigger canoe, one can see which concepts recur and so are likely to have been present in POc. We can readily reconstruct quite a detailed terminology for the parts of an outrigger canoe (Pawley & Pawley 1998) (Figure 15). The POc term for a canoe was *waga, and this is well supported across Oceania, as well as in the non-Oceanic languages of Wallacea. Below is just a small sample of reflexes from scattered Oceanic languages.

⁸ A major source of non-Oceanic cognates is Robert Blust’s online *Austronesian Comparative Dictionary* (Blust and Trussell, on going). Four collections of Oceanic reconstructions with supporting cognate sets have also been very useful, namely Clark (2009) for north and central Vanuatu, Lynch (2001) for south Vanuatu, Bender et al. (2003) for Micronesian and Clark & Biggs (2006) for Polynesian.

*POc *waga 'large sailing canoe; (generic) canoe'*

Admiralties:	Wuvulu	<i>wa</i>	'canoe'
N New Guinea:	Gedaged	<i>wag</i>	'large canoe that goes out on the high seas, has one or two masts and a large platform, ship or boat'
North New Guinea:	Gitua	<i>waga</i>	'canoe'
Papuan Tip:	Dobu	<i>waga</i>	'sailing canoe'
North-Central Vanuatu:	V'ënen	<i>na-uak</i>	'canoe'
	Taut		
New Caledonian:	Nyelâyu	<i>wanga-</i>	'canoe'
Micronesian:	Kiribatese	<i>wā</i>	'canoe'
Micronesian:	Marshallese	<i>wa</i>	'canoe'
Fijian:	Bauan	<i>waga</i>	'canoe'
Polynesian:	Tongan	<i>vaka</i>	'boat (generic)'
Polynesian:	Hawaiian	<i>waʔa</i>	'canoe'

The term *saman for 'outrigger' is also well supported.

*POc *saman 'outrigger float'*

Admiralties:	Seimat	<i>cam</i>	'outrigger float'
Admiralties:	Mussau	<i>samana</i>	'outrigger float'
North New Guinea:	Gedaged	<i>sam</i>	'outrigger float'
Meso-Melanesian:	Nehan	<i>haman</i>	'outrigger boom'
North-Central Vanuatu	Paamese	<i>a-sem</i>	'outrigger'
Micronesian:	Kiribatese	<i>rama</i>	'outrigger float'
Micronesian:	Marshallese	<i>tam</i>	'outrigger float'
Fijian:	Bauan	<i>ḏama</i>	'outrigger float, smaller hull of double canoe'
Polynesian:	Tongan	<i>hama</i>	'outrigger float, smaller hull of double canoe'
Polynesian:	Tikopia	<i>ama</i>	'outrigger'

The reconstructive method is the one described in §2.1.1, and Pawley & Pawley's reconstructions for these and many more terms are found in the first volume of *The lexicon of Proto Oceanic*. Note the additional meaning of *saman in Bauan Fijian and in Tongan, which we return to just below.

A third and final step in reconstructing a terminology is to see if it needs modification in the light of the reconstructions. In particular, there are cases where we have been unable to reconstruct a term where we had believed we should be able to. Thus some scholars have assumed that the Lapita people used double canoes like the Fijian *"drua*, that is, canoes with two hulls, to cross the vast distances of the Pacific, as double canoes are widespread in the Austronesian speaking area. So we might expect a term for it to occur in POc, but there is none. We can only reconstruct a term convincingly in Proto Central Pacific, namely *paqurua 'double canoe', from POc *paqus 'bind' and *rua 'two'. Traditional double canoes in Fiji have a larger and a smaller hull, implying that the double canoe developed by replacing the outrigger with a second dugout. Linguistic evidence supports this, since the Fijian and Tongan term for an outrigger is also used for the smaller hull of a double canoe, as shown above. In fact, when one puts together the available information, it is highly probable that the early Lapita expansion used ocean-going single-hull outrigger canoes like those recently still occasionally constructed in Melanesia, and that the double canoe was invented in the Fiji/Tonga/Samoa area and was then used to overcome the huge distances of eastern Polynesia (Blust 1999:82).

Another unexpected POc gap occurs in the semantic domain of root crops, which are among the main staple foods in Oceania. They include varieties of taro, yam and sweet potato, listed here with relevant POc reconstructions and their PMP origins.

*talo(s)	taro <i>Colocasia esculenta</i>	< PMP *tales
*m ^w apo(q)	taro <i>Colocasia esculenta</i>	< a Papuan source
*piRaq	giant taro <i>Alocasia macrorrhizos</i>	< PMP *biRaq
*qupi	greater yam <i>Dioscorea alata</i>	< PMP *qubi
	lesser yam <i>Dioscorea esculenta</i>	
	sweet potato <i>Ipomoea batatas</i>	
	American/Chinese taro <i>Xanthosoma sagittifolium</i>	

We know that the two last items on the list, sweet potatoes and American taro reached Melanesia from South America. But we did expect to be able to reconstruct terms for the five taro and yam varieties without much difficulty. Terms for taro are indeed readily reconstructed, but interestingly the PMP term for ‘taro’ (*Colocasia esculenta*), *tales, has been replaced in much of NW Melanesia, by *m^wapo(q), fairly clearly a borrowing from a Papuan language.⁹ This implies, perhaps, that Lapita agriculture was considerably influenced by the already ancient taro agriculture of the New Guinea mainland.

Two main species of yam are grown in New Guinea, the greater yam (*Dioscorea alata*) and the lesser yam (*Dioscorea esculenta*). Today, the greater yam in places has considerable ceremonial significance,¹⁰ but the lesser yam is a far more widespread food crop. Surprisingly, no POc term for lesser yam can be reconstructed.

On the other hand two Western Oceanic terms can be reconstructed, *mamisa and *kamisa. The fact that this pair seemingly entails the prefixes *ka- and *ma- immediately suggests that they reflect an earlier stative verb (such pairs are common in POc), perhaps one meaning ‘sweet’, as POc has the term *mamis ‘sweet’. The inference must be that the lesser yam did not grow in Lapita gardens, but grew wild in the rainforest and was domesticated somewhat later somewhere in NW Melanesia, perhaps not so long after the break-up of POc (Ross 2008).

Turning to animals, no POc term for ‘dog’ can be reconstructed, even though dogs are ubiquitous in Oceanic villages (Osmond & Pawley 2011). The Bismarcks lie in the realm of marsupial mammals like the wallaby, and it is clear that placental mammals like dogs, pigs and rats were introduced by human beings. However, there are no dog bones in Lapita archaeological sites, and POc speakers evidently didn’t have dogs. Dog bones do appear in sites across parts of Oceania between one and two thousand years ago, but in Vanuatu and New Caledonia no evidence of prehistoric dogs has been found at all (Lisa Matisoo-Smith, pers. comm.).

Of course, canoes, root crops and dogs are all material items, theoretically accessible to archaeology—although no Lapita-age canoe has even been found. However, terminological reconstruction can also give insights into POc categorisations of the world to which archaeology has little or no access, and some of these have the potential to tell us about Lapita social structure. Central to this are the terminologies of kinship and chieftainship, which are planned for vol. 6 of *The lexicon of Proto Oceanic*.

However, Ross & Osmond (2016) look at POc terms for human age cohorts, by which I mean terms equivalent to the English sequence *baby, toddler, child, teenager, adult*. POc speakers, however, evidently classified age cohorts rather differently:

⁹ See the short discussion in Ross (2008:265–266). Hays (2005) collects taro terms from the western half of New Guinea, and POc *m^wapo(q) was evidently borrowed from a member of Hays’ ‘mao’ set of taro terms. The set extends well into the eastern half of New Guinea (<http://transnewguinea.org/word/taro>, consulted 25 September 2015).

¹⁰ See the discussion in Ross (2008:256–258) for references.

*m ^w eRa	‘young person from birth to puberty’
*m ^w eRa-m ^w eRa	‘baby, very young child’ (diminutive of *m ^w eRa)
*tau paqoRu	‘marriageable young person’ (person + new)
*tam ^w ata	‘fully grown adult; (generic) person’
*matuqa	‘mature, full-grown, ripe, old (person), still vigorous’
*m ^w arap (V)	‘grow old’; (N) ‘old person’

POc lacks a sequence resembling *baby, toddler, child*. Instead they are all *m^weRa, although compounds involving a word meaning ‘little’ occur in many languages for small children. Instead, the important thresholds are puberty and marriage. There is a default assumption that a fully grown adult is married. The term *tam^wata ‘fully grown adult’ was also used as a life-form generic for ‘person’. Later adulthood falls into two stages, *matuqa, still economically useful, and *m^warap, no longer vigorous.

Terminological reconstruction also allows us to see how POc speakers categorised the natural world. Pawley (2011b) examines the reflexes of the POc term *ikan, usually glossed ‘fish’, and concludes that it was polysemous. In one sense, it denoted typical fish, but it was also a high-level generic denoting typical fish, sharks, rays, eels, cetaceans (whales, dolphins, porpoises), dugongs, turtles, crocodiles and possibly octopus and squid. But it didn’t include all sea creatures: some belonged apparently to the category *sisiq, which perhaps included all edible gastropods. Pawley also concludes that *manuk, often glossed ‘bird’, was a high-level generic for birds, bats and flying insects.

A careful study by Evans (2008) examines how plants are categorised in a sample of Oceanic languages, then reconstructs a POc ethnobotanical classification. POc evidently had no word for ‘plants’ in general. Instead it seems to have had five major life-form taxa:

*kayu	Tree or shrub, generic name for plants with woody stems and branches, probably not including palms or tree-ferns
*waRoc	General term for vines and creepers, plants with creeping or climbing growth structure
*pali[sj]i	Generic term for grasses and possibly also sedges and other grass-like plants
*limut or *lumut	Term for mosses, algae and seaweed
*taliŋa	(Which otherwise means ‘ear’) generic term for mushrooms and fleshy fungi (Evans 2008:111).

Her more detailed ethnobotanical classification is shown in Figure 16. Notably ‘bamboo’ and ‘pandanus’ are also top-level generics, presumably because they serve so many functions in Oceanic societies. Why the fan palm is a top-level generic, though, I do not know.

5 Conclusion

I have tried to show, admittedly without providing as much detail as I might like, that historical linguistics, carefully conducted, can make unique contributions to the interdisciplinary study of human prehistory in at least three respects:

- A phylogeny can provide information about movements of speakers.
- Contact-induced change can tell us about interactions between speaker groups, and therefore about the speaker groups themselves.
- Reconstructed lexicon can give us insight into the lifestyle and ways of thinking of its users.

I have left untouched the question of how one integrates linguistic findings with those of archaeology, genetics and archaeogenetics, physical anthropology and human ecology. Perhaps the most important comment here is that each discipline should be allowed to pursue its study using the methods in which its

practitioners are skilled, without claims being made from outside the discipline that its findings are 'wrong' because they don't concur with those of a sister discipline. Mismatches in results provide new research questions, and time and again, patience and careful study have eventually led to a cogent interdisciplinary synthesis.

References

- Bakker, Peter. 1994. Michif, the Cree-French mixed language of the Métis buffalo hunters in Canada. In *Mixed languages: 15 case studies in language intertwining*, ed. by Peter Bakker & Maarten Mous, 13–33. Studies in Language and Language Use 13. Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT).
- Bakker, Peter. 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press.
- Bellwood, Peter. 1984. A hypothesis for Austronesian origins. *Asian Perspectives* 26:109-117.
- Bellwood, Peter. 1995. Austronesian prehistory in Southeast Asia: Homeland, expansion and transformation. In *The Austronesians: historical and comparative perspectives*, ed. by Peter S. Bellwood, James J. Fox, and Darrell Tryon, 103–118. Canberra: Pacific Linguistics.
- Bellwood, Peter. 2002. Farmers, foragers, languages, genes: The genesis of agricultural societies. In *Examining the farming/language dispersal hypothesis*, ed. by Peter Bellwood and Colin Renfrew, 17–28. Cambridge: McDonald Institute for Archaeological Research.
- Bellwood, Peter. 2004. The origins and dispersals of agricultural communities in Southeast Asia. In *Southeast Asia: From prehistory to history*, ed. by Ian Glover and Peter Bellwood, 21–40. London: RoutledgeCurzon.
- Bender, Byron W., Ward H. Goodenough, Frederick Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Homin Sohn, Stephen Trussel, and Judith W. Wang. 2003. Proto-Micronesian Reconstructions--1. *Oceanic Linguistics* 42:1-110.
- Blust, Robert A. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawaii Working Papers in Linguistics* 9/2:1-15.
- Blust, Robert A. 1982. The linguistic value of the Wallace Line. *Bijdragen tot de Taal-, Land- en Volkenkunde* 138:231–250.
- Blust, Robert A. 1983–84. More on the position of the languages of Eastern Indonesia. *Oceanic Linguistics* 22-23:1–28.
- Blust, Robert A. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–293.
- Blust, Robert A. 1998. A note on higher-order subgroups in Oceanic. *Oceanic Linguistics* 37:182–188.
- Blust, Robert A. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica 1. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert. 2008. Remote Melanesia: One history or two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47:445–459.
- Blust, Robert A. 2009. The position of the languages of Eastern Indonesia: A reply to Donohue and Grimes. *Oceanic Linguistics* 48:36–77.
- Blust, Robert A., and Stephen Trussell. ongoing. *Austronesian comparative dictionary*. Honolulu: Department of Linguistics, University of Hawai'i. <http://www.trussel2.com/ACD/>

- 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.
- Clark, Ross. 1985. Languages of north and central Vanuatu: groups, chains, clusters and waves. In *Austronesian linguistics at the 15th Pacific Science Congress*, ed. by Andrew Pawley and Lois Carrington, 199–236. Pacific Linguistics C-88. Canberra: Pacific Linguistics.
- Clark, Ross. 2009. **Leo tuai: A comparative lexical study of North and Central Vanuatu languages*. Pacific Linguistics 603. Canberra: Pacific Linguistics.
- Clark, Ross, and Bruce G. Biggs. 2006. POLLEX: Polynesian lexicon. Electronic files, Department of Anthropology, University of Auckland. <http://pollex.org.nz/>.
- Cox, Murray P. 2013. Southeast Asian islands and Oceania: human genetics. In *The encyclopedia of global human migration*, ed. by Immanuel Ness, ch.37. Oxford: Blackwell.
- Delfin, Frederick, Sean Myles, Ying Choi, David Hughes, Robert Illek, Mannis van Oven, Brigitte Pakendorf, Manfred Kayser, and Mark Stoneking. 2012. Bridging Near and Remote Oceania: MtDNA and NRY Variation in the Solomon Islands. *Molecular Biology and Evolution* 29:545-564.
- Dempwolff, Otto. 1937. *Vergleichende Lautlehre des Austronesischen Wortschatzes* (Comparative phonology of the Austronesian lexicon). Berlin: Dietrich Reimer.
- 1: *Induktiver Aufbau einer indonesischen Ursprache* (Inductive reconstruction of a Malayo-Polynesian protolanguage), 1934. (Beihefte zur Zeitschrift für Eingeborenen-Sprachen 15)
- 2: *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen* (Deductive application of Proto Malayo-Polynesian to individual Austronesian languages), 1937. (Beihefte zur Zeitschrift für Eingeborenen-Sprachen 17)
- Donohue, Mark, and Tim Denham. 2010. Farming and language in Island Southeast Asia: Reframing Austronesian history. *Current Anthropology* 51:223–256.
- Donohue, Mark, and Charles E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia. *Oceanic Linguistics* 47:114-158.
- Evans, Bethwyn. 2008. Ethnobotanical classification. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 3: Plants*, ed. by Malcolm Ross, Andrew Pawley & Meredith Osmond, 53–84. Canberra: Pacific Linguistics
- Forster, Peter, and Colin Renfrew. 2014. The Pacific: DNA. In *The Cambridge world prehistory, 1: Africa, South and Southeast Asia and the Pacific*, ed. by Colin Renfrew and Paul Bahn, 561–565. Cambridge: Cambridge University Press.
- François, Alexandre. 2011. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1:175–246.
- François, Alexandre. 2014. Trees, waves and linkages: Models of language diversification. In *The Routledge handbook of historical linguistics*, ed. by Claire Bowerman and Bethwyn Evans, 161–189. London: Routledge.
- Geraghty, Paul. 1983. *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawaii Press.
- Geraghty, Paul. 1996. Problems with Proto Central Pacific. In *Oceanic studies: proceedings of the First International Conference on Oceanic Linguistics*, ed. by John Lynch and Fa'afu Pat, 83–91. Pacific Linguistics C-133. Canberra: Pacific Linguistics.
- Golovko, Eugenij V. 1994. Mednyj Aleut or Copper Island Aleut: an Aleut-Russian mixed language. In *Mixed languages: 15 case studies in language intertwining*, ed. by Peter Bakker and Maarten Mous,

- 113–121. *Studies in Language and Language Use* 13. Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT).
- Golovko, Eugenij V., and Nikolai B. Vakhtin. 1990. Aleut in contact: the CIA enigma. *Acta Linguistica Hafniensia* 22:97–125.
- Greenhill, Simon J., and Russell D. Gray. 2009. Austronesian language phylogenies: myths and misconceptions about Bayesian computational methods. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. by Alexander Adelaar and Andrew Pawley, 375–397. Pacific Linguistics 601. Canberra: Pacific Linguistics.
- Grimm, Jacob. 1848. *Geschichte der deutschen Sprache* (History of the German language). Leipzig: Hirzel.
- Hays, Terence E. 2005. Vernacular names for tuber in Irian Jaya: implications for agricultural prehistory. In *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, ed. by Andrew Pawley, Robert Attenborough, Jack Golson, and Robin Hide, 625–670. Pacific Linguistics 572. Canberra: Pacific Linguistics.
- Heggarty, Paul. 2013. Europe and western Asia: Indo-European linguistic history. In *The encyclopedia of global human migration*, ch.19, ed. by Immanuel Ness. Oxford: Blackwell.
- Jackson, Frederick H. 1983. The internal and external relationships of the Trukic languages of Micronesia. PhD dissertation, University of Hawai‘i.
- Kerswill, Paul. 2003. Dialect levelling and geographical diffusion in British English. In *Social dialectology. In honour of Peter Trudgill*, ed. by David Britain and Jenny Cheshire, 223–243. Amsterdam: John Benjamins.
- Kirch, Patrick V. 1997. *The Lapita peoples: Ancestors of the Oceanic world*. Oxford: Blackwell.
- Klamer, Marian, Ger P. Reesink, and Mirjam van Staden. 2008. East Nusantara as a linguistic area. In *From linguistic areas to areal linguistics*, ed. by Pieter Muysken, 95–149. Amsterdam: John Benjamins.
- Labov, William. 2001. *Principles of linguistic change, 2: Social factors*. Oxford: Blackwell.
- Lichtenberk, Frantisek. 1988. The Cristobal-Malaitan subgroup of Southeast Solomonian. *Oceanic Linguistics* 27:24–62.
- Lichtenberk, Frantisek. 1994. Reconstructing heterogeneity. *Oceanic Linguistics* 33:1–35.
- Lynch, John. 2000. Linguistic subgrouping in Vanuatu and New Caledonia. In *SICOL: Proceedings of the Second International Conference on Oceanic Linguistics, 2: Historical and descriptive studies*, ed. by Bill Palmer and Paul Geraghty, 155–184. Pacific Linguistics 505. Canberra: Pacific Linguistics.
- Lynch, John. 2001. *The linguistic history of southern Vanuatu*. Canberra: Pacific Linguistics.
- Lynch, John. 2004. The Efate-Erromango problem in Vanuatu subgrouping. *Oceanic Linguistics* 43:311–338.
- McConvell, Patrick, and Felicity Meakins. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25:9–30.
- Muysken, Peter. 1997. Media Lengua. In *Contact languages: A wider perspective*, ed. by Sarah Grey Thomason, 365–426. Amsterdam: John Benjamins.
- Nichols, Johanna. 1998. The Eurasian spread zone and the Indo-European dispersal. In *Archaeology and language, 2: Correlating archaeological and linguistic hypotheses*, ed. by Roger M. Blench and Matthew Spriggs, 220–266. London: Routledge.
- O'Shannessy, Carmel. 2005. Light Warlpiri: A new language. *Australian Journal of Linguistics* 25:31–57.

