

TRACES OF CONTACT
IN THE LEXICON

Austronesian and Papuan Studies

Edited by
Marian Klamer
Francesca R. Moro

BCD
SD

BRILL

Traces of Contact in the Lexicon

Brill Studies in Language Contact and the Dynamics of Language

Series Editor

Robert Nicolai (*University of Nice Sophia Antipolis, Nice, France*)

Editorial Board

Umberto Ansaldo (*Curtin University, Perth, Australia*)

Peter Auer (*Albert Ludwig University of Freiburg, Freiburg, Germany*)

Marianne Mithun (*University of California, Santa Barbara, USA*)

Patrick Seriot (*University of Lausanne, Lausanne, Switzerland*)

VOLUME 4

The titles published in this series are listed at brill.com/bscd

Traces of Contact in the Lexicon

Austronesian and Papuan Studies

Edited by

Marian Klamer
Francesca R. Moro



BRILL

LEIDEN | BOSTON



This is an open access title distributed under the terms of the CC BY-NC-ND 4.0 license, which permits any non-commercial use, distribution, and reproduction in any medium, provided no alterations are made and the original author(s) and source are credited. Further information and the complete license text can be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

The terms of the CC license apply only to the original material. The use of material from other sources (indicated by a reference) such as diagrams, illustrations, photos and text samples may require further permission from the respective copyright holder.

This volume is an output of the VICI Grant Research Project *Reconstructing the past through languages of the present: the Lesser Sunda Islands* (project number # 277-70-012) funded by the Dutch Research Council (NWO) and awarded to Marian Klamer, 2014–2019.

The Library of Congress Cataloging-in-Publication Data is available online at <http://catalog.loc.gov>
LC record available at <http://lccn.loc.gov/2022052554>

Typeface for the Latin, Greek, and Cyrillic scripts: “Brill”. See and download: brill.com/brill-typeface.

ISSN 2214-5613

ISBN 978-90-04-52893-2 (hardback)

ISBN 978-90-04-52945-8 (e-book)

Copyright 2023 by Marian Klamer and Francesca R. Moro. Published by Koninklijke Brill nv, Leiden, The Netherlands.

Koninklijke Brill nv incorporates the imprints Brill, Brill Nijhoff, Brill Hotei, Brill Schöningh, Brill Fink, Brill mentis, Vandenhoeck & Ruprecht, Böhlau, V&R unipress and Wageningen Academic. Koninklijke Brill nv reserves the right to protect this publication against unauthorized use.

This book is printed on acid-free paper and produced in a sustainable manner.

Contents

Acknowledgements VII
List of Figures and Tables VIII
Notes on Contributors XIV

- 1 Lexical Borrowing in Austronesian and Papuan Languages: Concepts, Methodology and Findings 1
Marian Klamer and Francesca R. Moro

PART 1

Ancient and Pre-modern Contact

- 2 Lexical Influence from South Asia 25
Tom G. Hoogervorst
- 3 Traces of Pre-modern Contacts between Timor-Alor-Pantar and Austronesian Speakers 57
Marian Klamer
- 4 Phonological Innovation and Lexical Retention in the History of Rote-Meto 101
Owen Edwards
- 5 The Mixed Lexicon of Lamaholot (Austronesian): A Language with a Large Lexical Component of Unknown Origin 140
Hanna Fricke
- 6 Entwined Histories: The lexicons of Kawaimina and Maka Languages 180
Antoinette Schapper and Juliette Huber

PART 2

Modern and Contemporary Contact

- 7 Detecting Papuan Loanwords in Alorese: Combining Quantitative and Qualitative Methods 213
Francesca R. Moro, Yunus Sulistyono and Gereon A. Kaiping
- 8 Multilateral Lexical Transfer among Four Papuan Language Families: Border, Nimboran, Sentani, and Sko 263
Claudia Gerstner-Link
- 9 Spanish Suffixes in Tagalog: The Case of Common Nouns 307
Ekaterina Baklanova and Kate Bellamy
- 10 The Structural Consequences of Lexical Transfer in Ibatan 348
Maria Kristina S. Gallego
- 11 The Effects of Language Contact on Lexical Semantics: The Case of Abui 392
George Saad
- Index 421

Acknowledgements

This volume is an output of the VICI Grant Research Project *Reconstructing the past through languages of the present: the Lesser Sunda Islands* (project number # 277-70-012) funded by the Dutch Research Council (NWO) and awarded to Marian Klamer, 2014–2019. The data collection and analyses presented in the five chapters by Klamer, Edwards, Fricke, Moro et al. and Saad have been supported by the VICI project.