- Osmond, Meredith, and Andrew Pawley. 2011. Mammals, reptiles, amphibians. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 4: Animals*, ed. by Malcolm Ross, Andrew Pawley, and Meredith Osmond, 217–269. Canberra: Pacific Linguistics.
- Pawley, Andrew. 1999. Chasing rainbows: Implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 95–138. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica 1. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Pawley, Andrew. 2003. Locating Proto Oceanic. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 2: The physical world*, ed. by Malcolm Ross, Andrew Pawley, and Meredith Osmond, 17–34. Canberra: Pacific Linguistics.
- Pawley, Andrew. 2008. Where and when was Proto Oceanic spoken? Linguistic and archaeological evidence. In *Language and text in the Austronesian world: Studies in honour of Ülo Sirk*, ed. by Yury A. Lander and Alexander K. Ogoblin, 47–71. Studies in Austronesian Linguistics 06. Munich: Lincom Europa.
- Pawley, Andrew. 2011a. On the position of Bugotu and Gela in the Guadalcanal-Nggelic subgroup of Oceanic. *Oceanic Linguistics* 50:1–24.
- Pawley, Andrew. 2011b. Were turtles fish in Proto Oceanic? Semantic reconstruction and change in some terms for animal categories in Oceanic languages. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 4: Animals*, ed. by Malcolm Ross, Andrew Pawley, and Meredith Osmond, 421–452. Canberra: Pacific Linguistics.
- Pawley, Andrew, and Medina Pawley. 1998. Canoes and seafaring. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 1: Material culture*, ed. by Malcolm Ross, Andrew Pawley, and Meredith Osmond, 173–209. Canberra: Pacific Linguistics.
- Rask, Rasmus Kristian. 1818. *Undersøgelse om det gamle Nordiske eller Islandske Sprogs Oprindelse* (Essay on the origin of the ancient Norse or Icelandic tongue). Copenhagen: Gyldendal.
- Reid, Lawrence A. 1994. Possible non-Austronesian lexical elements in Philippine Negrito languages. *Oceanic Linguistics* 33:37–72.
- Reid, Lawrence A. 2013. Who are the Philippine Negritos? Evidence from language. *Human Biology* 85: Article 15.
- Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Pacific Linguistics C-98. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2002a. The history and transitivity of western Austronesian voice and voice-marking. In *The history and typology of western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross, 17–62. Pacific Linguistics 518. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2002b. Final words: research themes in the history and typology of western Austronesian languages. In *The history and typology of western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross, 2002:451–474. Pacific Linguistics 518. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2007. Calquing and metatypy. *Journal of Language Contact: Thema* 1:116–143.
- Ross, Malcolm. 2008. Staple foods: root crops, bananas, breadfruit and sago. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 3: Plants*, ed. by Malcolm Ross, Andrew Pawley, and Meredith Osmond, 255–292. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2012. In defense of Nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13:1253–1330.
- Ross, Malcolm. 2013. Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact* 6:5–47.

- Ross, Malcolm. 2014. Reconstructing the history of languages in northwest New Britain: Inheritance and contact. *Journal of Historical Linguistics* 4:84–132.
- Ross, Malcolm, and Meredith Osmond. 2016. People: gender, age cohorts and marital status. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society 5: People: Body and mind*, ed. by Malcolm Ross, Andrew Pawley, and Meredith Osmond, 37-73. Canberra: Asia-Pacific Linguistics.
- Ross, Malcolm, Andrew Pawley & Meredith Osmond, eds. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*.
1: *Material culture*, 1998, Canberra: Pacific Linguistics.
2: *The physical world*, 2003, Canberra: Pacific Linguistics.
3: *Plants*, 2008, Canberra: Pacific Linguistics.
4: *Animals*, 2011, Canberra: Pacific Linguistics.
5: *Peope: Body and mind*, 2016, Canberra: Asia-Pacific Linguistics.
- Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel* 90:99-151.
- Schleicher, August. 1861. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen* (Compendium of the comparative grammar of the Indo-European languages). Weimar: Hermann Böhlau.
- Stiles, Patrick V. 2013. The pan-West Germanic Isoglosses and the subrelationships of West Germanic to other branches. *NOWELE* 66:5–38.
- Thurston, William R. 1996. Amara: An Austronesian language of northwestern New Britain. In *Studies in languages of New Britain and New Ireland, 1: Austronesian languages of the North New Guinea cluster in northwestern New Britain*, ed. by Malcolm Ross, 197–248. Pacific Linguistics C-135. Canberra: Pacific Linguistics.
- Tryon, Darrell. 1976. *New Hebrides languages: An internal classification*. Pacific Linguistics C-50. Canberra: Pacific Linguistics.
- Valentin, Frédérique, Florent Détroit, Matthew Spriggs, and Stuart Bedford. 2016. Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: Implications for Remote Oceanic settlement and Lapita origins. *Proceedings of the National Academy of Sciences of the USA* 113:292-297.
- Wouk, Fay, and Malcolm Ross, eds. 2002. *The history and typology of western Austronesian voice systems*. Pacific Linguistics 518. Canberra: Pacific Linguistics.
- Wright, Duncan, Tim Denham, Denis Shine, and Mark Donohue. 2013. An archaeological review of western New Guinea. *Journal of World Prehistory* 29:326–358.
- Xu, Shuhua, Irina Pugach, Mark Stoneking, Manfred Kayser, Li Jin, and the HUGO Pan-Asian SNP Consortium. 2012. Genetic dating indicates that the Asian–Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proceedings of the National Academy of Sciences* 109:4574–4579.

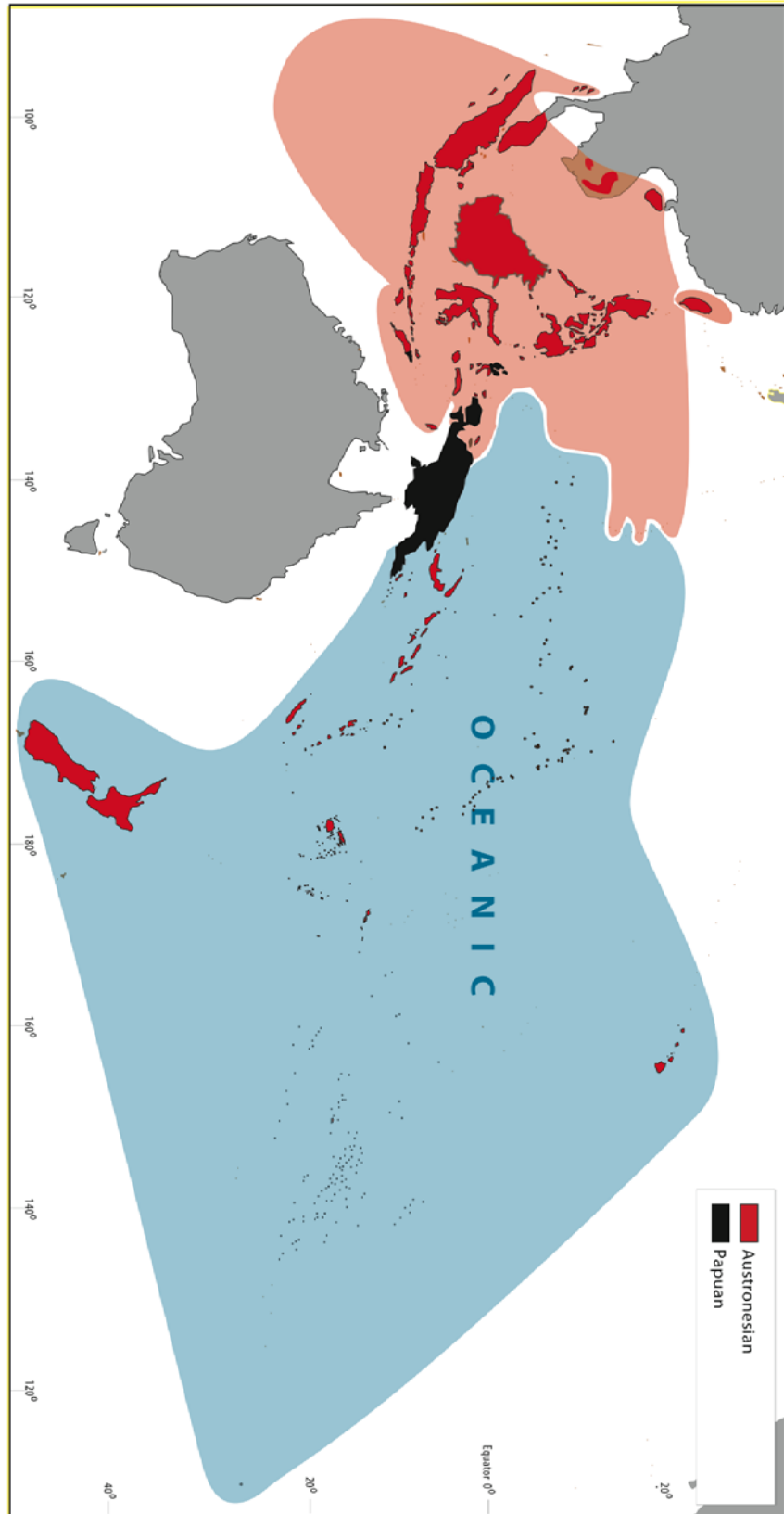


Figure 1: The Oceanic subgroup of Austronesian: 500 out of 1200 languages

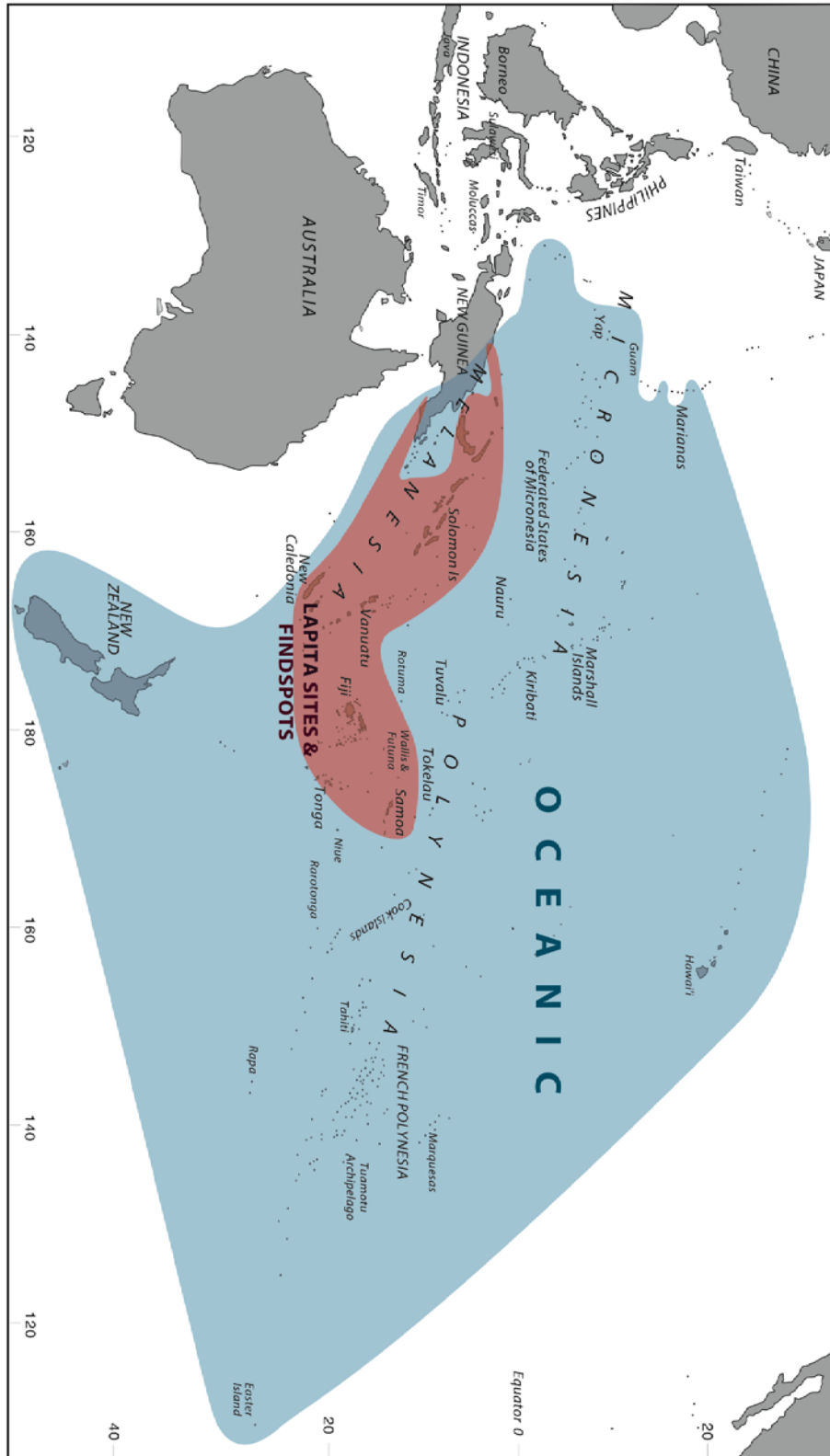


Figure 2: Lapita archaeological sites

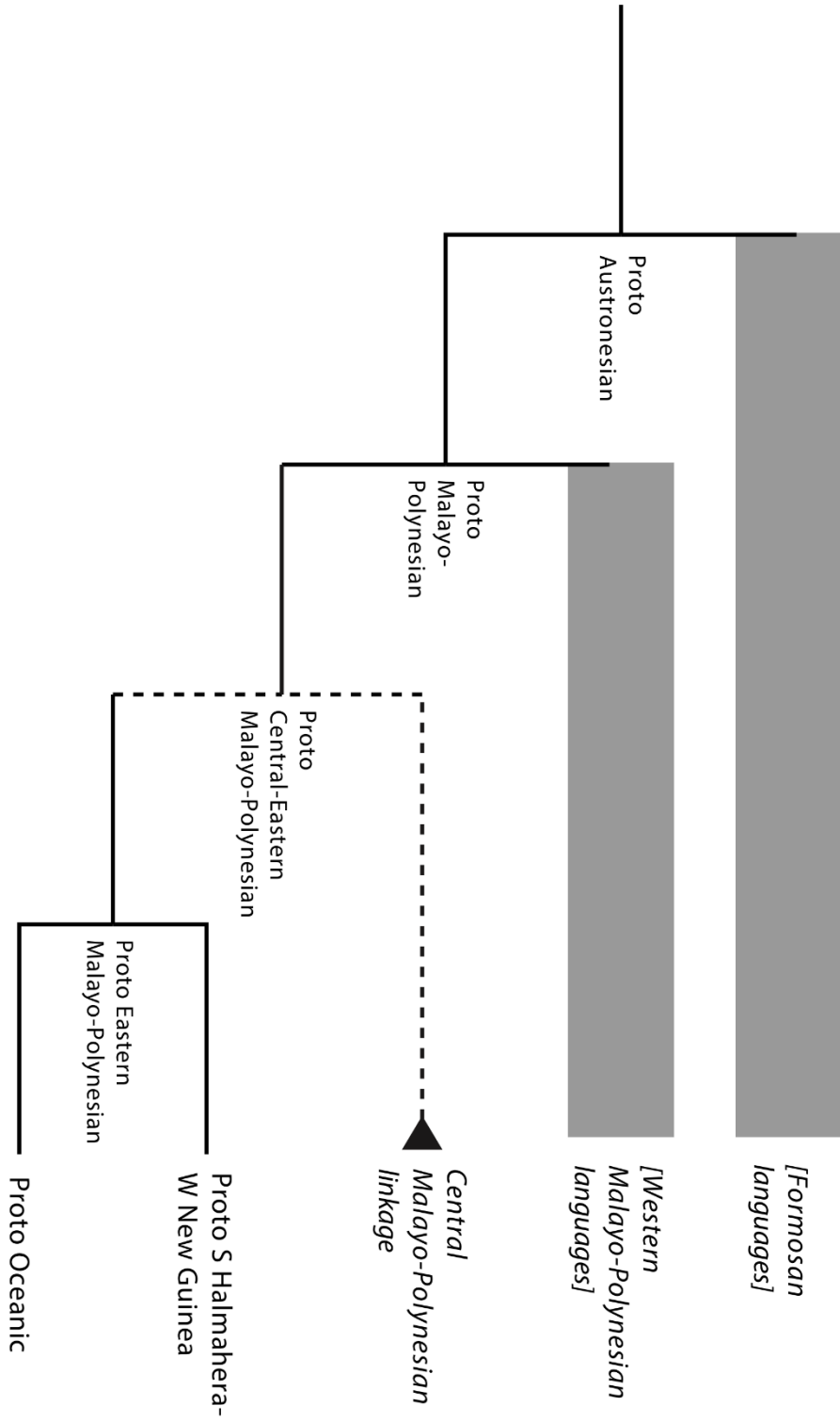


Figure 3: Austronesian phylogeny

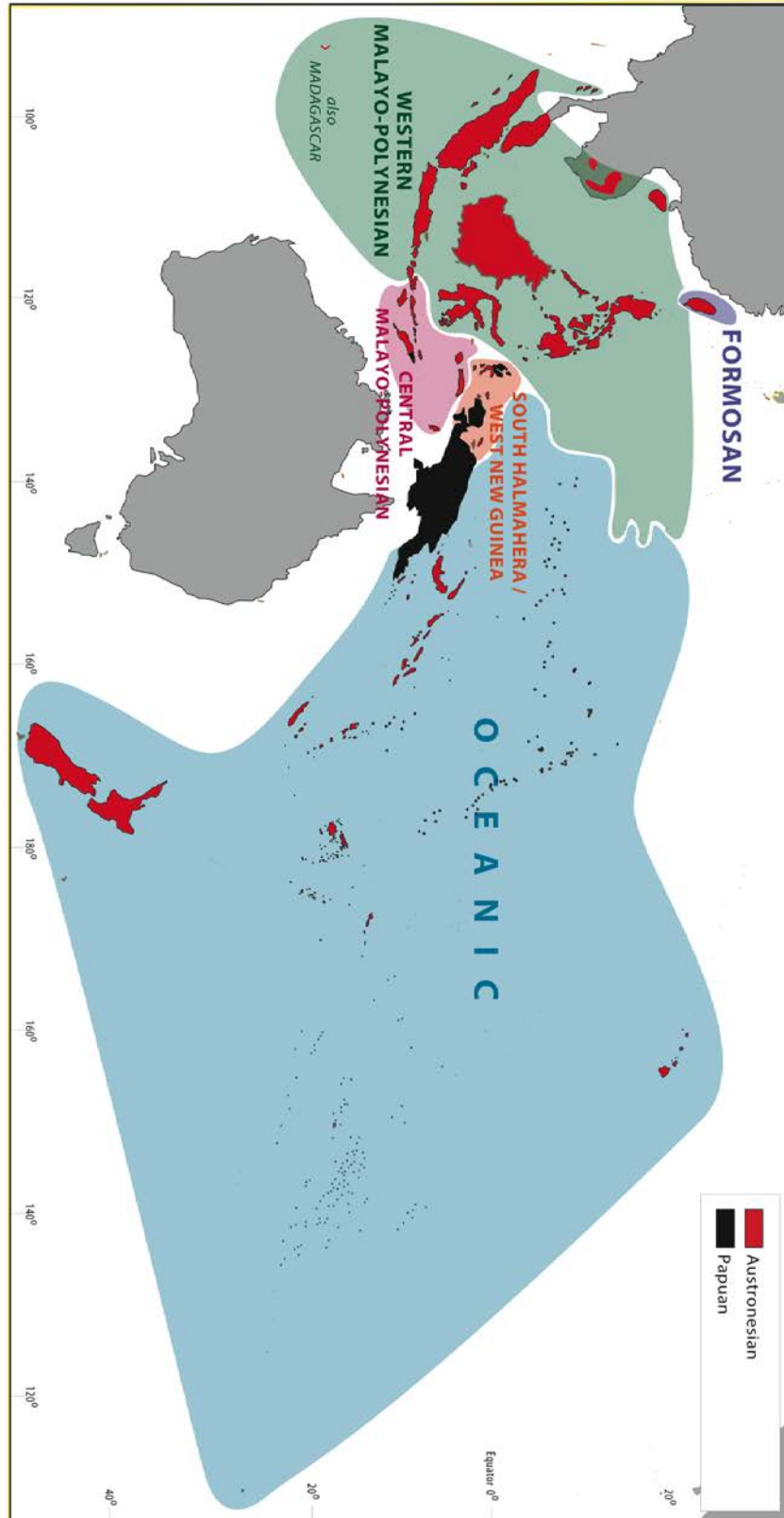


Figure 4: Austronesian groupings

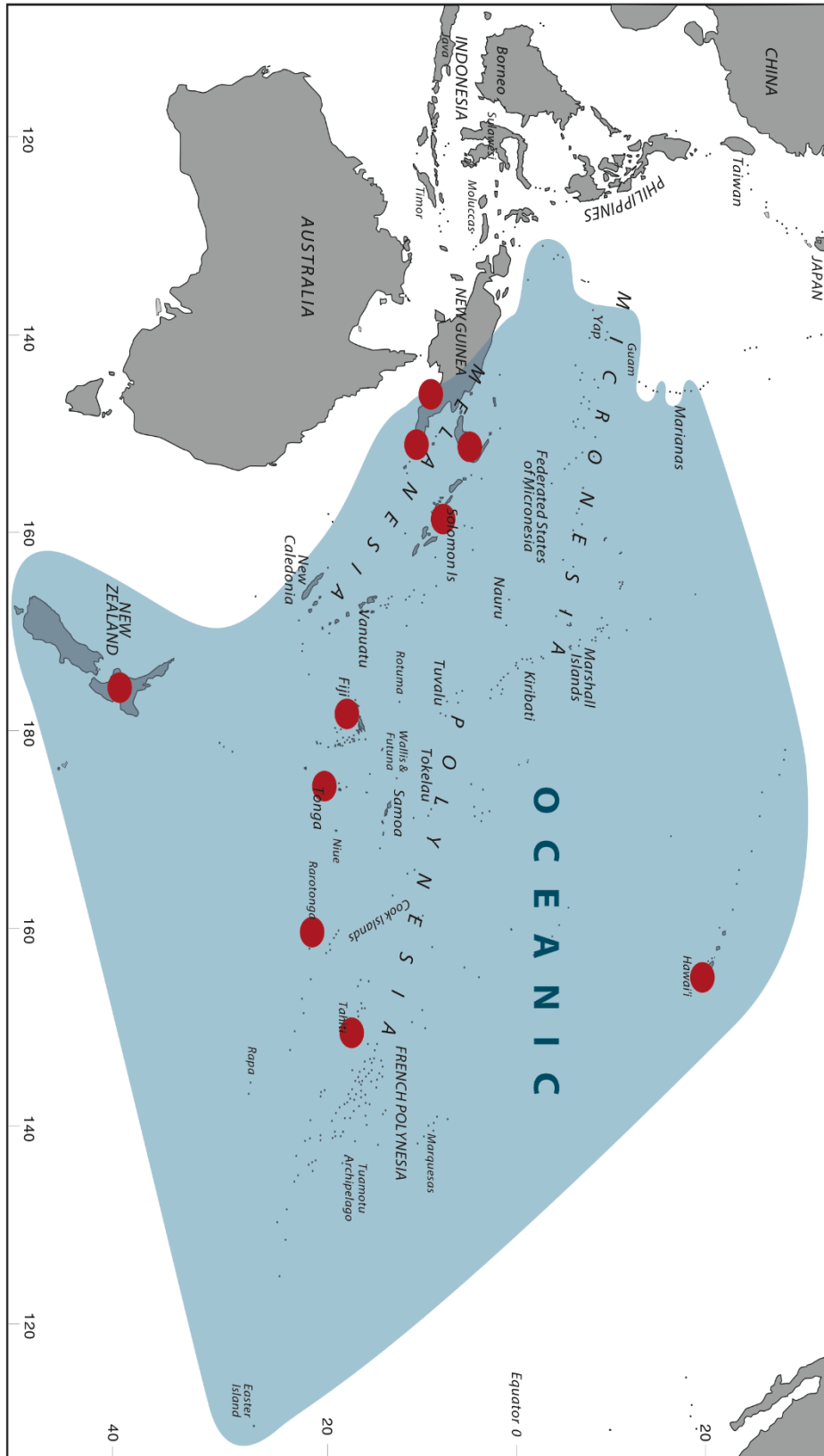


Figure 5: Locations of the languages in Tables 1–4

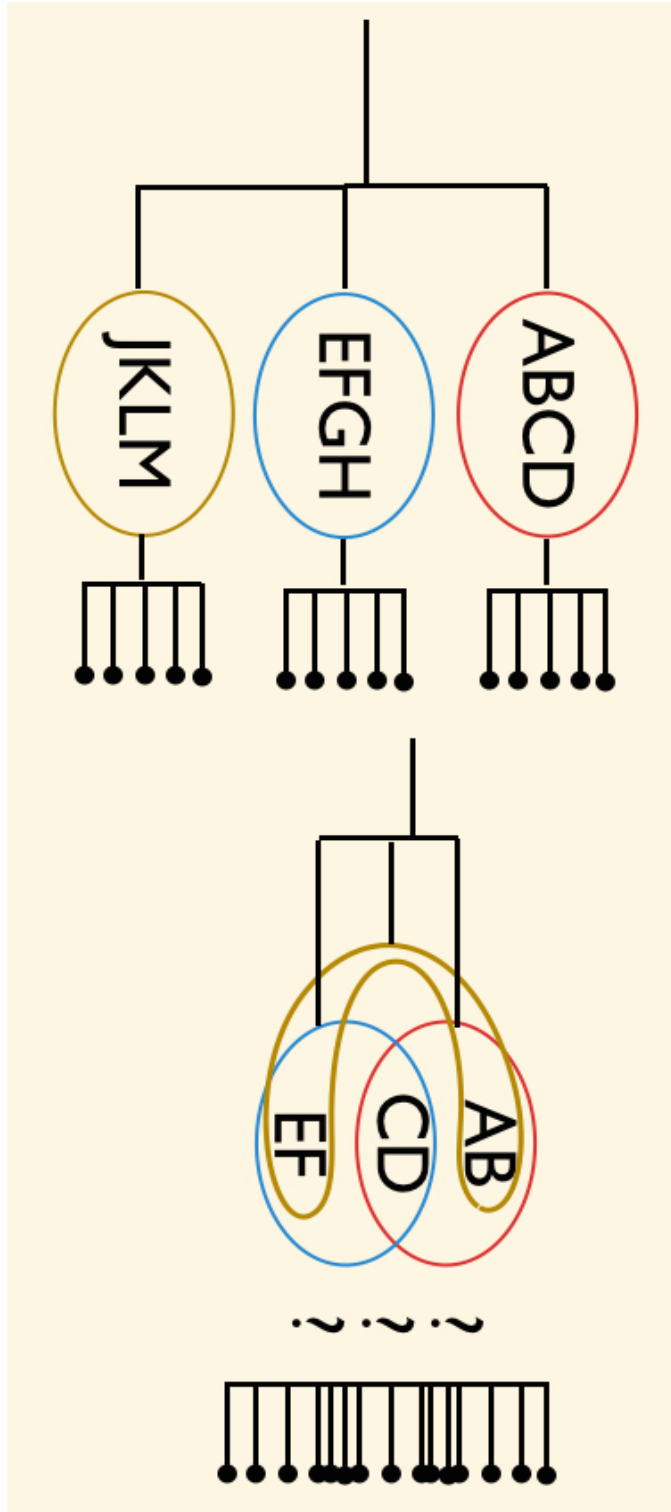


Figure 6: Schematic diagram of a tree and a linkage

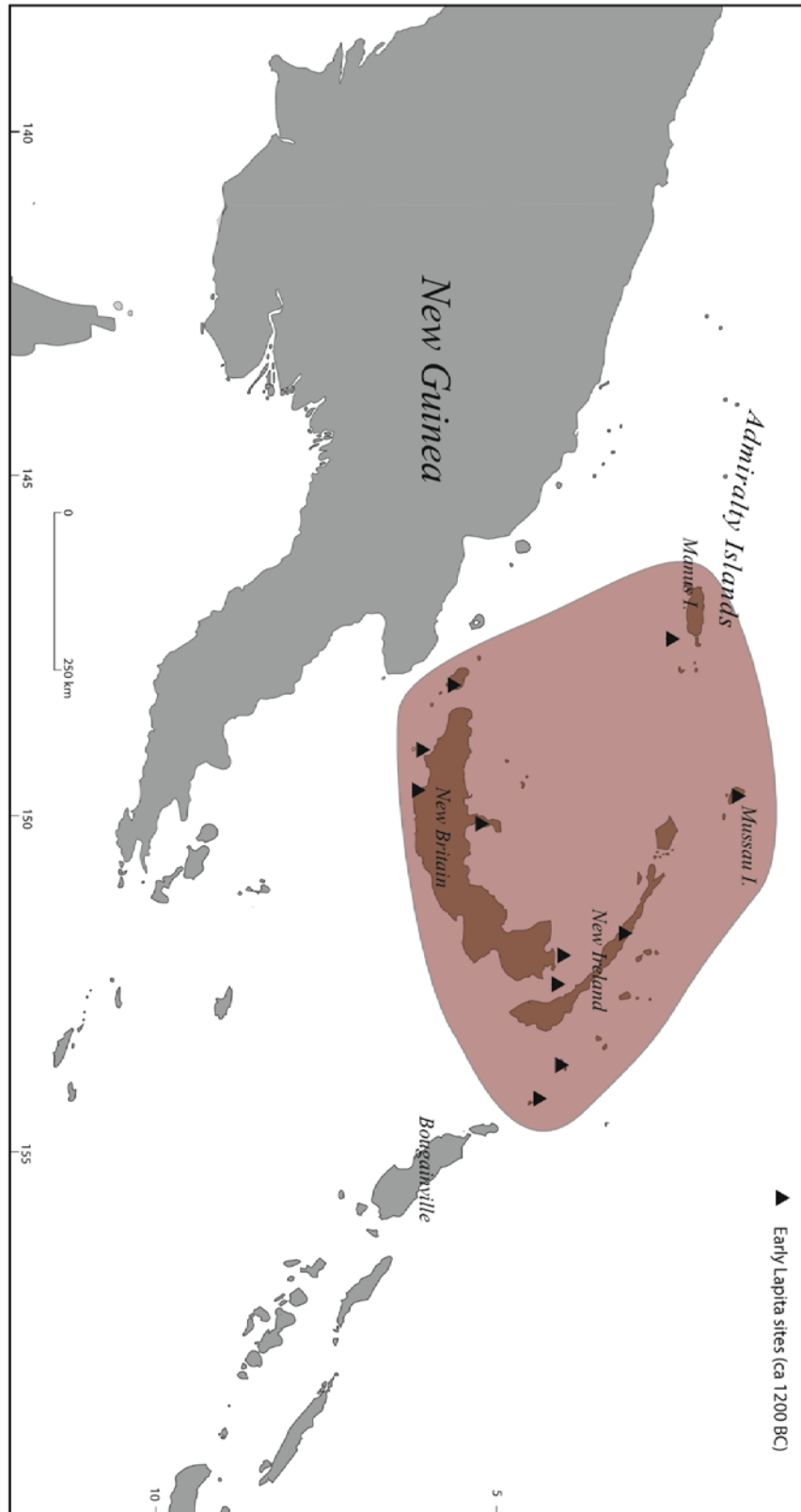


Figure 7: Early Lapita sites