The chapters in this volume are a selection of the papers that were presented at the workshop *Lexical traces of contact*, which was organised as part of the 11th Conference on Austronesian and Papuan Languages and Linguistics (APLL11), held in Leiden, The Netherlands, June 13–15 2019.

Each chapter has been reviewed single-blind by at least two reviewers, both internal and external. The volume editors are grateful to the following colleagues for providing reviews and helpful comments on the various chapters (in alphabetical order): Sander Adelaar, Laura Arnold, Owen Edwards, Alex Elias, Bethwyn Evans, Hanna Fricke, Kristina Gallego, Arlo Griffiths, Tom Hoogerworst, Juliette Huber, František Kratochvíl, Naonori Nagaya, Maria del Carmen Parafita Couto, Laurie Reid, George Saad, and Jiang Wu.

The Open Access publication of this volume was made possible through an Open Access book publication grant from the Dutch Research Foundation NWO, and a grant from the University of Insubria, Italy.

Figures and Tables

Figures

- 1.1 Locations of languages or language areas discussed in the chapters of this volume, by their chapter number 17
- 3.1 Location of the Timor-Alor-Pantar languages in Indonesia 57
- 3.2 The Timor-Alor-Pantar languages 58
- 3.3 The three subbranches of the Timor-Alor-Pantar family (Holton et al. 2012; Holton and Robinson 2017b; Schapper et al., 2017) 59
- 3.4 The MP subbranches of AN languages on Timor (Edwards 2018b; 2019; 2020a; 2020b) 63
- 4.1 Domains and Languages of Rote Island 103
- 4.2 Self identified Varieties of Meto 104
- 4.3 Proto Rote-Meto strata 106
- 4.4 Frequencies of majority reflexes 128
- 5.1 The Flores-Lembata languages and their linguistic context 143
- 5.2 The Flores-Lembata languages and their genealogical affiliation 143
- 6.1 The relations of the Papuan languages of the Timor-Alor-Pantar family 182
- 6.2 Linguistic map of Timor and surrounds (adapted from Schapper 2020b) 183
- 7.1 Alorese spoken on Alor and Pantar 215
- 7.2 Genealogical classification of Alorese 216
- 7.3 The 13 Alorese varieties spoken on Alor and Pantar, labeled according to village name 217
- 8.1 Language map 264
- 8.2 Putative migration routes of the Kilmeri, Nimboran, Sentani, and Sko people 297
- 10.1 The location of Ibatan 350
- 11.1 Map of indigenous languages of Alor and Pantar 395
- 11.2 Early language exposure among the three Light Abui groups 395
- 11.3 Proportion of mismatches for 'see', 'fall over', 'wake up' 412

Tables

- 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set 7
- 1.2 Lexical borrowing in Thomason's borrowing scale (based on Thomason 2001: 70–71) 11