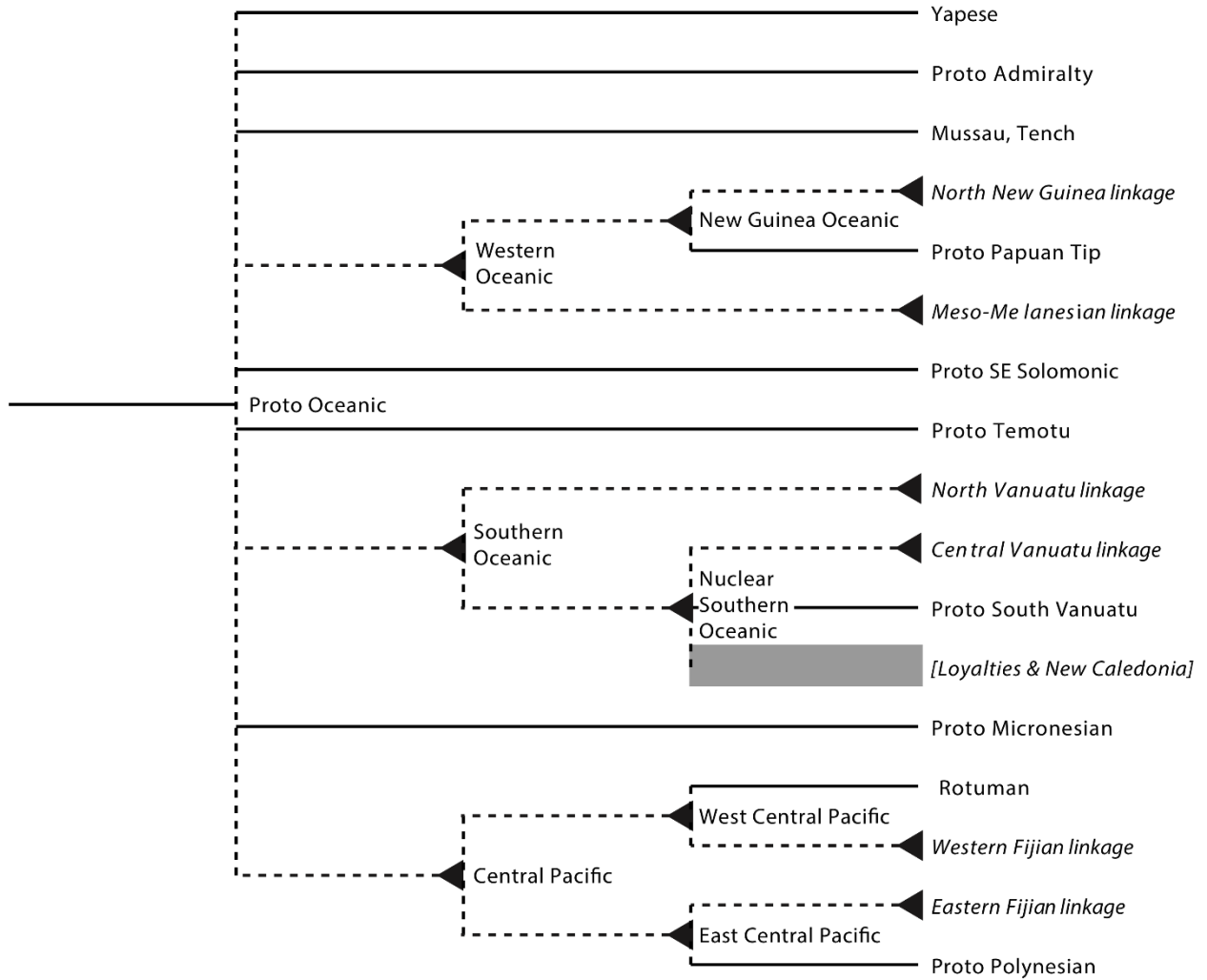


Figure 8: Oceanic phylogeny

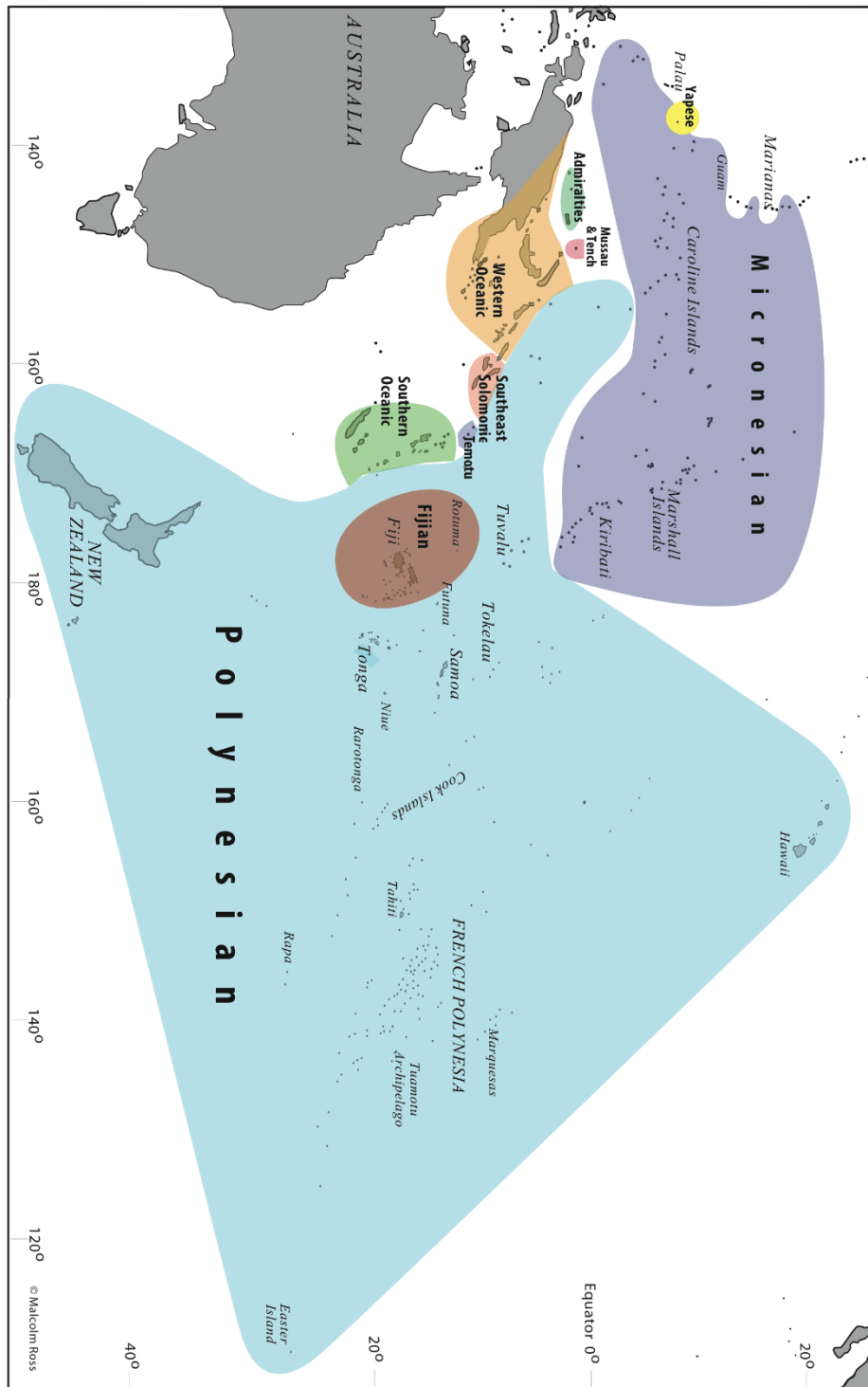


Figure 9: Oceanic groupings

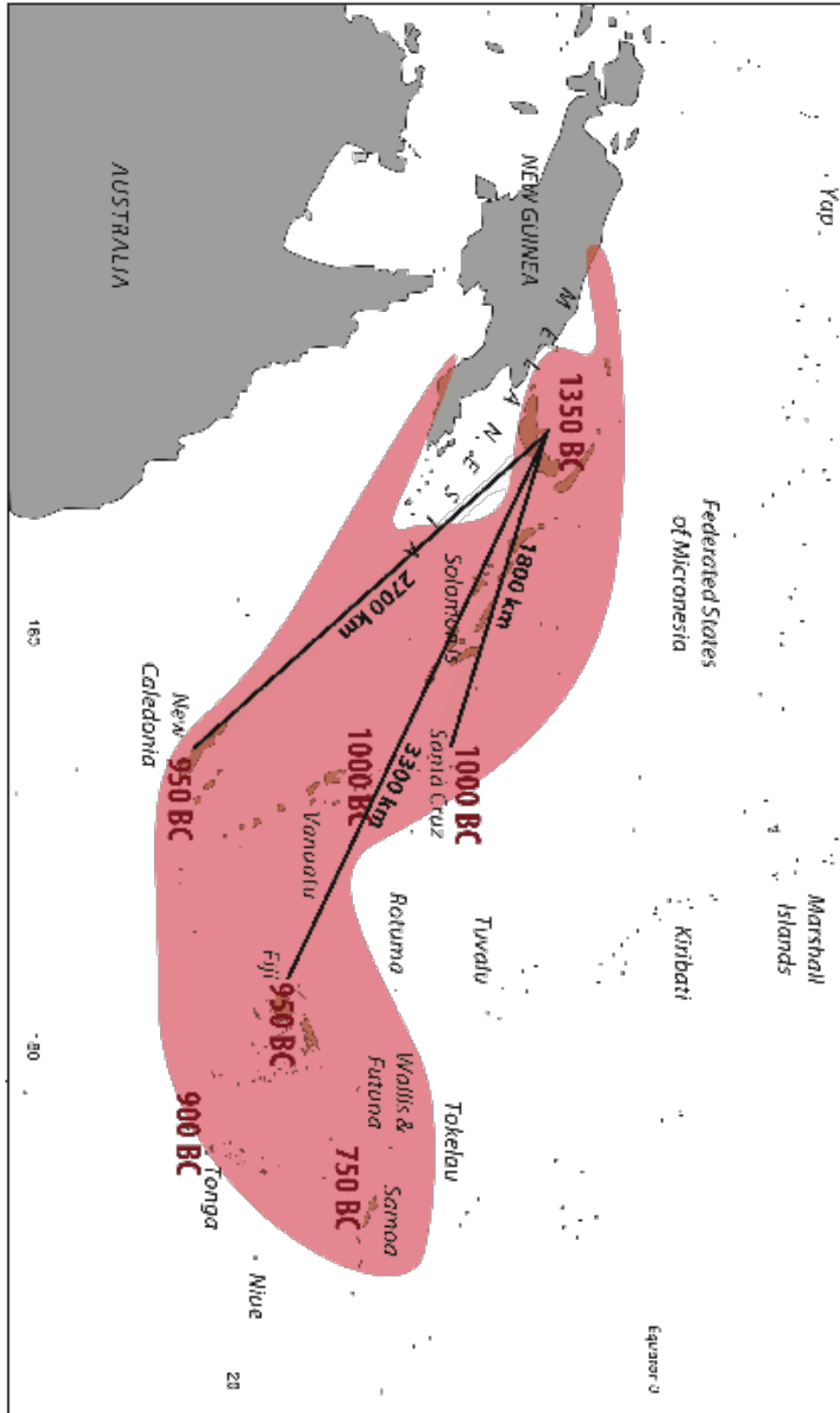


Figure 10: Lapita dates and distances

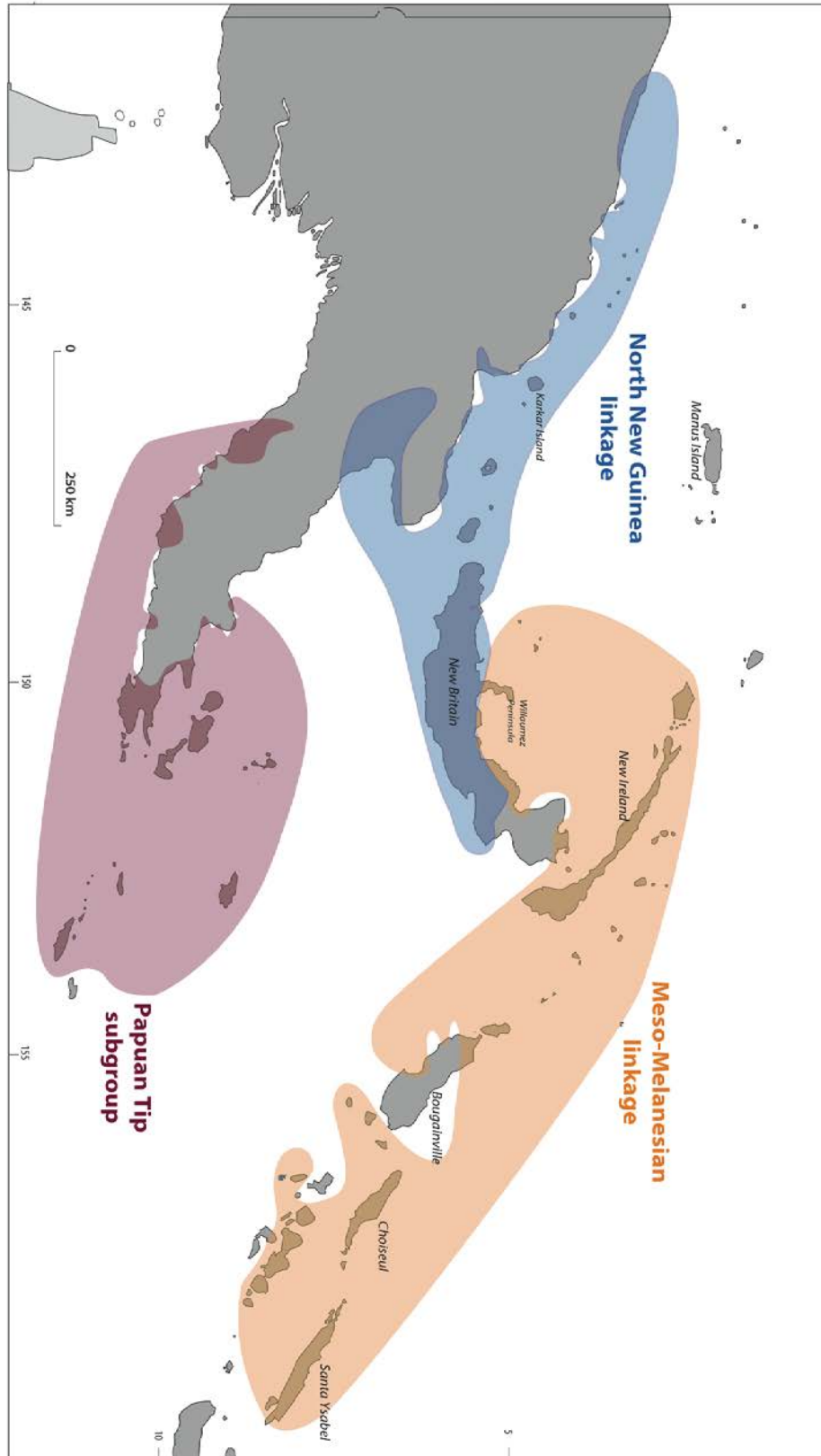


Figure 11: Groupings within Western Oceanic

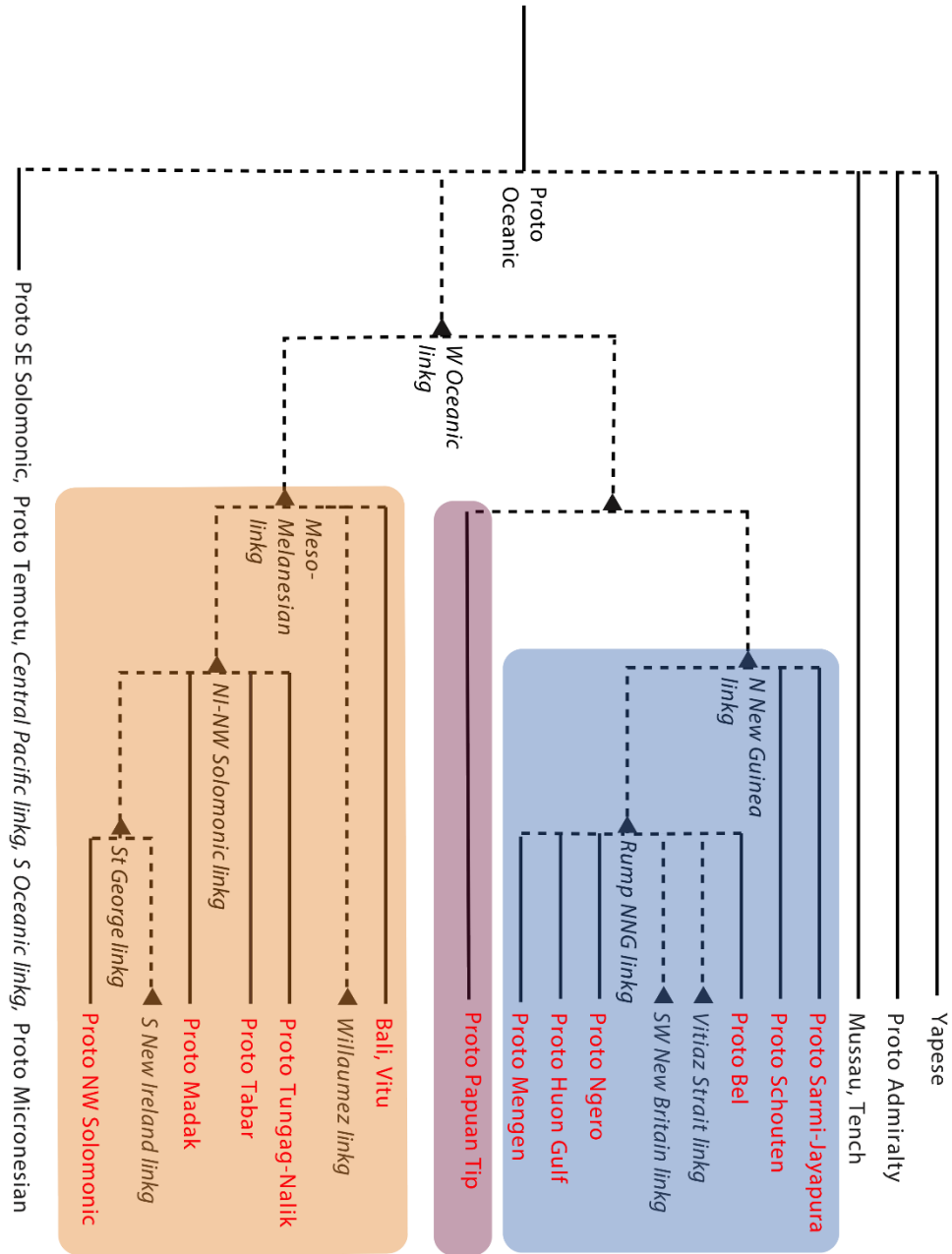


Figure 12: Oceanic phylogeny, showing the internal groupings of Western Oceanic

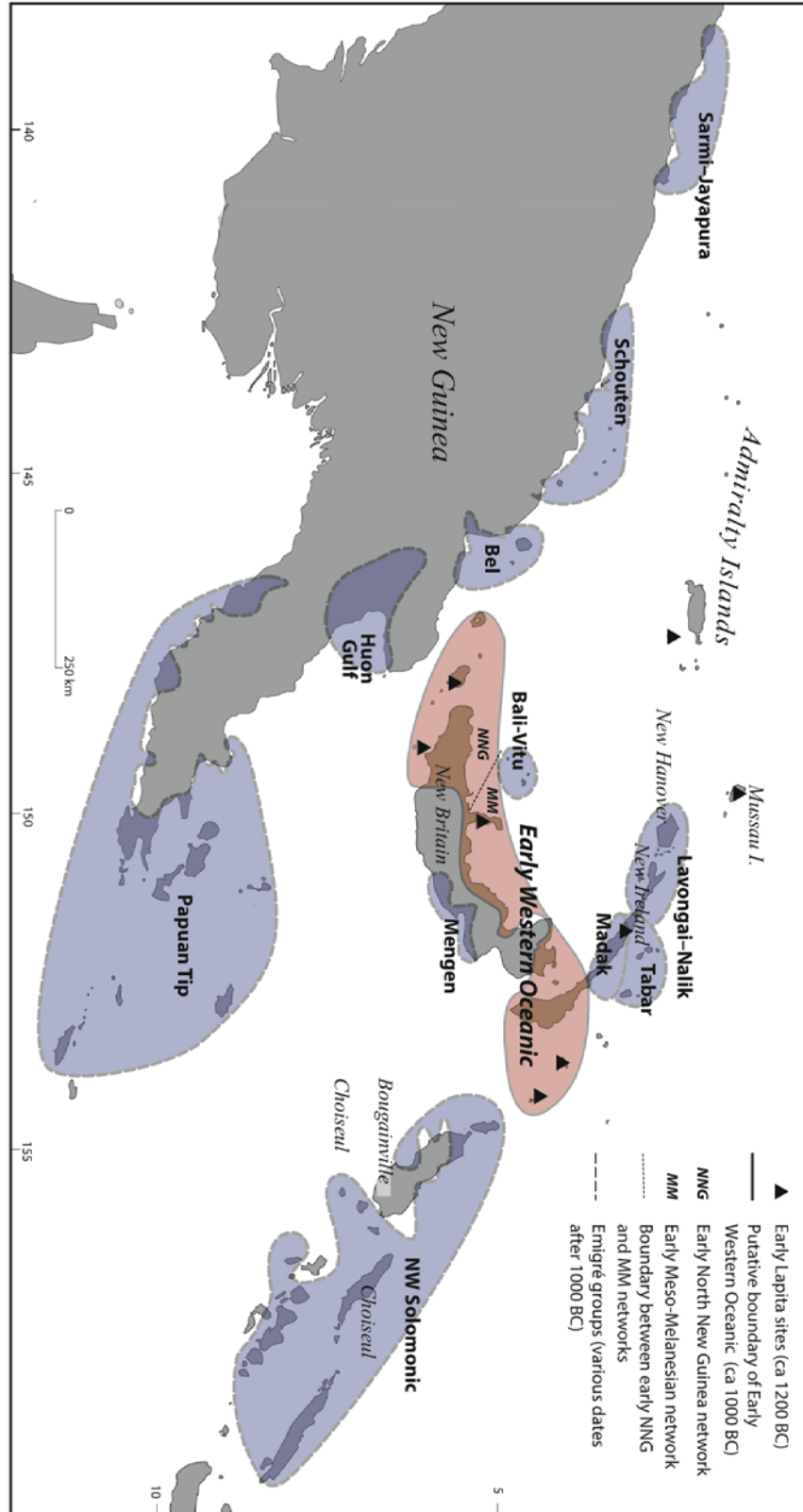


Figure 13: Early Lapita sites, subgroups within Western Oceanic, and the rumps of the North New Guinea and Meso-Melanesian linkages

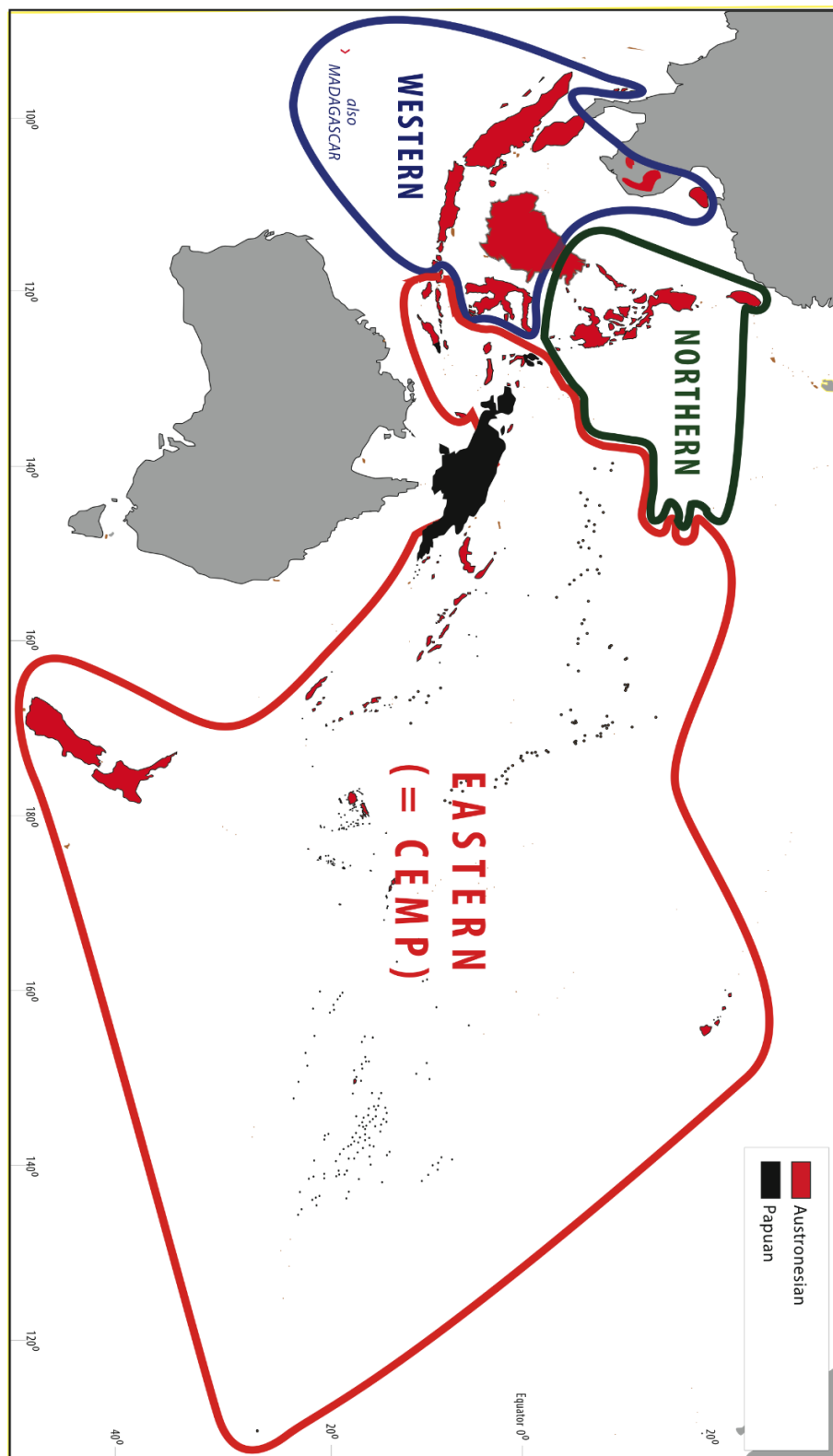


Figure 14: Broad typological regions within Austronesian

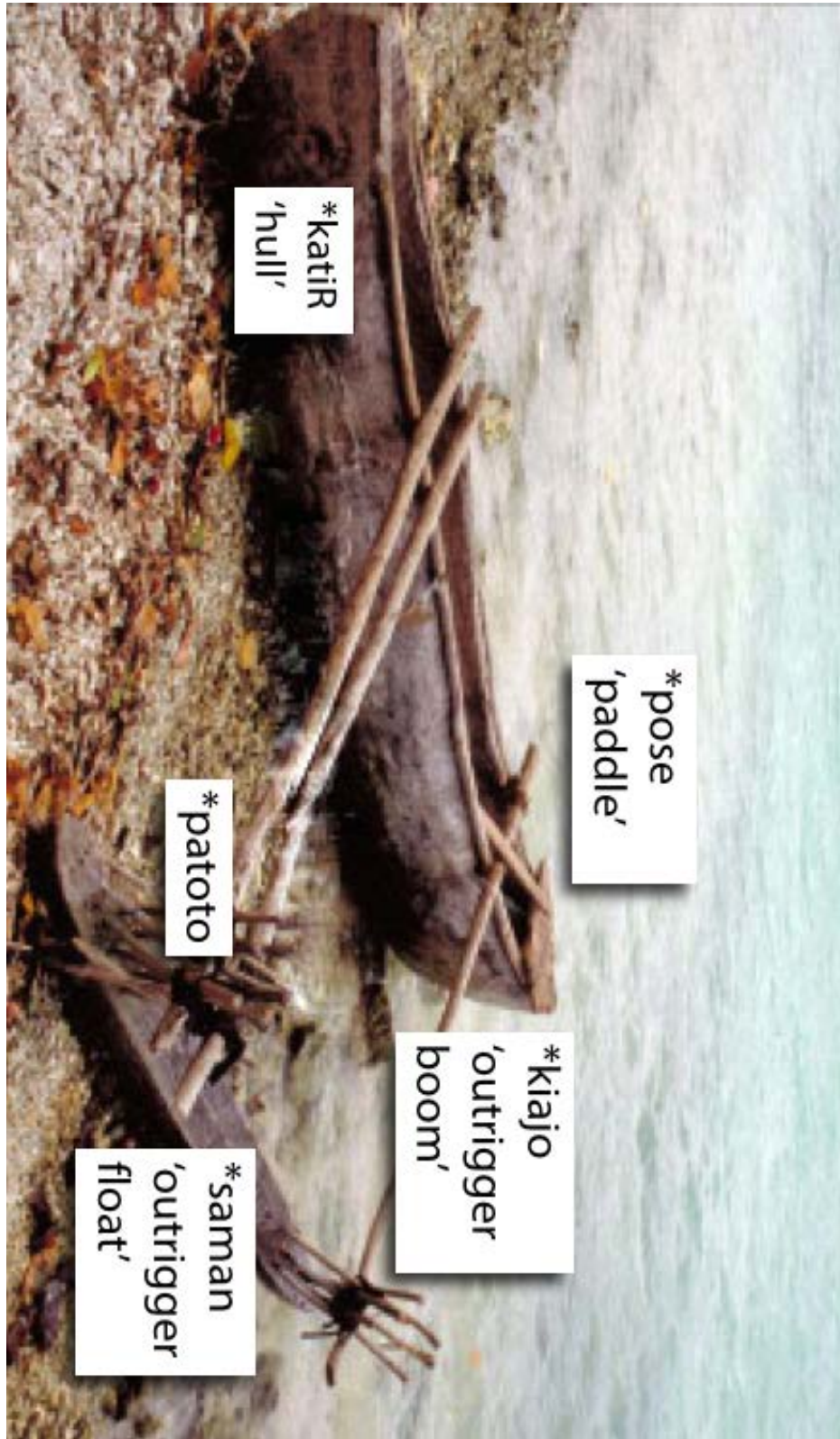


Figure 15: Terms for canoe (*waga) parts
(Additional terms include *tuku 'mast', *layaR 'sail', *limas 'bailer', *jauq 'anchor', *jila 'sail boom'.)

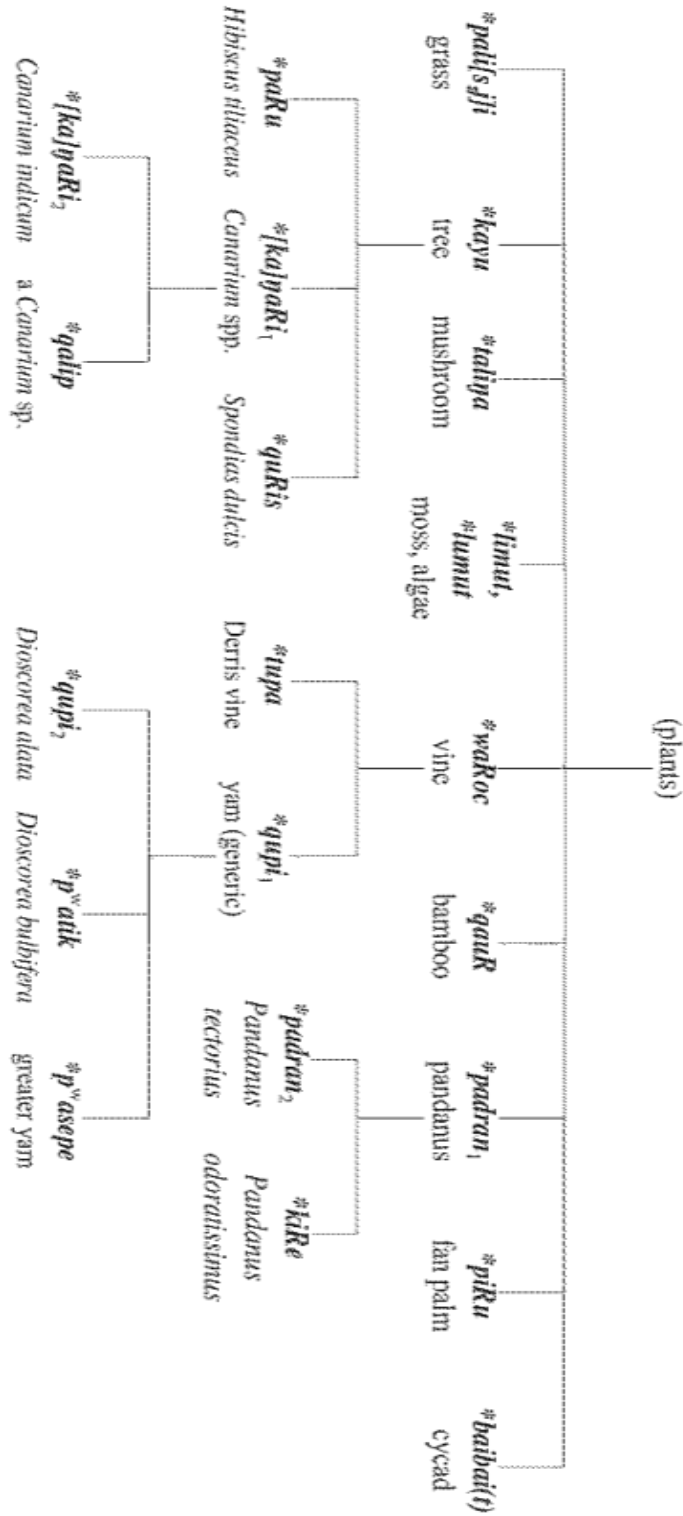


Figure 16: : POc ethnobotanical classification (Evans 2008:83)

CLASSIFYING OLD RAPA: LINGUISTIC EVIDENCE FOR CONTACT NETWORKS IN SOUTHEAST POLYNESIA

Mary Walworth

Max Planck Institute for the Science of Human History

walworth@shh.mpg.de

Abstract

The historical classification of Old Rapa, a Polynesian language spoken on the island of Rapa Iti, has never been thoroughly investigated. Based on the author's recent documentation of the language, this paper provides the first detailed historical investigation of Old Rapa; the results of which reveal a number of unique features in Old Rapa with respect to other Eastern Polynesian languages. Through a comparative analysis, evidence is provided for an especially close relationship between Mangaian and Old Rapa, as well as for shared innovations between Old Rapa and Rarotongan, Mangarevan, and Rapanui. Furthermore, the new linguistic information provided here indicates that there was an ongoing micro contact network between Rapa Iti and Mangaia. This network eventually expanded to include Rarotonga, Mangareva, and Rapa Nui.

Keywords: Polynesia, contact networks, historical classification

ISO 639-3 code: rap

1. Introduction¹

Old Rapa, the indigenous language of Rapa Iti (French Polynesia), is a severely endangered Polynesian language. It is spoken today by only a very few members of the community (mostly elders) and has been almost completely replaced by a Tahitian-Old Rapa mixed language called Reo Rapa (Walworth 2015). The historical relationships of Old Rapa have never been thoroughly analyzed due to the lack of documentation and description of the language. As a result, Old Rapa's close genetic affiliations have been more or less assumed based on very limited data or casual observations. However, upon closer examination through my own field investigations since 2012, it is evident that Old Rapa exhibits a number of unique features with respect to other Eastern Polynesian languages. The existence of these features merits further investigation in order to understand Old Rapa's classification within EP. This paper addresses the results of such an inquiry, and demonstrates the ways in which Old Rapa's unique linguistic qualities can lead to understanding the language's specific genetic affiliation as well as the Rapa people's prehistoric contacts.

This paper primarily highlights the more unique features of Old Rapa and then discusses preliminary comparative observations with other related languages. First, I summarize references to Old Rapa's genetic affiliation in historical observations and in the current literature on Polynesian

¹ I am indebted to Robert Blust and Yuko Otsuka for their encouragement to write this paper (originally a chapter in my dissertation) and for their suggestions for improvement on earlier versions. I would also like to thank Malcolm Ross and Russell Gray for their interest and support; as well as Terry Hunt, Ross Clark, and Albert Davletshin for their comments on the data within. Special appreciation is due to the native speakers of the languages discussed, with whom I consulted regularly: in Rapa Iti - Te'a Tamata, Ma'urei Angia, Takura Angia, Teuira Vahine, and Lionel Watanbe; in Mangareva - Monica Paheo, Bruno Schmidt, and Ena Manuireva; in Rapa Nui - Petero Huke and Mr. Tongariki; additional thanks to Sally Nicholas who provided data for Ma'u'ke. The research for this paper was funded in part by the Bilinski Dissertation Fellowship and also by the National Science Foundation Documenting Endangered Languages Program. All errors in my interpretation or citation of the data are my own.

languages. I then discuss evidence for Old Rapa as a Central Eastern Polynesian² language and examine its possible CEP internal relationships. Third, I explain some of Old Rapa's aberrant features, and discuss how these may demonstrate prehistoric relationships (either genetic or contact-based) with certain other Polynesian languages. Finally, I propose a scenario for Rapa Iti prehistory: early migration from the Southern Cook Islands, subsequent development of a wide-ranging contact sphere that extended to include other areas of South Polynesia, and then a period of significant isolation up until European contact.

2. References to Old Rapa's genetic classification

A number of explorers, missionaries, and early researchers who arrived in Rapa Iti during the nineteenth and twentieth centuries made informal observations about the Rapa language with respect to other Eastern Polynesian (EP) languages. While these accounts cannot be viewed as scientific or technically linguistic, they are important to consider as they indicate intelligibility with other EP languages and offer support that Old Rapa is a very different language than the more widely spoken Reo Rapa. Observations on the language based on contact with Old Rapa speakers prior to heavy Tahitian influence indicate its clear relationship with other CEP languages, and identify it as notably different than Tahitian.

2.1. Historical observations

Vancouver was the first to describe his encounters with people from Rapa Iti. His first impressions of the language he heard in Rapa Iti led him to declare that the island was definitely "part of the Great South Sea nation"; however, he noted that a Hawaiian man traveling with him was unable to understand the Rapa people (Vancouver 1798:75). Stutchbury remarked more specifically "they [Rapa people] do not speak the New Zealand or Tahitian language but something resembling the Marquesan" (1996:71–72). Davies (Newbury 1961:280) also remarked on the dissimilarity of Old Rapa and Tahitian, reporting that two men who were taken from Rapa Iti onto his ship in 1825 could not understand much Tahitian upon arrival in the Society Islands. Furthermore, in 1828, Cuming (Richards 2007:6-7) observed that while "the language of the islanders without doubt had the same [distant] origin with those of the Society Islands...[in their] language, manner and customs [they] differ materially from the inhabitants of the Society Islands of which they had not any knowledge until the arrival of the native teachers from Otaheite...They could not understand them at first."

In 1829, missionaries Pritchard and Simpson (Richards 2007:7) also noted the differences of Old Rapa from Tahitian: "The Rapan [language] in many respects is different from the Tahitian dialect. From the frequent use of the k and the ng or gn, it appears more to resemble New Zealand [Māori] or the Marquesans." Ellis (1838:364) noted that Old Rapa sounded more like Māori than Tahitian. Finally, Hale (1846:141) wrote that the language of Rapa must come from the Cook Islands, as it was nearly identical to Rarotongan.

More linguistically oriented observations about Old Rapa come from Stokes and Schooling in the twentieth century. Stokes (1955:316), while he did not suggest any specific genetic affiliation for Old Rapa, did note that by the time he had arrived on the island in 1921, men were speaking a language that resembled Tahitian, or a mixed Tahitian-Rapa language. Women, on the other hand, were still speaking the older language, and Stokes observed that it was much different than the Tahitianized language that the men used. Finally, in his 1981 sociolinguistic survey, Schooling (1981:22) set out to quantify the extent to which Tahitian had influenced other French Polynesian languages. In this study, he did not go to Rapa Iti; however, he spoke with Rapa people in Tahiti. Based on his observations, he stated that "Rapan" was a language closely related to Marquesan, with Mangarevan influences. He

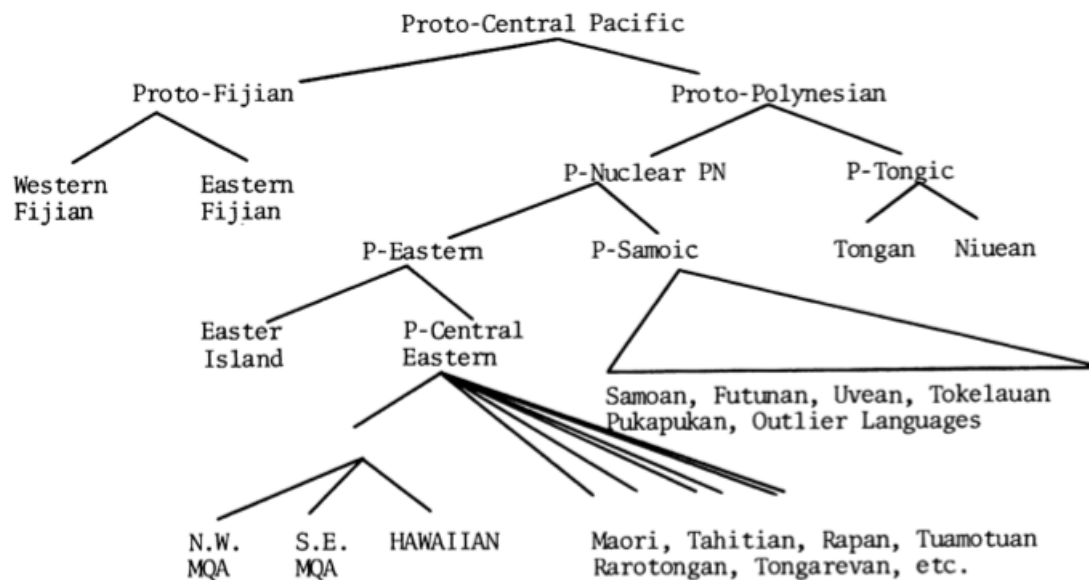
² Abbreviations for language names and language groups are as follows throughout the rest of the paper: Aitutaki = ATK, Central Eastern Polynesian = CEP, Eastern Polynesian = EP, Hawaiian = HWN, Ma'uke = MKE, Mangaian = MIA, Mangarevan = MGV, Marquesan = MQS, Moriori = MOR, OR = Old Rapa, Proto Central Eastern Polynesian=PCE, Proto Eastern Polynesian = PEP, Proto Nuclear Polynesian = PNP, Proto-Polynesian = PPN, Rapanui = RN, Rarotongan = RAR, Tahitian = TAH, Tuamotuan dialects = TUA.

furthermore wrote that the Rapa language was “sufficiently different that neither a speaker of another Australs dialect, nor a Tahitian would understand [it] on first hearing it.”

2.2. Current linguistic literature

In the current body of literature on historical relationships of Polynesian languages, there is little reference to Rapa Iti. The few that mention Rapa’s language agree that it is an Eastern Polynesian language, though projections on its precise placement within EP are varied. Green (1966:27–28)³ included “Rapan”⁴ among the Marquesic languages, citing four lexical correspondences: *taeti* ‘child’, *nga'u* ‘bite’, *rongo'uru* ‘ten’, and *kami'a* ‘canoe’. Wilson disputed two of these four as genuinely Marquesic in 2010 (293, 298), and in 2012 (350–351) rejected the claims that any of these items are markedly Marquesic. Pawley and Green (1974:44), listed “Rapan” among traditionally Tahitic languages, separate from the traditionally Marquesic languages (Figure 1). Marck (2000:185), like Pawley and Green, identified Rapa as a Tahitic language.

Figure 1: Pawley and Green’s (1974:44) classification of Central Pacific languages

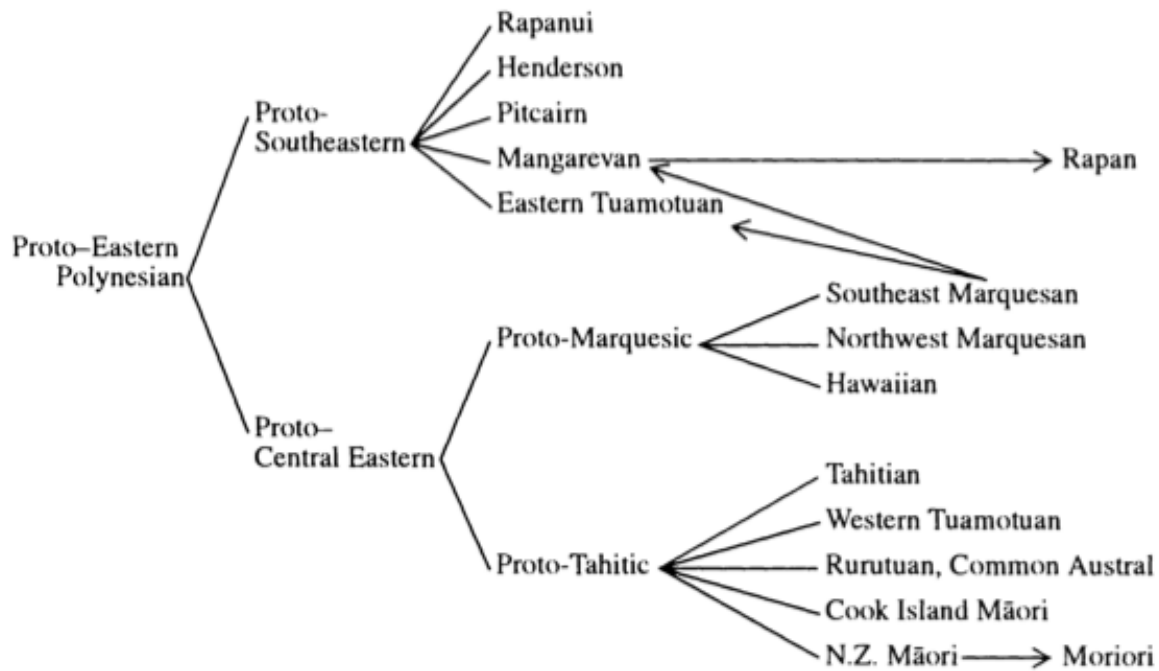


Fischer (2000) classified Old Rapa as Marquesic, and then, in 2001, hypothesized that it is actually part of a South-Eastern Polynesian (SEP) subgroup, a direct descendant of an older form of Mangarevan that had undergone Marquesic influence. Under this hypothesis, he claimed that Rapa Iti had been settled directly from Mangareva.