1.3	Contact settings and lexical borrowing in the contributions of this volume	14
2.1	Rejected South Asian borrowings in Malay	27
2.2	MIA and NIA loans in Malay and Javanese	30
2.3	The preference for i-forms in Sanskrit loanwords	30
2.4	Possible i-forms of MIA loanwords	31
2.5	NIA loanwords displaying the suffix -ī	32
2.6	Indo-Aryan loanwords introduced by speakers of Tamil	32
2.7	Diachronic shifts in meaning	34
2.8	Semantic shifts in loanwords transmitted through Malay	34
2.9	Phonological integration in Leti and Rote	35
2.10	Sanskrit words often found as loanwords in Maritime Southeast Asia	36
2.11	South Asian loans in Cham borrowed through Khmer	37
2.12	Sanskrit loans substituting inherited vocabulary in Malay	38
2.13	Proto Moklenic reconstructions borrowed from Indo-Aryan languages	39
2.14	Tamil loanwords in Old Javanese	39
2.15	Tamil loanwords in Austronesian languages	40
2.16	Malayalam loanwords in Austronesian languages	40
2.17	South Asian loanwords found in Acehnese but not in other Austronesian languages	42
2.18	Tamil loanwords in Karo Batak	42
2.19	Sanskrit loanwords found in Toba Batak but not in other Austronesian languages	43
2.20	South Asian loanwords in Nias	44
2.21	Sanskrit loanwords in Tagalog	45
2.22	Sanskrit loanwords in Malagasy	45
2.23	South Asian loanwords in languages of Borneo	46
2.24	South Asian loanwords in Ternate and Galela	47
2.25	South Asian loanwords in Biak	47
2.26	South Asian loanwords in Tetun	48
2.27	South Asian loanwords featuring Malay sə-	49
2.28	Indo-Aryan loanwords not found in Malay	50
4.1	Cognate sets shared between Rote-Meto and Helong	107
4.2	Proto Malayo-Polynesian consonants	111
4.3	Proto-Rote-Meto consonants	112
4.4	PRM consonants by strata	114
4.5	PRM vowels by strata	115
4.6	Implosives and prenasalised plosives #_	116
4.7	Implosives and prenasalisation according to strata	117
4.8	Reflexes of voiced coronals in PRM	119

- 4.9 Potential coinages and borrowings in strata 121
- 4.10 Comparison of basic vocabulary and entire lexicon 122
- 4.11 PRM reconstructions by semantic fields 123
- 4.12 Regular and irregular sound changes 125
- 4.13 PRM consonants undergoing unconditioned splits 127
- 4.14 PMP nasal-stop clusters 129
- 4.15 Reflexes of PMP *b in PRM 130
- 4.16 Examples PMP *b > PRM *f, *b, *b, *mb 130
- 4.17 PMP *wa > PRM *o, *wa, fa 132
- 4.18 Lole assimilation of voiced plosives 135
- 5.1 Non-IPA symbols in PMP forms in Blust and Trussel (2010) and in this chapter 146
- 5.2 Two types of lexeme sets in the Flores-Lembata languages 147
- 5.3 Proto Flores-Lembata vowel inventory 149
- 5.4 Proto Flores-Lembata consonant inventory 149
- 5.5 Attested sound changes in the Flores-Lembata subgroups 150
- 5.6 PFL reconstructions with PMP source (n=173) 152
- 5.7 PFL reconstructions without PMP source (n=37) 158
- 5.8 AN and non-AN lexemes in individual varieties of the Flores-Lembata subgroups 161
- 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185) 162
- 5.10 Regional comparison of AN and non-AN components of the lexicon (Elias 2020, 331; Edwards this volume; Fricke 2019c, 248–249) 170
- 5.11 Parallel lexemes in Central Lembata Lamaholot (Fricke 2019a) 173
- 6.1 TAP etyma in Kawaimina languages 186
- 6.2 ET etyma in Kawaimina languages 190
- 6.3 Selected reflexes of #kafo ‘eight’ across eastern Timor and southern Maluku 193
- 6.4 Shared Maka-Kawaimina lexicon 194
- 6.5 Shared Makasae-Kawaimina lexicon 196
- 6.6 Examples of etyma borrowed through Makasae 199
- 6.7 Body part terms marked with *-kai* in Makasae, Waima’a and Naueti 201
- 6.8 Human nouns with *-kai* 204
- 7.1 Relevant patterns of distribution of lexically similar forms in languages of the region, and the corresponding borrowing or inheritance history of such a form. 220
- 7.2 Examples of a lexical similarity set associated with the concept ‘to breathe’ generated using automatic comparison. 221
- 7.3 Examples of a meaning mismatch due to word order. 222