³ Based on data from Stokes 1955.

⁴ Often, Old Rapa is called “Rapan” by outsiders. This is not a term used by local people or by speakers of Old Rapa.

Figure 2: Fischer's 2001 SEP hypothesis



Fischer's SEP hypothesis is flawed as it is based on a group of languages that is extremely understudied, including two languages (Henderson and Pitcairn) that are only presumed to have been spoken and for which no actual records exist. Furthermore, the data he used to support this hypothesis are limited to Mangarevan. For these reasons, among others, his hypothesis has been widely disputed (Rutter 2002; Marck 2002; Wilson 2012:351–352).

3. Old Rapa as a CEP language

The previous assessments of Old Rapa are varied, and do not provide a clear classification of the language. The only commonality among the historical observations and linguistic categorizations is that Old Rapa is most certainly an Eastern Polynesian language. What is critically undecided is its more exact membership within Eastern Polynesian. Based on my present study, Old Rapa appears to share the same innovations as other CEP languages and can thus be classified as such. Table 1 demonstrates Old Rapa's consonant reflexes from PEP and PCE.

Table 1: Consonant reflexes of Proto Polynesian, Proto Eastern Polynesian, and Proto Central Eastern in Old Rapa

PPN	*p	*t	*k	*m	*n	*ŋ	*ʔ	*f	*s	*h	*w	*l	*r
PEP	*p	*t	*k	*m	*n	*ŋ	*ʔ	*f	*s	∅	*w	*r	*r
PCE	*p	*t	*k	*m	*n	*ŋ	*ʔ	*f	*s	∅	*w	*r	*r
OR	p	t	k	m	n	ŋ	∅	ʔ	ʔ	∅	v	r	r

In addition to these consonant reflexes, Old Rapa exhibits most of the defining characteristics of CEP languages. The following have been identified as strong evidence for PCE:⁵ *tahito 'old, ancient' (semantic innovation from Green 1966:17-18); and the phonological innovation PEP *faf > PCE > *waf. Examples of this are provided in Table 2.

⁵ See Walworth (2014:262–263) for further discussion and summary of PCE's defining characteristics. I no longer use Green's *kite 'to know, to see' as a PCE innovation and instead list this to be a PEP innovation based on evidence of a reflex in Rapanui: *tikea* 'to see' (Greenhill and Clark 2011). The Rapanui reflex was previously overlooked due to the metathesis that has occurred in this form.

Table 2: PEP *faf- to PCE *waf-

PEP	PCE	Gloss
*fafa	*wafa	‘carry on back’
*fafie	*wafie	‘firewood’
*fafine	*wafine	‘woman’
*fafa	*wafa	‘mouth’
*fafii	*wafii	‘wrap food in a leaf’
*fafo	*wafo	‘outside’

Marck (2000:132) identified an additional five sporadic sound changes in PCE. These appear in Table 3.

Table 3: Sporadic sound changes in PCE (Marck 2000:132)

PEP	PCE	Gloss
*ŋu-feke	*mu-feke	‘squid’
*ŋau	*ŋahu	‘chew, bite’
*faahua	*paahua	‘ <i>Tridacna</i> (giant clam)’
*kai	*koi	‘sharp’
*kau-natu	*kau-nati	‘fire-plow’

Finally, Green (1985:12) and Marck (1996) presented nine grammatical innovations for PCE: *tei ‘present position’; *ina(a) fea ‘when (past)’; *le(‘)ila ‘there, aforementioned place’; *noo/naa ‘possessive particle’; *me ‘and, with, plus’; *taua ‘that aforementioned’;⁶ *aanei ‘interrogative’; *vai ‘who’; and *vau ‘1st person singular’.

Old Rapa exhibits Green’s (1966) PCE semantic innovation *tahito as *ta’ito* ‘old’. PEP *faf- to PCE *waf- is not evident, as Old Rapa has uniquely innovated forms for ‘woman’ (OR *pē’ā*), ‘wrap-up’ (OR *veinga*) and ‘firewood’ (OR *rārā*) and retains PPN *ngutu for ‘mouth’. Old Rapa does not exhibit forms that resemble PCE *wafo or PEP *fafo for ‘outside’ (OR *rāpae*), nor PCE *vafa or PEP *fafa for ‘carry on back’ (OR *amo*).⁷

Of Marck’s five sporadic sound changes, Old Rapa possesses *mī’eke* ‘type of squid’, *nga’u* ‘bite’, and *koikoi* ‘sharp, pointy’. Practical explanations can be sought for the absence of Marck’s two additional sporadic sound changes: Old Rapa has a unique innovation for ‘fire-plow’ and ‘*Tridacna*’ are not found in Rapa Iti’s cool waters (pers. comm. with local Rapa fisherman).

Regarding the grammatical innovations outlined by Green and Marck, Old Rapa demonstrates the following reflexes:⁸ PCE *tei as OR *ti* ‘immediately, here, now’; PCE *noo/naa as *nō/nā* ‘genitive particle’; and PCE *vai as OR *vai* ‘who’. Old Rapa does not exhibit reflexes of PCE *me, *ina(a) fea, or *vau. For PCE *ina(a) fea, Old Rapa has merged the past and present interrogative forms for ‘when’ and uses *a’ea* for both. Old Rapa forms for ‘first person singular’ (OR *ou*) and ‘with, plus’ (OR *ma*), have been retained from PPN *au and *ma, respectively. The retention of PPN *au for ‘first person singular’ is shared only with Mangarevan among the CEP languages.

The evidence presented in this section demonstrates that Old Rapa is a CEP language; but what of its further classification within CEP? This is more difficult to ascertain. Many scholars, as noted in previous sections, have classified Old Rapa as a Marquesic language, or in the case of Fischer (2001), a language descended directly from Mangarevan, after Marquesic “intrusion.” Others have categorized Old Rapa as Tahitic. However, this categorization is likely due to the similarities to

⁶ *taua ‘retrospective definitive’ has been reconstructed for Proto-Tahitic by Greenhill and Clark (2011), as well as for PCE as indicated here, but due to evidence of a related form in the traditionally “Marquesic” languages of Hawaiian (*ua*) and Mangarevan (*tou*); as well as evidence in Rapanui for the form *tou*, I posit a *taua reconstruction for PEP and not for PCE. This is further discussed in section 4.2.6.

⁷ OR *va’a* ‘carry a baby on the back’ may be representative of Old Rapa’s retention of PCE’s dual phonological innovations.

⁸ See Walworth (2015:74–166) for examples and discussions on the functions of these grammatical markers.

Tahitian that Reo Rapa exhibits, given that it is heavily mixed with Tahitian. Based on my data and analyses, Old Rapa does not exhibit any particular features that would classify it under either of the traditional CEP subgroups, which, in any case, have been recently challenged (see Walworth 2014). This is not to say that Old Rapa does not demonstrate any particular relationships with other CEP languages. To the contrary, it has striking linguistic similarity to several other CEP languages, but not under the traditional subgrouping framework. In the sections that follow, I discuss these potential relationships while highlighting some of Old Rapa’s more unusual features.

4. Comparative analyses of some Old Rapa features

In this section, I examine some of the phonological, grammatical, and lexical features of Old Rapa, paying particularly close attention to those features that represent a departure from most of the other CEP languages, either as innovations or as retentions.

4.1. Phonological features

This section addresses Old Rapa’s consonant reflexes and the other CEP languages that exhibit the same reflexes from PCE. Furthermore, this section highlights a sporadic vowel change in Old Rapa that is shared with other EP languages.

4.1.1. Consonant reflexes

Old Rapa’s consonant reflexes from PCE are identical to those of Rarotongan, Mangaian, Ma’uke, Aitutaki, and Mangarevan. Rarotongan, Mangaian, Ma’uke, and Aitutaki are languages spoken in the Southern Cook Islands (approximately 900 NM northwest of Rapa Iti). Mangarevan is spoken in the Gambier Islands (approximately 570 NM northeast of Rapa Iti). These shared consonant reflexes are striking, as this group of languages represents the largest group of EP languages to share identical consonant reflexes.

Table 4: *Consonant reflexes of PEP and PCE in OR, RAR, MIA, MGV, ATK, and MKE*

PEP	*p	*t	*k	*m	*n	*ŋ	*ʔ	*f	*s	*w	*r
PCE	*p	*t	*k	*m	*n	*ŋ	*ʔ	*f	*s	*w	*r
OR	p	t	k	m	n	ŋ	∅	ʔ	ʔ	v	r
RAR	p	t	k	m	n	ŋ	∅	ʔ	ʔ	v	r
MIA	p	t	k	m	n	ŋ	∅	ʔ	ʔ	v	r
MGV	p	t	k	m	n	ŋ	∅	ʔ	ʔ	v	r
ATK	p	t	k	m	n	ŋ	∅	ʔ	ʔ	v	r
MKE	p	t	k	m	n	ŋ	∅	ʔ	ʔ	v	r

4.1.2. Sporadic sound change

One sporadic sound change of PNP *k to *t* is observed in OR *tauru* ‘tree top’ (PNP *kauru ‘tree top’). This change is shared with Mangarevan and Rapanui.

4.2. Grammatical features

This section highlights several of Old Rapa’s grammatical words that are historically unusual. These include: the perfective aspect marker *ka*, adverbial *tuai*, adverbial *ta’anga*, negative past *ki’ere*, negative non-past *kāre*, and definite *tō*.

4.2.1. Perfective aspect

Most Eastern Polynesian languages denote the perfective aspect using a reflex of PPN *kua ‘perfective aspect marker’ (Clark 1976:30).

Table 5: PPN *kua reflexes in some CEP languages⁹

TAH	MAO	RAR	MSQ	HAW	TUA
'ua	kua	kua	'ua	'ua	kua

In Old Rapa, however, the perfective is marked by *ka*. A form *kua* does occur but can only be used with a small group of intransitive verbs that incorporate a subject, and it appears to have a deeper ‘past’ connotation that contrasts with the *ka* perfective (Walworth 2015:102-103). There are two possible explanations for the change PPN *kua > Old Rapa *ka*: (1) Old Rapa underwent an irregular phonological change from PPN *kua (loss of [u]), resulting in *ka* for the perfect marker. This change is not exhibited in any other EP language. (2) OR *ka*, as a perfective marker, represents a semantic innovation from PPN *kaa, which Clark (1976:30) reconstructs as ‘future’ or ‘inceptive’ aspect. This innovation may be part of a shared innovation with Māori, Manganian, and Mangarevan.

Bauer et al. (1993; 1997) and Harlow (2012) offered evidence for a TAM marker *ka* in Māori,¹⁰ with, however, varied interpretations of its function. Harlow (1989) wrote that the particle *ka* serves only to mark that a phrase is verbal and denotes no tense, aspect, or modal value. He expanded on this in 2012: “When no adverbial or previous [tense-aspect] marking determines a tense, the default reading of *ka* is temporally present, aspectually aorist” (137). Bauer et al. (1997:85) wrote that *ka* has more of a non-specific aspect function and can be used to indicate past, future, or present tense.

Manganian exhibits a similar *ka*; however, it is unclear whether it denotes perfective or is more non-specific and oriented to the surrounding context. The example below from the *Manganian Dictionary* indicates a perfective translation, but provides no context, making it difficult to surmise if its semantic value is truly perfective or if it is contextually based.

- (1) *ka* *ʻitonga te kuru*
TAM bruise DEF breadfruit
‘The breadfruit is bruised’.

(Manganian Dictionary, 2013)¹¹

Based on analysis of published Manganian texts (Reilly 1993) as well as examples from the *Manganian Dictionary* (2013), it appears that Manganian typically uses *kua* to denote the perfective. It follows, then, that *ka* in Manganian may function as it does in Māori, as a non-specific aspect marker.

Finally, there is evidence of a somewhat ambiguous aspect marker *ka* in Mangarevan. According to Ena Manuireva (pers. comm. 2014), this *ka* can be used to express the future and perfective. For example, a form of saying ‘goodbye’ in Mangarevan is *ka no'o koe*, literally, ‘you (2S) should stay’ or ‘you (2S) will stay’. According to Mr. Manuireva, *ka* in this case can mean both imperative and future, and can be interchanged with the imperative marker *a* or the future marker *e*. The inexact value of *ka* may indicate that it has functions as a ‘non-specific’ aspect marker in Mangarevan as well.

This evidence of a *ka* aspect marker in Māori, Manganian, and Mangarevan may indicate that a non-specific marker was a shared innovation in Māori, Manganian, Mangarevan, and Old Rapa. Over time, Old Rapa replaced perfective *kua* with non-specific *ka*. Trace evidence of *kua* does exist in Old Rapa, however it has a slightly different semantic function from *ka*, as illustrated in (2a) and (2b).

- (2a) *kua ngaro*
PFV disappear
‘It disappeared’.

⁹ Unless otherwise noted, forms cited are taken from the following sources throughout this paper: Tahitian from Lemaître 1973; Maori from Williams 1971 (cited in Greenhill and Clark 2011); Rarotongan from Buse 1995; Mangarevan from Tregear 1899; Manganian from the *Manganian Dictionary*; Marquesan from Dordillon 1904; Hawaiian from Pukui & Elbert 1957; and the Tuamotuan dialects from Stimson & Marshall 1964.

¹⁰ TAM = tense aspect mood

¹¹ I have added the interlinear gloss and free translation.

- (2b) *ka ngaro*
 PFV disappear
 ‘It just disappeared’.

This trace evidence further supports the replacement explanation in Old Rapa; had a phonological change occurred (**kua* > *ka*), evidence of a *kua* form in the Old Rapa corpus would be unlikely.

4.2.2. Adverbial *tuai*

Another notable feature in Old Rapa is *tuai* ‘absolutely, definitely’ which demonstrates a semantic shift from PPN **tuai* ‘old’, as well as a grammatical shift (becoming definitively adverbial). Reflexes for PPN **tuai* are not found in any other EP language and similar shifts are not noted in only one other PN language - Niuean, a distantly related Tongic language. In Old Rapa, *tuai* functions as an adverb that carries perfective connotations in that it emphasizes that an action has indeed been carried out. In Niuean, *tuai* has a primarily perfective aspect function (Seiter 1980:2), but is syntactically adverbial. Niuean’s placement of *tuai* is post-verbal, an atypical location for a Polynesian aspect marker, but the prototypical position of a Polynesian adverb.

- (3) *hau tuai e tehina haau.*
 come PERF ABS brother your
 ‘Your little brother has come’.

(taken from Seiter 1980:8)

According to Seiter (1980:8), the perfective in Niuean may be marked by a co-occurrence of the perfective aspect markers *kua* and *tuai*. This is, in fact, the most common way to mark perfective in Niuean.

- (4) *kua ligi tuai e au e kapini tī ma-au.*
 PERF pour PERF ERG I ABS cup tea for-you
 ‘I’ve poured a cup of tea for you’.

(taken from Seiter 1980:8)

There is nothing else by way of particular linguistic similarity that would point to a subgrouping relationship between Niuean and Rapa Iti, nor are these two identical changes likely to have arisen independently in both languages, so this single connection is more likely contact related. Strong evidence indicates similarities between Niuean and EP languages, which are typically attributed to borrowing through contact with the Cook Islands (Clark 1979; Marck 2000; Otsuka 2006). If EP features were borrowed into Niuean from contact with the Cooks, the existence of the adverbial functioning *tuai* in Old Rapa provides evidence that Rapa Iti was to some extent involved in this contact network.

4.2.3. Adverbial *ta'anga*

Most reflexes of PPN **tafaŋa* ‘naked, bare, clear’ in EP languages retain the adjectival function and semantic value of ‘naked, bare, clear’: TAH *taha'a* ‘naked’ (Fare Vāna'a 1999); MAO *tahanga* ‘naked, empty’; RAR *taa'aka* ‘naked, bare, empty-handed, destitute’; MSQ *tahaka* ‘clear, open, discovered’; HAW *kohana* ‘naked’; HAW *kaahana* ‘clearing (as in a forest)’; TUA *tahanga(hanga)* ‘clear, naked, obvious’. In addition to the previously stated meaning and function, Maori and the Tuamotuan dialects also exhibit reflexes of **tafaŋa* with adverbial function and extended meaning. Stimson and Marshall (1964) reported an adverbial reflex with a wide semantic range in some of the Tuamotuan dialects, *tahanga* ‘for a little while, just a moment, a little, moderately, suddenly, surely,

certainly, positively’. For Maori, Williams (1971; cited in Greenhill and Clark 2011) also reported *taahanga* to have the meaning ‘moderately, a little’.

Old Rapa also exhibits an adverbial reflex of PPN *tafaŋa, and additionally appears to have undergone semantic change from the PPN meaning to ‘only, simply, continuously’. The semantic and functional changes from PPN *tafaŋa exhibited in Old Rapa are shared with Mangarevan (*ta’anga*, pers. comm. Mangarevan consultants 2015) and Rapanui *tahanga* ‘simply, only, continuously’.¹² Based on this data, I would not suggest that the shared semantic and functional innovation in Old Rapa, Mangareva, and Rapanui signals a subgrouping relationship between these three languages; however, it may be evidence for some contact between them.

4.2.4. Negative past *ki’ere*

Old Rapa’s marker for negative past constructions is *ki’ere*. This form is not evidenced elsewhere in Polynesia. However, it appears to be a compounded reflex of PCE *kihai ‘negative’ and PEP *ŋere ‘deprived of’. This would certainly be a unique construction for a Polynesian negative form, but is not improbable. This would mean that Old Rapa retained only the *ki portion of PCE *kihai, as *ki*, and compounded it with *ngere*, resulting in *ki-ngere*. At some later point, the velar nasal was reduced to glottal stop, under the influence of the Tahitian reflex of PPN *ŋere, *’ere*. Old Rapa does borrow the nominal negative from Tahitian, *e ’ere*, which clearly incorporates the Tahitian reflex of PPN *ŋere. It is not unlikely then that an older Old Rapa form of *ki-ngere* might have experienced a similar shift, resulting in *ki’ere*.

Māori and Ma’uke are the only other CEP languages that have retained PCE *kihai. Based on evidence from Clark (1976:95) and Ma’uke linguist Sally Nicholas (pers. comm. 2014), both Māori and Ma’uke demonstrate a reflex of *kihai (*kihai* and *ki’ai*, respectively) to mark the negative past, thus sharing Old Rapa’s semantic value of the *kihai reflex.

4.2.5. Non-past negative *kāre*

Clark (1976:98–100) reconstructed PPN *kole as a verb that indicated ‘lacking’ or ‘non-existence’. However, he remarked that its presence as a negative marker was only apparent among EP languages. He wrote, “Outside of this subgroup, not only is it unknown as a form of NEG, but plausible cognates of any sort are hard to find” (1976:98). For this reason, I find it more suitable to reconstruct a verb of non-existence, *kore (incorporating the merger of PPN *r and *l), only as far back as PEP. Among CEP languages, PEP *kore “fused” with either PCE *e ‘future’ or PCE *ka ‘non-future’, which produced past, present, and future negative markers in EP languages (Clark 1976).¹³

¹² Personal communication with a Rapa Nui speaker, August 2013; also evidenced in Churchill (1912:254) ‘only, solely, alone, wholly, with- out stopping, always, quite, a sort of superlative’.

¹³ It is important to note that Clark (1976:30–33) did not explicitly reconstruct PPN *ka in his discussion of PPN tense-aspect markers. He does however reference this *ka “tense-marker” to mean ‘non-future’ in his analysis of PPN *kole and his treatment of CEP reflexes (p.99). Though he does not specify to what proto-language *ka is reconstructable, I interpret from the data he provided that *ka ‘non-future’ can be reconstructed for PCE.

Table 6: Reflexes of PCE *ka/*e + *kore in some CEP languages (Clark 1976:99)

	Past	Present	Future
MAO	<i>(kiihai)</i>	<i>kaahore</i>	<i>e kore</i>
RAR	<i>kaare</i>	<i>kaare</i>	<i>kaare</i>
TAH	<i>'aore</i>	<i>'aore</i>	<i>'e 'ore</i>
MVA	<i>e kore</i> <i>kakore</i>	<i>e kore</i> <i>kakore</i>	<i>e kore</i>
MQS	<i>'a'o'e</i>	<i>'a'o'e</i>	<i>'a'o'e</i>
HAW	<i>'a'ole</i>	<i>'a'ole</i>	<i>'a'ole</i>

Clark (1976:100) additionally noted that in all CEP languages the tense + *kore form emerges as a negative existential. Old Rapa's reflex of *kore, *kāre* can also function in this way.

Old Rapa's non-past negative *kāre*, as Clark stated to be true for other CEP languages, likely derives from an earlier PCE *ka + *kore merger. The resulting form, *kakore, then underwent a sporadic deletion of [k], followed by an assimilation of [o] to [a]. This assimilation resulted in geminate [a], thus producing an apparent long [ā] in Old Rapa:

PCE *tense + *kore > POR *kakore > *kaore* > *kāre*

This identical series of sound changes appears to have also occurred in Rarotongan and Mangaian, which exhibit *kāre* to indicate some form of the negative. As shown in Table 6, Rarotongan uses this form for past, present, and future negative constructions. It can, of course, following Clark's observation for all CEP languages, also be used to mark the negative existential (see the Dictionary of Cook Islands Languages 2014). In Mangaian, due to lack of documentation, the function of *kāre* is not readily clear. However, the form does appear as a negative and seems to be derived from the same sound changes from PCE as in Old Rapa and Rarotongan (example (5)).

- (5) *kāre ra i ariki-'ia.*¹⁴
 NEG DEIC PFV accept-PASS
 '[He] was not accepte'.

4.2.6. Definite *tō*

In Old Rapa, *tō* functions as a definite article that is heavily discourse driven. This particular form is not found in any other Central Eastern Polynesian language; however, I believe it is semantically related to PEP **taua*, which is evidenced in several other CEP languages (Table 7). Thus, the Old Rapa form is presumably phonologically derived from PEP **taua* through the following sound changes: a sporadic loss of final *a, *au > *ou*, and subsequent monophthongization *ou* > *ō*. This exact series of changes from PEP **taua* is exhibited only in Rapa Iti among the CEP languages. However, it is very important to note that both Rapanui and Mangarevan exhibit the form *tou* as a reflex of PEP **taua* (Langdon and Tryon 1983:23 for Rapa Nui; Tregear 1899:106 for Mangarevan). The Rapanui and Mangarevan forms appear to have possibly gone through the same first two sound changes as Old Rapa. Identical sporadic sound changes such as these can provide compelling evidence for historical relationships, and in this case further suggest certain prehistoric contact between Rapa Iti, Mangareva, and Rapa Nui.

¹⁴ Text from Reilly 2007; interlinear gloss added.

Table 7: Reflexes of PEP **taua* in some CEP languages (taken from Greenhill and Clark 2011)

Language	Reflex	Description
HAW	<i>ua</i>	Refers to a previously mentioned noun
MOR	<i>wa</i>	‘those’
MAO	<i>taua</i>	‘that, aforementioned’
RAR	<i>taua</i>	Demonstrative and relative pronoun; ‘that aforementioned’
TAH	<i>taua</i>	‘aforementioned’ (pers. comm. Jack Ward); ‘this/that’, when used with DEIC <i>ra</i> (Lemaitre 1973)
TUA	<i>taua</i>	‘that, the aforesaid’
MIA	<i>taua</i>	‘that/those aforementioned’ (as demonstrated in Reilly 1993)

4.3. Lexical innovations

Old Rapa exhibits a significant number of basic lexical items that cannot be reconstructed for Proto-Polynesian. My study is not the first to comment on these as unusual phenomena. In John F. G. Stokes’s 1955 article “Language in Rapa,” he noted several lexical items without cognates anywhere else in Polynesia. Kieviet and Kieviet (2006:6–10) also remarked on some of this unusual terminology and offered “parallels,” otherwise known as cognates, in some other Polynesian languages. The list of Rapa innovations has expanded through my recent linguistic work on the island (Walworth 2015:186–189). Table 8 provides my current and complete list of Rapa lexical innovations. This table also indicates PPN reconstructions for the same gloss and provides other possibly related higher-level reconstructions. These innovations represent either unique forms or unique semantic shifts in Old Rapa.

Table 8: List of Rapa lexical innovations

Gloss	Rapa Innovation	Reconstructions in PPN (unless noted otherwise)	Related Forms (PPN, unless noted otherwise) ¹⁵
advance, charge	<i>mātu</i>	*qoso	*oma ‘be swift’ + *atu ‘directional - away from speaker’
armpit, tickle	<i>ketekete</i>	PCE *keke ‘armpit’; PPN *ma-gene ‘tickle’	
back	<i>moko</i>	*tuqa ‘back’	
banana	<i>tautau</i>	PEP *m(a,e)ika	*tau ‘hang, be suspended’
buttocks	<i>kōmi</i>	PCE *remu	
calm sea	<i>karamate</i>		*kale ‘A wave that ripples or breaks, rather than a swell’ + *mate ‘die, dead’
canoe	<i>kāmi’a</i>	*waka	
change direction	<i>tīkoni</i>		*koni ‘move around’
children	<i>puki</i>	*tamariki ¹⁶	
<i>Cordyline terminalis</i>	<i>karokaro</i> , <i>kaukaro</i> ¹⁷	*ti	*kalokalo ‘flower species’; *kau stalk, stem’
<i>Corokia collenetei</i>	<i>raupata</i>		*laupata ‘tree species’; *lau ‘leaf’; PCE *naupata ‘scaevola plant’ ¹⁸

¹⁵ Proto-forms were taken from Greenhill and Clark 2011, unless otherwise noted.

¹⁶ *Tamariki* is also used to mean ‘children’ in Old Rapa.

¹⁷ These two terms are generally used as synonyms today; however, all of my elder consultants report that *karokaro* is the young leaf, curled in the center of the plant; *kaukaro* refers to the entire plant.

earth oven which is covered by volcanic stones	<i>ko'otu</i> ¹⁹	*umu	
eel	<i>takaviri</i>	*pusi	PCE *takawiri 'turn, twist'
escape	<i>moka</i>	*sola	
eyebrow	<i>kene'u mata</i>	*tuke-mata	*mata 'eye'
family	<i>kōpū</i>	*saqa 'clan'	*koopuu 'gullet, stomach, belly, guts'
fire	<i>ngara'u</i>	*afi	*ŋarafu 'charcoal'
fire plow	<i>'ikā</i>		*sika 'make fire'
firewood	<i>rārā</i>	*fafie	*raqa-raqa 'small branch'
fishing net	<i>ngake</i>	*kupuŋa	
forest	<i>raro rākau</i>	PPN *wao; PCE *ŋasere	*lalo 'below, under' + *lakau 'tree'
fresh water	<i>kōta'e</i>	*wai	*tafe 'to flow, especially of a current'
fresh water source, waterfall	<i>kōringiringi</i>		*liŋi 'pour'
generation, divide	<i>kopanga</i>		
go	<i>naku</i>	*saqele	
hair of head	<i>rauka'a</i>	*lau-qulu	*lau 'leaf' + *kafa 'braided fibers'
high fort, fortified village	<i>pāre</i>		*pale 'defense'; *pa 'enclosure, fence'
immediate family	<i>puki'anga</i>		
indeed	<i>noti</i>	*foki	
large, numerous	<i>ngare</i>	*lasi	
learn	<i>'aikete</i>	*ako	*kai 'eat' + PEP *kite
learn	<i>āikete</i>	*ako	
make a path in the woods	<i>tāmoka</i>	*taa 'cut, chop, carve'	
man	<i>rua</i>	*taqane	*rua 'two'; *lua-ni 'associate, companion'
nose	<i>pitā'u</i>	*isu	
overflowing of river	<i>karea</i>		
parent	<i>karakua</i>	*matuqa	*koromatua 'elderly or wise person'
peel	<i>'oni</i>	*fore	*soni 'incise, cut into'
pretty	<i>mānea</i>		*mana-qia 'handsome lothario'; *maneqa 'play'
ridge	<i>taratika</i>	*tuqa-siwi	*tala 'pointed object'; PEP *tika 'straight'
river	<i>mangavai</i>	*wai-tahe	*maŋavai 'tributary'
river bank	<i>tupe</i>		
salt water, sea	<i>kara, kare</i>	*tahi; *miti	*kale 'a wave that ripples or breaks, rather than a swell'

¹⁸ According to Tiffany Laitame (pers. comm. 2014), a biologist and member of the Rapa Iti community, *Scaevola* and *Corokia* are not biologically similar, nor do they have any surface similarities. Furthermore, *Scaevola* is not found on Rapa Iti.

¹⁹ Stokes (1955) lists this as *kauatu*.

see	<i>noko</i>	*kite	
skinny	<i>mokōʻī</i>		
sleep	<i>komo</i>	*mohe	
small	<i>kakaio</i>	*riki	
small animals	<i>kororio</i>	*riki	
small sea cucumber	<i>'akaekae</i>	*loli ²⁰	
small taro bundles	<i>tīromi</i>		
south wind	<i>mākiki</i>	*toŋa	
speak	<i>'akaero</i> ²¹	PEP *kii	*reo 'voice'
spider	<i>kopitoru</i>	*lewelewe	
split into equal parts	<i>panga'a</i>		
spoken word	<i>koai</i>	*kupu	
stone, rock	<i>koni'i</i>	PCE *po-fatu	
storm	<i>tonou; touno</i>	*afaa	
sweet (smell and taste)	<i>kakona</i>	*maŋalo	*kona 'satiated, satisfied'
taro	<i>mīkaka</i>	*talo	
taro leaf	<i>raupaka</i>		*laupata 'tree species'; *lau 'leaf'; *pata 'spotted'
taro species	<i>'ara'ara</i>		*farafara 'plant species'
tear (n)	<i>karavai</i>	*loqi-mata	*wai 'fresh water'; *kala 'sting'
valley	<i>tīkoko</i>		*kookoo 'hole'
weak	<i>mōkiki</i>	*ŋawari	
wet taro-bed	<i>roki</i>		*loki 'enclosed area' ²²
wild sugar cane	<i>tāmi'a</i>	*too	
wild taro	<i>matae</i>		
woman	<i>pē'ā</i>	*fafine	*pela 'mud, muddy, dirt, filth'
young taro	<i>kāvake</i>		*kawake 'moon, month'

4.3.1. Social motivation for innovation

While some of the items on this list are entirely new forms, without any clear provenance (*pē'ā*, *karakua*, *kopanga*, *panga'a*, *mīkaka*, *akaekae*, *koni'i*, *pitā'u*, *moko*, *kōmi*, *āikete*, *koai*, *nākū*, *mātu*, *mokōʻī*, *kakaio*, *noko*, *komo*, *kororio*, *ko'otu*, *kāmi'a*, *karea*, *moka*, *tāmoka*, *tāmi'a*, *tīromi*, *ngare*, *ngake*, *kopitoru*, *matae*), the majority of Old Rapa's innovations appear to have been derived from other terms that likely already existed in the Old Rapa lexicon. The “other” reconstructed forms in column four in Table 8 offer the reconstructions for lexical items that may have had reflexes in Old Rapa, but have undergone unique semantic innovations.

Stokes (1955:320–321) remarked on his list that these innovations in Old Rapa were nearly all terms of “ordinary life,” otherwise known as basic vocabulary items. This remains true of my more extensive list. Most of the aberrant vocabulary found in Old Rapa is “basic” in that it is vocabulary that relates to the everyday life, activities, and environment of Rapa Iti culture. These types of terms are the least likely to change so drastically from prototypical forms barring some significant social motivation. This is particularly the case among Polynesian languages where basic vocabulary tends to be conservatively retained.

²⁰ Old Rapa does have a reflex of PCP *loli, *rori*, meaning 'larger sea cucumbers'.

²¹ This looks like a metathesized form of *reo* with a fossilized causative prefix *'aka-*. *Ero* does not have any meaning alone.

²² Hawaiian shares a semantic cognate, *lo'i*, for enclosed wet taro-bed.

So what was the motivation for deriving and inventing new forms for words that presumably already existed in the Old Rapa lexicon? Stokes (1955:319) suggested that the unique terms found in Rapa must be indicative of a “custom of word-changing.” While he admits not finding local confirmation of such a custom having existed in Rapa, he suggests that it is the only possible explanation for the changing of basic terminology. Stokes provided two main reasons for his hypothesis: (1) because many of the unique Old Rapa terms can be “derived from other terms” and (2) because relics of terms more consistent with Proto-Polynesian reflexes can be found in Old Rapa.²³ Stokes argued, then, that these terms represented “local word-coining” due to social pressure, or some sort of speech taboo.

Speech taboos were a common practice among Polynesians (see, for Tonga: Haugen and Philips 2010; Sāmoa: Duranti 1992; Tahiti: Ahnne 1926; Peltzer 1994), and were usually used to mark respect for the aristocracy or religious leaders. The most noted of these systems for lexical modification, as also described by Stokes (322–329), were the “chief’s language” in Sāmoa and the Tahitian *pi’i* system.²⁴ Both systems are practices for expressing respect for chiefs and gods. In Sāmoa, it was a system of deference, where certain terms and metaphoric expressions were coined for use only in reference to chiefs. This created a more formal speech register. For example, a chief is not “sick” (although perhaps seriously ill) but is “indisposed,” “weary,” “turned aside,” “wrapped in covering” and so forth...[a king] does not “wake” (ala) but does maleifua, perhaps “emits a cough” (Newell 1911:89, cited in Stokes 1955:322). In the Tahitian *pi’i*, certain words or sounds that were similar to the names of chiefs or gods were not allowed to be used by the general public. A term or even a syllable that was part of a chief’s name could be prohibited for use in regular speech or to refer to anything other than the chief himself. As a result, new words were coined (by the royal family) or borrowed from other nearby languages (Ahnne 1926; Stokes 1955:324) to replace the chiefly sounding terms and syllables throughout the language.

Stokes’s hypothesis was that there was “no doubt” a similar system of language restriction in Rapa Iti. Local oral traditions, however, do not support this idea. Stokes (1955:326) reported that the Rapa royals he spoke with denied such a system existing in Rapa Iti. I, too, was unable to confirm with any elder consultants that such a system ever existed in Rapa. I am nonetheless inclined to agree with Stokes’s hypothesis, due to Rapa Iti’s history of clan division (Hanson 1970; Hanson and Ghasarian 2007; Stokes 1930). Oral history (Stokes 1930; pers. comm. with Rapa Iti elders) indicates that there were at one time twelve different, and opposing, clans (*kōpū*) in Rapa Iti. Each of these clans had claim to a fort (*pāre*) and the valley land below the *pāre*. According to multiple legends, Rapa’s clans were in constant conflict, each trying to appropriate more land and resources from the others. Archaeological research also offers evidence of a warring culture in Rapa Iti. Anderson et al. (2012:253) wrote of clan warfare:

It first becomes apparent archaeologically with the establishment of the Noogorupe and Ruatara fortifications between AD 1300 and 1400, the new need for defensive architecture implying the beginning of stronger status rivalry between competing polities than existed at the time of initial colonisation or developed soon afterward. If it is accepted, as ethnographic data suggest, that the flat-topped towers at the centres of large forts were places of chiefly habitation, then competing chiefly polities existed on Rapa by the 18th century. *Pare* at that time tended to have more defensive features (e.g. Morongo Uta, Potaketake, Kapitanga), some of which cut through existing architecture, suggesting improvements to defensive structure. The higher elevation refugia sites (Ngapiri and Pukumia) also suggest increased warfare late in the Rapan sequence. The overall trend in fort construction, from two in the 14th century, gradual increases into the 17th century and an accelerated burst through the 18th century,

²³ While Stokes does not provide much evidence, he was correct in his assumption. I have found reflexes of some PPN basic vocabulary in place names and in people’s names in Rapa Iti indicating that these terms may have at one time been part of the lexicon, perhaps prior to the language innovations. Some examples are: *Tevaitau* lit. ‘the fresh-water fight’, where two clans supposedly fought over a claim to a fresh water source; *Teumukopuki* lit. ‘the oven place for children’, where (according to legend) a cannibalistic giant cooked children; and *va’ine*, used in the married names of some elder women.

²⁴ Stokes refers to the Tahitian system as *pi*, but the name is actually *pi’i*, meaning ‘to call’ in Tahitian.

suggests that conflict and the threat of war increased through the sequence. The most likely reasons for this increase were either direct population growth or indirect population pressure on resources, such as agricultural land.

The leaders of Rapa's multiple clans, in trying to assert authority and negotiate space on such a small and crowded island, may have used language restrictions to create socio-political boundaries and clan demarcations. Perhaps as the population became more unified, the unique vocabulary of the more powerful clans persisted.

4.3.2. Evidence of Old Rapa's unique vocabulary in other PN languages

Regardless of why Rapa's innovative vocabulary may have been coined, its existence is important to investigate. Due to the divergence of Rapa's innovative vocabulary from Proto-Polynesian and Proto-Eastern Polynesian, attestations of similar forms in other Polynesian languages provide convincing evidence for pre-historic relationships with other island communities. The languages that share some of Rapa Iti's lexical or semantic innovations are the languages of Rapa Nui, the Southern Cook Islands, Mangareva, and Mangaia. Rapanui shows evidence of the semantic innovation of *kakona* 'sweet smelling', as well as the form innovations *matu* 'to advance', *poki* 'child', and *honi* 'peel' (Rapanui consultants).²⁵ Among the languages of the Cook Islands, Penryhn demonstrates three shared semantic innovations with Old Rapa: *kona* meaning 'sweet', *taha rua* 'person', and *kōpū tangata* 'family, relative'; Manihiki shows one semantic innovation *kōpu tangata* 'extended family'.²⁶ Rarotongan shows evidence of the following innovated terms: *ngake* 'a small scoop net with a handle'; *mokotua* 'back'; *mānea* 'beauty, agreeable to sight'; and *kōpū* 'family' (Buse 1995). Mangarevan shares several innovations: *koko* 'valley', *noti* 'indeed', *roupaka* 'small leaves of taro for eating', *matu* 'go follow'; *rua* 'spouse, partner in a couple' and *kami'a* 'canoe' or 'trunk hallowed out to make an outrigger'; and two possibly related innovations: *ko'otu* 'rocky extremity' and *tīkoni* 'clubfoot' (Mangarevan consultants, 2013 and 2015; Janeau 1908; Tregear 1899). The language of Mangaia has the greatest number of shared lexical innovations with Rapa Iti.²⁷

Table 9: Mangaia's shared innovations with Rapa Iti

back	<i>moko</i>
canoe	<i>kami'a</i>
children	<i>puke'anga</i>
<i>Cordyline terminalis</i>	<i>karokaro</i>
family, clan	<i>kōpū</i> (Walter and Reilly 2010)
fire plow	<i>'ikā'ia</i>
forest	<i>raro rākau</i>
fresh water	<i>kota'e</i>
fresh water source	<i>koringiringi</i>
generation, divide	<i>kopanga</i>
immediate family; household	<i>puke'anga</i>
many	<i>ngare</i>
name of a wind	<i>makiki</i>
nose	<i>pita'u</i>
old person	<i>inaina</i>
oven	<i>kauatu</i>
parent	<i>karakua</i> ²⁸

²⁵ *Honi* (RN) and *'oni* (OR) appear to be retentions of PEP 'incise, cut into' and a subsequent semantic narrowing to 'peel' from PEP 'cut into'.