7.4 Lexical similarity set associated with the concept ‘fish trap’. 223

7.5 Lexical similarity set associated with the concept ‘bed’. 224

7.6 Lexical similarity set associated with the concept ‘to fold’. 224

7.7 Lexical similarity set associated with the concept ‘to pull’. 225

7.8 Lexical similarity set associated with the concept ‘to wash’. 226

7.9 Lexical similarity set associated with the concept ‘to pray’. 227

7.10 Lexical similarity set associated with the concept ‘adultery’. 227

7.11 Lexical similarity set associated with the concept ‘digging stick’. 228

7.12 Lexical similarity set associated with the concept ‘garden’. 229

7.13 Lexical similarity set associated with the concept ‘rattan’. 230

7.14 Lexical similarity set associated with the concept ‘root’. 231

7.15 Lexical similarity set associated with the concept taro. 232

7.16 Lexical similarity set associated with the concept ‘coral rock’. 233

7.17 Lexical similarity set associated with the concept ‘mud’. 234

7.18 Lexical similarity set associated with the concept ‘gravel’. 235

7.19 Lexical similarity set associated with the concept ‘dolphin’. 236

7.20 Lexical similarity set associated with the concept ‘monitor lizard’. 237

7.21 Lexical similarity set associated with the concept ‘ten’. 239

7.22 Lexical similarity set associated with the concept ‘angry’. 241

7.23 Lexical similarity set associated with the concept road. 242

7.24 Lexical similarity set associated with the concept ‘younger sibling’. 243

7.25 Lexical similarity set associated with the concept ‘to bury’. 244

7.26 Lexical similarity set associated with the concept ‘heart’. 246

7.27 Lexical similarity set associated with the concept ‘to breathe’. 247

7.28 Lexical similarity set associated with the concept ‘small’. 248

7.29 Lexical similarity set associated with the concept ‘to close’. 249

7.30 Lexical similarity set associated with the concept ‘to hide’. 250

7.31 Lexical similarity set associated with the concept ‘dirty’. 252

7.32 Distribution of loanwords based on geographic groups. 253

7.33 Alorese varieties with their number of loanwords. 254

8.1 The Border family including Elseng 265

8.2 The Nimboran family (Foley 2018:446) 267

8.3 The Sentani family (Foley 2018:438) 267

8.4 The Sko family (Foley 2018:399; Donohue 2004:16; 18) 269

8.5 ‘water, rain, river’ 288

8.6 ‘tree, wood’ 289

8.7 ‘leaf’ 290

8.8 ‘arrow’ 291

8.9 Identified transfers in numbers and word class 292

9.1 Tagalog native derivation of the class ‘names of persons and objects’ 315

9.2	Characteristics of nouns with <i>-ero/a</i> in the Tagalog dataset of 1900s	318
9.3	Characteristics of hybrid formations with <i>-ero/a</i> in contemporary Tagalog	320
9.4	Semantic distribution of CLs and HF s with <i>-ero/a</i> in contemporary Tagalog	321
9.5	Characteristics of nouns with <i>-ista</i> in the Tagalog historical dataset	322
9.6	Characteristics of nouns with <i>-ista</i> in the contemporary Tagalog dataset	324
9.7	Comparison of native and Spanish strategies of agentive derivation in Tagalog	327
9.8	Characteristics of nouns with <i>-ito/a</i> and <i>-ilyo/a</i> (<i>-illo/a</i>) in the historical dataset	329
9.9	Characteristics of nouns with <i>-ito/a</i> and <i>-ilyo/a</i> in contemporary Tagalog	332
9.10	Characteristics of nouns with <i>-ete</i> in contemporary Tagalog	334
9.11	Scale of directness of affix borrowing (Seifart, 2015: 527, fig. 3)	335
9.12	Summary of distribution of agentive <i>-ero/a</i> in Tagalog historical data	336
9.13	Hybrids with <i>-ero/a</i> in contemporary Tagalog. Pattern adopted from Bakker and Hekking (2012: Table 7)	339
10.1	Native and non-native affixes and stems (Maree 2007:173)	354
10.2	Ilokano verbal morphology	357
10.3	Loanwords from different source languages occurring with <i>mag-</i>	360
10.4	Ibatan verbal morphology for native stems	361
10.5	Ibatan verbal morphology for non-native stems	362
10.6	Distribution of durative formations indicated in the Ibatan dictionary by Maree et al. (2012)	364
10.6	Distribution of durative formations indicated in the Ibatan dictionary by Maree et al. (2012)	364
10.8	Restricted distribution of the non-native durative paradigm vis-à-vis the native paradigm	366
10.9	Further morphological derivations involving the durative paradigms	367
10.10	Pairs of near-homophonous native and non-native forms in Ibatan	371
10.11	Forms that involve native and non-native prefixes in free variation	372
11.1	Age-groups used in this study	396
11.2	Breakdown of participants	398
11.3	Coding of event domains	398
11.4	Verbs of perception	400
11.5	Frequency of visual perception verbs (Kratochvíl corpus)	401
11.6	Proportion of mismatches for <i>-ien-</i> 'see [-CONTROL]' target	401
11.7	Verbs of falling	403