²⁶ Taken from *The Dictionary of the Cook Islands Languages*.

²⁷ All terms were taken from the Mangaian Dictionary Project's online database between Nov. 2013 and Jan. 2014, unless otherwise noted.

²⁸ *Karakua* appears only in the context of adoptive parents or in-laws.

pretty	<i>mānea</i>
skinny	<i>mokōī</i>
small	<i>kakaio</i>
small (for animal)	<i>kororio</i>
small fishing net	<i>ngake</i>
small taro bundles	<i>tīromi</i>
stone, rock	<i>koni'i</i> ('weapon of stone')
sweet-smelling	<i>kakona</i>
taro	<i>mikaka</i>
taro leaf	<i>paka</i>
taro species	<i>'ara'ara</i>
tickle	<i>ketekete</i>
to come and go	<i>naku</i>
to go	<i>matu</i>
to learn	<i>'aikete</i>
to lie down	<i>komo</i>
to see	<i>nokoia; nokoroa</i> ²⁹
tribe	<i>vaka</i> (Walter and Reilly 2010)
wet taro-bed	<i>roki</i>
wild taro	<i>matae</i>
woman	<i>pe'ā</i>

4.4. Mangaian and Old Rapa

Sections 4.1–4.3 outlined some of Old Rapa's divergent features and showed the results of a typological investigation of similar phenomena in other Polynesian languages. The results of this investigation suggest a strong relationship between Mangaian and Old Rapa. Uniquely shared innovations as well as identical consonant reflexes from PEP indicate that these two languages may be more closely related to each other than to the rest of the languages in the EP group. Their identical sound correspondences alone signal shared development, but what is perhaps more convincing are their shared grammatical features and extensive shared basic vocabulary innovations. The nature of the shared features between Mangaian and Old Rapa points to an especially close relationship between the two languages.

The shared innovations between these two languages suggest that there may have been a direct settlement from one of these islands to the other. Recent archaeological dates for initial settlement of Rapa Iti are around 1200 AD (Kennett et al. 2006, 2012:196, 201), with a marked increase in population around 1400 AD (Kennett et al. 2012:201). Mangaia's settlement appears to be slightly earlier, between 1040 and 1220 (Walter and Reilly 2010). The periods of settlement for both islands overlap, meaning that it is very unlikely that there was direct settlement from one island to the other. Thus, the development of shared linguistic features is unlikely to have occurred out of a direct settlement scenario. This, however, does not discount the possibility that the two languages still may have shared an original source. Given the settlement dates, I propose that the shared features of the Rapa and Mangaian languages developed out of prolonged contact rather than through stationary isolation in a homeland and subsequent migration from that homeland. This kind of continued contact would have facilitated the development and sharing of linguistic features between the two speech communities. Continued contact into later periods of Rapa Iti's clan divisions and fort developments would also have allowed for innovations to have been exchanged between the two islands.

Further support for a close relationship between Mangaian and Old Rapa comes from lexical and historical attestations of similar, and unique, social structures. First, the terms *vaka* 'clan' and *kōpū* 'family' have undergone a significant semantic change in both languages. These changes are

²⁹ *Nokoia* is likely *noko'ia*, with a passive suffix. *Nokoroa* is likely *noko roa* 'to see far'. The base-word is thus the same as Rapa Iti's *nook* 'to see'.

significant as they represent how early Mangaian and Rapa Iti populations may have been organized under their *rangatira* 'chief', as well as how they were divided. The use of these terms in identical ways in both languages points to a shared system of clan division and social stratification. Additionally, both Mangaia and Rapa Iti have stories of women warriors, something unique in Polynesian history. High-ranking women and chiefly women were certainly not uncommon in many parts of Polynesia (Gunson 1987); however, warrior women are extremely rare. In both Rapa Iti and Mangaia, however, women warriors seem to have been commonplace, perhaps pointing to a shared social structure between the two island communities. On Rapa Iti, there are two large, erected slabs of rock at opposing ends of the large A'urei Bay. According to a local historian, the taller of the two was to measure boys for war. The second and shorter stone was to measure girls for war. In both cases, if a child's shoulders reached the height of the stone, he or she was ready for battle. This same local historian stated that Rapa had women warriors who were in charge of guarding their clan's taro beds. These women were called *irari*. Reilly (2001) describes women in Mangaia also fighting in lines of battle. He wrote, "Women were clearly capable fighters who worked in a complementary wartime partnership with their husbands" (2001:160).

5. A South Polynesian contact sphere

Based on the linguistic evidence alone, the relationship between Old Rapa and Mangaian reflects both shared inheritance and maintained contact. Rapa Iti and Mangaia share an original source, and through continued waves of contact between the two speech communities, interaction was maintained to the point where these communities were participating in each other's political and social systems. A localized contact sphere persisted between these two islands. Additionally, based on the shared features between Old Rapa and other languages, I hypothesize that their local contact sphere was only one part of a larger contact network that stretched across southeast Polynesia, from the Southern Cooks to Rapa Nui, including Rapa Iti and Mangareva.

If Rapa Iti and Mangaia were involved in a two-way interaction sphere, it follows that people from Rapa Iti would have been voyaging to Mangaia. Due to the close proximity of the Southern Cooks, it is not unlikely that these groups were also interacting with people from Rapa Iti. Linguistically, this is demonstrated in the shared features between Old Rapa and other Southern Cooks languages. Rarotongan, for example, demonstrates identical consonant reflexes from PPN, and shares many of Old Rapa's grammatical innovations, as well as a handful of Old Rapa's lexical innovations. Other languages of the Southern Cooks are not well documented enough to investigate grammatical and lexical correlations. However, it is certain that consonant reflexes from PPN for at least Ma'uke and Aitutaki are also identical to those of Old Rapa. Furthermore, as discussed in section 4.2.2, Old Rapa shares the unique grammatical marker *tuai* with Niuean. Niue is not part of the Southern Cooks, but, as previously stated, Niuean borrowed extensively from EP languages via contact with the Southern Cooks languages. If Rapa Iti voyagers were regularly involved in a Southern Cooks contact sphere, they may have had contact with Niuean speakers as well, leading to the incorporation of this shared item into Old Rapa.

I have also noted shared linguistic features between Old Rapa and Rapanui, as well as Old Rapa and Mangarevan. These shared features are not as extensive as those between Old Rapa and Mangaian, but they cannot be ignored as evidence for at least some sort of isolated language contact, if not a period of shared development. Moreover, many of these shared features overlap and are shared between all three languages. Lexically, Rapanui, Mangareva, and Old Rapa do not show as compelling evidence for subgrouping as do Mangaian and Rapa Iti; however, the identical sound changes that occurred from PEP **taua* may lend credence to the idea of a period of shared development between these three languages as well.

6. Conclusions

Old Rapa is a Central Eastern Polynesian language that demonstrates a very close affinity with Mangaian through sound correspondences, sporadic sound changes, shared grammatical innovations, and a number of shared lexical innovations. If the archaeological dates are accurate, the time-depth

from the settlement of Mangaia to the initial settlement of Rapa Iti was likely not great enough to have allowed for their complex shared developments to have occurred. Thus, it is unlikely that Rapa Iti was settled in one pulse from Mangaia. It is more likely that Mangaia and Rapa Iti share an original source, and that the shared linguistic features between the two languages spoken on these islands were developed within a micro-contact sphere. This contact network became part of a larger interaction network with the other Southern Cook Islands, Mangareva, and possibly even Rapa Nui, wherein linguistic features were exchanged and possibly developed. At some point, the smaller spheres of contact ceased to exist, and Rapa Iti remained isolated until Western contact in the nineteenth century. Support for this proposal of isolation comes from Old Rapa's truly distinctive features. These include the marker for past negative *ki'ere* and the use of *ka* as a perfective aspect marker. Perhaps contact stopped due to the aggressive in fighting on Rapa Iti, or perhaps there was simply less of a need to exchange with other islands as later generations became more settled. These reasons are entirely speculative, however, and there is no way of knowing why Rapa Iti's interaction with other islands ended.

There are two wider implications of this prehistoric contact scenario. First, a long-distance contact network in which linguistic features were developed argues against the traditional PCE subgroups, Marquesic and Tahitic. This is perhaps a more minor issue, as evidence for these subgroupings has been proved unsubstantial (see Walworth 2014). Second, a southern contact sphere in which Rapa Nui participated argues against the long-held theory that Rapa Nui was significantly isolated (Fischer 1992; Kirch and Green 2001, among others) during periods of long distance voyaging between all of the other east Polynesian islands. This is critical, as this long period of isolation accounts for the Rapa Nui language's conservative retentions from PEP, and lack of membership in the PCE subgroup. This problem requires a deeper investigation into the language of Rapa Nui, as well as other under-studied languages of south Polynesia. Further research on these languages will offer a clearer picture of historical relationships in the region.

References

- Ahnne, Edouard. 1926. De la coutume du *pi'i* et des modifications qu'elle apporta au vocabulaire tahitien. *Bulletin des études Océaniques* 11:6–10.
- Anderson, Atholl, Douglas Kennett, and Eric Conte. 2012. A Bayesian AMS 14C chronology for the colonisation and fortification of Rapa Island. In *Taking the high ground: the archaeology of Rapa, a fortified island in remote east Polynesia*, 247–256. Canberra: Australian National University.
- Bauer, Winifred, William Parker, and Te Kareongawai Evans. 1993. *Māori*. United Kingdom: Routledge.
- Bauer, Winifred, William Parker, Te Kareongawai Evans, and Te Aroha Noti Teepa. 1997. *The Reed reference grammar of Maori*. Auckland: Reed Books.
- Buse, Jasper. 1995. *Cook Islands Maori dictionary*. Canberra: Australian National University Press.
- Churchill, William. 1912. *Easter Island: The Rapanui speech and the peopling of Southeast Polynesia*. Washington: the Carnegie Institute.
- Clark, Ross. 1976. *Aspects of Proto-Polynesian syntax*. Auckland: Linguistic Society of New Zealand.
- Clark, Ross. 1979. Language. In *The prehistory of Polynesia*, ed. by Jesse D. Jennings, 249–70. Canberra: Australian National University Press.
- The Dictionary of the Cook Islands Languages. <http://cookislandsdictionary.com/> [accessed between November 2013 and January 2014].
- Dordillon, Ildefonse René. 1904. *Grammaire et dictionnaire de la langue des Iles Marquises*. Paris: Imprimerie Belin Freres.
- Duranti, Alessandro. 1992. Language in context and language as context: The Samoan respect vocabulary. In *Rethinking context: Language as an interactive phenomenon*, ed. by Alessandro Duranti and Charles Goodwin, 77–99. Cambridge: Cambridge University Press.

- Ellis, William. 1838. *Polynesian Researches, during a residence of nearly eight years in the Society and Sandwich Islands, Vol III*. London: Fisher, Son and Jackson.
- Fare Vāna'a. 1999. *Dictionnaire Tahitien-Français*. Tahiti: Académie Tahitienne.
- Fischer, Steven Roger. 1992. Homogeneity in Old Rapanui. *Oceanic Linguistics* 31(2):181–190
- . 2000. Preliminary evidence for Proto Austral, excluding Rapan. In *Leo Pasifika: Proceedings of the Fourth International Conference on Oceanic Linguistics*, Niue, 5th-9th July 1999, ed. by Steven Roger Fischer and Wolfgang B. Sperlich, 100-111. Auckland: Institute of Polynesian Languages and Literatures.
- . 2001. Mangarevan doublets: Preliminary evidence for Proto-Southeastern Polynesian. *Oceanic Linguistics* 40(1):112–124.
- Gill, William Wyatt. 1876. *Life in the southern isles*. London: The Religious Tract Society.
- Green, Roger C. 1966. Linguistic subgrouping within Polynesia: The implications for prehistoric settlement. *The Journal of the Polynesian Society* 75(1):6–38.
- . 1985. Subgrouping of the Rapanui language of Easter Island in Polynesia and its implications for East Polynesian prehistory. *Working Papers in Anthropology, Archaeology, Linguistics, Māori Studies* 68.
- . 1988. Subgrouping of the Rapa Nui language of Easter Island in Polynesian and its implications for East Polynesian prehistory. In *First International Congress, Easter Island and East Polynesia*, ed. by C. Cristino F., P. Vargas C., R. Izaurieta S., and R. Budd P., 37–57. Santiago: Instituto de Estudios, Universidad de Chile.
- Greenhill, Simon J., and Ross Clark. 2011. POLLEX-Online: The Polynesian Lexicon Project Online. *Oceanic Linguistics* 50(2):551–559.
- Gunson, Niel. 1987. Sacred women chiefs and female ‘headmen’ in Polynesian history. *The Journal of Pacific History* 22(3):139–172.
- Hale, Horatio. 1846. *United States exploring expedition during the Years 1838, 1839, 1840, 1841, 1842 under the command of Charles Wilkes, U.S.N. ethnography and philology*. Philadelphia: C. Sherman.
- Hanson, Allan. 1970. *Rapan lifeways*. Boston: Little, Brown and Company.
- Hanson, Allan, and Christian Ghasarian. 2007. The land belongs to everyone. *The Journal of Polynesian Society* 116(1):59–72.
- Harlow, Ray. 1989. *Ka: The Māori injunctive*. In *VICAL 1: Papers from the Fifth International Conference on Austronesian Linguistics*, ed. by Ray Harlow and Robin Hooper, 197–210. Auckland: Linguistic Society of New Zealand.
- . 2012. *Māori: A linguistic introduction*. Cambridge: Cambridge University Press.
- Haugen, Jason D. and Susan Philips. 2010. Tongan chiefly language: The formation of an honorific speech register. *Language in Society* 39(5):589–616.
- Janeau, Vincent Ferrier. 1908. *Essai de Grammaire de la langue des Iles Gambier*. Braine-le-Compte: Imprimerie Zech et Fils.
- Kennett, Douglas, Atholl Anderson, Matthew Prebble, Eric Conte, and John Southon. 2006. Prehistoric human impacts on Rapa, French Polynesia. *Antiquity* 80:340–354.
- Kennett, Douglas, Brendan J, Culleton, Atholl Anderson, and John Southon. 2012. A Bayesian AMS 14C chronology for the colonisation and fortification of Rapa Island. In *Taking the high ground: The archaeology of Rapa, a fortified island in remote east Polynesia*, 189–201. Canberra: Australian National University.
- Kieviet, Paulus, and Antje Kieviet. 2006. *Lexique Rapa-Français*. Unpublished manuscript. SIL International.
- Kirch, Patrick V. 2000. *On the road of the winds: an archaeological history of the Pacific Islands before European contact*. Berkeley: University of California Press.

- Kirch, Patrick V., and Roger C. Green. 2001. *Hawaiki, ancestral Polynesia: An essay in historical anthropology*. Cambridge: Cambridge University Press.
- Langdon, Robert, and Darrell Tryon. 1983. *The language of Easter Island: Its development and Eastern Polynesian relationships*. Laie, Hawaii: Institute for Polynesian Studies.
- Lemaître, Yves. 1973. *Lexique du Tahitien contemporain*. Paris: Orstom.
- Mangaian Dictionary. <http://taramangaia.com> [accessed between November 2013 and January 2014].
- Marck, Jeff. 1996. Eastern Polynesian subgrouping today. In *Oceanic culture history: Essays in honour of Roger Green*, ed. by Janet Davidson, Geoffrey Irwin, Foss Leach, Andrew Pawley, and Dorothy Brown, 491–511. Wellington: New Zealand Journal of Archaeology.
- Marck, Jeff. 2000. *Topics in Polynesian language and culture history*. Canberra: Pacific Linguistics.
- Marck, Jeff. 2002. Comments on S.R. Fischer's 'Mangarevan doublets: Preliminary evidence for Proto-Southeastern Polynesian'. *Oceanic Linguistics* 41(1):225–231.
- Newbury, Colin Walter. 1961. *The history of the Tahitian mission, 1799-1830, by John Davies*. Cambridge: Cambridge University Press.
- Otsuka, Yuko. 2006. Niuean and Eastern Polynesian: A view from syntax. *Oceanic Linguistics* 45(2):429–56.
- Pawley, Andrew, and Roger Green. 1974. Dating the dispersal of the Oceanic languages. *Oceanic Linguistics* 12(2):1–67.
- Peltzer, Louise. 1994. Le pi'i. *Bulletin des études Océaniques* 261–262:28–35.
- Pukui, Mary K., and Samuel H. Elbert. 1957. *Hawaiian-English dictionary*. Honolulu: University of Hawai'i Press.
- Reilly, Michael P.J. 1993. A political succession text from Mangaia. *The Journal of the Polynesian Society* 102(1):71–98.
- Reilly, Michael P.J. 2001. Women in Mangaian society. *The Journal of Pacific History* 36(2):149–161.
- Reilly, Michael P.J. 2007. Transforming Mangaia's spiritual world: Letters from the early Christian community of Oneroa. *The Journal of the Polynesian Society* 116(1):35–58.
- Richards, Rhys. 2007. The earliest foreign visitors and their massive depopulation of Rapa-iti from 1824 to 1830. *Le Journal de la Société des Océanistes* 118:3–10.
- Rutter, Lawrence Kenji. 2002. The validity of Proto-Southeastern Polynesian. *Oceanic Linguistics* 41(1):232–237.
- Savage, Stephen. 1962. *A dictionary of the Māori language of Rarotonga*. Wellington: The Department of Island Territories.
- Schooling, Stephen. 1981. *A linguistic and sociolinguistic survey of French Polynesia*. SIL International.
- Seiter, William J. 1980. *Studies in Niuean syntax*. New York: Garland Publishing.
- Smith, Percy S. 1910. Easter Island (Rapa-Nui) and Rapa (Rapa-Iti) Island. *The Journal of the Polynesian Society* 19(4):171–175.
- Stimson, J. Frank, and Donald Stanley Marshall. 1964. *A dictionary of some Tuamotuan dialects of the Polynesian language*. Salem, MA: The Peabody Museum of Salem.
- Stokes, John F.G. 1930. Ethnology of Rapa Island. Unpublished manuscript on file in the Bishop Museum, Honolulu, five volumes.
- Stokes, John F.G. 1955. Language in Rapa. *Journal of the Polynesian Society* 64:315–340.
- Stutchbury, Samuel. 1996. *Science in a Sea of Commerce: The Journal of a South Seas Trading Venture (1825-27)*. Reprint by David Branagan. Sydney: Oxford University Press.
- Tregear, Edward. 1899. *A dictionary of Mangareva*. Wellington: The New Zealand Institute.

- Vancouver, George. 1798. *A voyage of discovery to the North Pacific Ocean*, volume 1. London: Paternaster-Row.
- Walter, Richard, and Michael Reilly. 2010. A prehistory of the Mangaian chiefdom. *The Journal of the Polynesian Society* 119(4):335–375.
- Walworth, Mary. 2014. Eastern Polynesian: The linguistic evidence revisited. *Oceanic Linguistics* 53(2):257–273.
- Walworth, Mary. 2015. *The language of Rapa Iti: Description of a language in change*. PhD dissertation. Honolulu: University of Hawai‘i at Mānoa.
- Wilson, William H. 1982. *Proto-Polynesian possessive marking*. Canberra: Pacific Linguistics.
- Wilson, William H. 2010. A Tuamotuan challenge to the subgrouping of Hawaiian. In *A journey through Austronesian and Papuan linguistic and cultural space: Papers in honour of Andrew K. Pawley*, ed. by John Bowden, Nikolaus P. Himmelmann, and Malcolm Ross, 291–318. Canberra: Pacific Linguistics.
- Wilson, William H. 2012. Whence the East Polynesians? Further linguistic evidence for a Northern Outlier source. *Oceanic Linguistics* 51(2):289–359.