- 11.8 Frequency of falling verbs (Kratochvíl corpus) 404
- 11.9 Production data: Proportion of mismatches for *-quoil/-kaai* 'fall over
[-ELEVATION]' target 405
- 11.10 Change of state verbs 407
- 11.11 Frequency of change of state verbs (Kratochvíl corpus) 409
- 11.12 Production data: Proportion of mismatches for *tein/-minang* 'wake up
[-CHANGE OF POSTURE]' target 410

Notes on Contributors

Ekaterina Baklanova

is Senior Research Fellow in Linguistics at the Institute of Asian and African Studies of Lomonosov Moscow State University, where she also obtained her PhD in 2009. Her major scholarly interests concern Tagalog/Filipino linguistics, particularly language contact and interference, morphology and corpus studies. Her research activities also include History of the Philippine Literature, Contemporary Philippine Literature and Malay Linguistics. She teaches Tagalog/Filipino Language, Linguistics and the Philippine Literature at the IAAS of MSU. She has written about lexical and structural borrowing from Asian and European languages, contact-induced changes and various morphological features in Tagalog, as well as about the Philippine literature in multiple journal articles and book chapters.

Kate Bellamy

is a Postdoctoral Researcher at Leiden University (The Netherlands), where she also obtained her PhD in 2018. Her research focuses on code-switching and multilingualism, particularly the role of grammatical gender in mixed nominal constructions. She investigates these topics, as well as questions of lexical semantics and word formation particularly in relation to P'urhepecha, a language isolate spoken in Michoacán (Mexico). Her code-switching research also extends to Tsova Tush-Georgian and Spanish-English contact situations. She has recently published articles in the *International Journal of Bilingualism*, and *Languages*, and book chapters in *The Acquisition of Gender: Crosslinguistic Perspectives*, and *The Linguistics of Olfaction* (both John Benjamins).

Owen Edwards

is a postdoctoral researcher at Leiden University. He studied linguistics at the Australian National University during which he carried out fieldwork on several languages of Indonesia. His linguistic interests include morphology, phonology, historical linguistics, and Austronesian linguistics.

Hanna Fricke

earned her PhD in Linguistics (2019) from the University of Leiden in the Netherlands. Her research interests lie in descriptive linguistics, typology, historical linguistics and language contact with a focus on Austronesian languages in eastern Indonesia. She has taught courses on language description and linguistic field methods. Currently she is a data steward for the Faculty of Humanities at the University of Amsterdam in the Netherlands.

Maria Kristina Gallego

is an Assistant Professor at the Department of Linguistics, University of the Philippines Diliman. She finished her PhD at the School of Culture, History and Language at the College of Asia and the Pacific, Australian National University in 2022. She is currently working on the documentation of Ibatan, a language spoken in Babuyan Claro, Philippines, through an Individual Graduate Scholarship funded by the Endangered Languages Documentation Programme. She has presented and published various papers on language contact and change, focusing on the Austronesian languages of the Philippines and Taiwan.

Claudia Gerstner-Link

Dr. phil. habil., Privatdozentin at Munich University, is a semanticist and typologist pursuing a semantically based approach to linguistics and language documentation. Her dissertation was on generics and habituals. Her interests in semantics include: role-based grammatical relations, lexical semantics, metaphoric language use in indigenous languages; orientation in space, orientation in time. These themes were investigated in her comprehensive typological case study *A Grammar of Kilmeri* (2018), which draws on extensive fieldwork on the Papuan language of Kilmeri. Her current projects include an online publication of a dictionary of Kilmeri in the open access series “Dictionaria” at Leipzig University.

Tom Hoogervorst

is a senior researcher at the Royal Netherlands Institute of Southeast Asian and Caribbean Studies. His research interests include historical linguistics, language contact, lexical borrowing, youth languages, and language history. Most of his publications focus on Malay and/or Javanese. Among his publications are the monographs *Southeast Asia in the Ancient Indian Ocean World* (Archaeopress 2013) and *Language Ungoverned: Indonesia's Chinese Print Entrepreneurs, 1911–1949* (Cornell University Press 2021), the co-edited volume *Sinophone Southeast Asia: Sinitic Voices across the Southern Sea* (Brill 2021), the co-edited special edition *Languages of Nusantara* (Wacana 2021), and several articles and book chapters on loanwords in Southeast Asia and the Indian Ocean World.

Juliette Huber's

research to date has focused on the Eastern Timor languages of the Timor-Alor-Pantar family, in particular Makalero. She has published on various aspects of grammar, landscape conceptualization and spatial language, as well as the

historical relations between those languages. She is currently studying the endangered Hakka varieties of the Chinese-descended communities of Timor.

Gereon Kaiping

is a Postdoctoral researcher in GIScience at the University of Zürich. After a PhD in Computational Engineering and Design from the University of Southampton (UK), he now works on the methodological side of inter-disciplinary research in language diversity. His interest is to develop tools for linguistic data and computational models for phylogenetics in order to capture and infer various processes of linguistic change, population dispersal, and language contact.

Marian Klamer

is Professor of Austronesian and Papuan Linguistics at Leiden University. She has published on a wide range of topics, including morphology, typology, grammaticalization, language contact, and historical reconstruction. Klamer is the author of (sketch) grammars of two Austronesian languages (Kambera, 1998, Alorese, 2011) and three Papuan languages (Teiwa, 2010, Kaera, 2014, and Sentani, to appear), as well as a number of edited volumes. Over the years, Klamer has led a range of research projects studying the grammatical variety and contact-induced change in Austronesian and Papuan languages spoken by small-scale communities in eastern Indonesia.

Francesca R. Moro

is Postdoctoral Research Fellow at the University of Napoli L'Orientale, Italy. In 2016 she completed her PhD at Radboud University Nijmegen on Ambon Malay spoken as heritage language in the Netherlands. From 2015 to 2019 she joined the NWO Vici Research Project *Reconstructing the past through languages of the present: the Lesser Sunda Islands* at Leiden University, where she investigated contact-induced language change in Alorese, an Austronesian language surrounded by Papuan languages. Moro has published articles on contact phenomena between Ambon Malay and Dutch, as well as between Alorese and its Papuan neighbours in the *International Journal of Bilingualism*, *Oceanic Linguistics*, and the *Journal of Language Contact*. Currently she teaches Indonesian language at the BA and MA level, and does research on Filipino as heritage language in Italy.

George Saad

is an Assistant Professor of Indonesian linguistics at Palacký University Olomouc, the Czech Republic. In 2020, he received his PhD from Leiden Uni-

versity for research investigating language contact and change in Abui (Papuan; Indonesia). He has previously conducted research on Kamang (Papuan; Indonesia) at the University of Amsterdam. He has published a sketch grammar of Abui as well as having written a sketch grammar of Shuar (Chicham; Ecuador). His current interests include descriptive linguistics, language contact and change, indigenous multilingualism, language variation, and language revitalization.

Antoinette Schapper

is a linguist specialising in the description and typology of Papuan and Austronesian languages, with a particular focus on the languages of Wallacea. She is currently writing a book on the history of the Papuan languages of Timor, Alor and Pantar.

Yunus Sulistyono

is a lecturer at the Department of Indonesian Education, Universitas Muhammadiyah Surakarta in Indonesia. He teaches courses in phonology, discourse analyses, and comparative linguistics of Indonesian local languages. He obtained his PhD from Leiden University, The Netherlands in 2022. His PhD research concerns the historical reconstruction of Alorese, an Austronesian language spoken in eastern Indonesia, using a combination of historical comparative linguistics and oral history.

Lexical Borrowing in Austronesian and Papuan Languages: Concepts, Methodology and Findings

Marian Klamer and Francesca R. Moro

Introduction

A fundamental idea in linguistics is that similarities between geographically close languages are not accidental, but point to a shared history of their speakers. Either, the speakers descend from a common ancestor, and the similar features were passed down the generations; or they are, or once were, in mutual contact, and adopted features from each other. This volume studies the latter type of contact-induced similarities, focussing on lexical borrowing.

Lexical borrowing involves the transmission of lexical material from one language to another. Lexicon is easily borrowed, and the lexicon of a language can provide important traces of the social and cultural past of its speakers (Ross 2013). For example, loanwords often signal contact in particular socio-semantic domains such as governance, technology, religion or trade at specific moments in time, and the contact may be datable by the spread of loanwords through a group of languages and level of integration into individual languages. As one of the most widespread and extensively documented form of contact-induced language change (Grant 2015), lexical borrowing is probably the most fruitful part of a language to look at in search of traces of a past history of contact.

Island South East Asia and New Guinea are ideal regions in which to study language contact. The region hosts thousands of languages and has a long history of contact through trade and marriage exchanges, or by culturally dominant groups, both colonial and indigenous. Coupled with the sharp lexical and typological contrasts between the Austronesian and non-Austronesian (Papuan or Indo-European) languages spoken in the region, this provides numerous opportunities to study many different types of language contact situations. The present volume studies language contact particularly in the Philippines, Indonesia, Timor-Leste, and New Guinea.

Although linguistic research on language change induced by contact between Austronesian and Papuan languages is increasing, the number of studies is still rather limited, and their scope varies. Most publications on lexical borrowing describe how a single language is influenced by a (regionally

or nationally) dominant language—recent examples from the region include Saad, Klammer & Moro (2019); Klammer & Saad (2020). Studies incorporating a wider set of Austronesian and Papuan languages typically study the borrowing or ‘diffusion’ of grammatical features (Ross 1996; Dunn et al. 2008; Foley 2010), sometimes in order to define so-called ‘linguistic areas’ (Klammer, Reesink & van Staden 2008; Ewing & Klammer 2010; Schapper 2015; Holton & Klammer 2017). The two edited volumes published so far on contact-induced change in the Austronesian world, namely *Language contact and change in the Austronesian world* (Dutton & Tryon 1994) and *Language change in Austronesian languages* (Ross & Arka 2015) focus mainly on Austronesian languages and discuss various types of (contact-induced) change not restricted to the lexical domain. The volume by Andersen (2003), *Language contacts in prehistory: studies in stratigraphy*, includes only one example of an Austronesian language, the language Rotuman (Fiji). Articles specifically centred on borrowing in the lexicon of Austronesian or Papuan languages include Reid (1994) on possible non-Austronesian lexical elements in Philippine Negrito languages, Terrill (2003) on lexical stratigraphy in the central Solomon Islands, Edwards (2018a) on lexical stratigraphy in Timor, Robinson (2015) on Austronesian borrowings in Alor-Pantar languages, and Gasser (2019) on borrowed colour and flora/fauna terminology in North-western New Guinea.

The current volume similarly focusses on borrowing of lexicon, including both Austronesian and Papuan languages, while expanding the geographical focus to include both Island SE Asia and New Guinea. Compared to existing studies it is innovative in three respects. First, most contributions study borrowing of lexicon *across* family borders. For example, Papuan lexicon entering Austronesian languages, Austronesian lexicon entering Papuan languages, lexicon transferring from one Papuan language family into another, or lexicon from an Indo-European language entering an Austronesian language. Second, some chapters (e.g., the chapters by Edwards and Fricke) systematically examine the entire lexicon of a set of Austronesian languages, focussing on the words that can *not* be shown to have an Austronesian origin. Third, most contributions address the question what loanwords can tell us about the social history of the speaker populations. This question is crucial in Island SE Asia and New Guinea where written historical records and archaeological evidence is very much lacking in most regions. The study of loanwords can provide a window to contact events that happened in the past.

This introductory chapter is organized as follows. In section 1, we give an overview of the concept of loanword, how to define it, the different types of loanwords, and the processes leading to lexical borrowings (1.1). We then discuss methods and practical considerations for detecting loanwords (1.2), and

the data types and data sets that can be used in research on loanwords (1.3). In section 2, we review some of the current models of language contact, relating specific contact settings to amounts and types of lexical borrowings. Section 3 introduces the volume by offering an overview of the chapters.

1 Lexical Borrowing: Concepts, Methods and Data Sets

1.1 *Concepts*

A central concept in this volume is the concept of loanword, which can be defined as ‘a word that at some point in the history of a language entered its lexicon as a result of borrowing’ (Haspelmath 2009: 36). The process of borrowing comprises all kinds of transfer or copying of linguistic elements from a source language (SL) into a recipient language (RL), including lexemes, derivational morphology, (morpho-)syntactic and lexical-semantic structures. Most contributions in this volume (i.e., Hoogervorst; Klamer; Edwards; Gerstner-Link; Moro, Sulistyono & Kaiping; Fricke; Schapper & Huber) are concerned with the borrowing of lexemes, two are concerned with the borrowing of derivational morphology (Baklanova & Bellamy; Gallego), and one investigates contact-induced semantic changes in the lexicon (Saad).

Traditionally, languages in contact are viewed to directly influence each other in two ways: ‘borrowing’, affecting the lexicon; and ‘interference’ affecting the grammar (Weinreich 1953). Van Coetsem (1988; 2000) adds a psycholinguistic dimension to these two processes of transfer, which he refers to as ‘borrowing’ and ‘imposition’, introducing the notion of agentivity of the speaker, and the relative dominance of languages in contact in the individual. While the direction of the transfer of linguistic material is always from source language SL to RL, the agent involved in the transfer is either the RL speaker or the SL speaker, depending on which language is their dominant language. A speaker is generally dominant in the language in which she is most proficient or fluent, which is usually, but not necessarily, her first language (van Coetsem 1988: 13). In Van Coetsem’s terms, ‘borrowing’ is then by speakers who show ‘RL agentivity’ and adopt elements from one or more SL into their dominant RL, while ‘imposition’ is the result of speakers who show ‘SL agentivity’ by transferring features of their dominant SL onto the RL. In this volume, examples of both processes are discussed. ‘Borrowing’ with RL agentivity would be involved when a speaker of a Timor-Alor-Pantar (TAP) language uses words originating from an Austronesian language (Klamer), or when a speaker of an Austronesian language uses words from a TAP language (Schapper & Huber; Moro et al.). An example of ‘imposition’ with SL agentivity would be when a speaker of an

Austronesian language uses derivational morphology from another Austronesian language (**Gallego**) or from a non-Austronesian language (**Baklanova & Bellamy**). In terms of contact-induced outcomes, borrowing typically results in transfer of lexicon to the RL, while imposition typically results in phonological or morpho-syntactic changes in the RL (see section 2).¹

The word from the SL that served as a model for the loanword in the RL may be called the source word, which may be morphologically simplex or complex. If it is complex, typically the internal structure of the word is lost when it enters the RL. This is in fact one of the ways in which the direction of borrowing can be established: if we attest similar lexemes across two or more languages, and the word is morphologically analyzable in language A, but not in language B, then A is likely to be the SL (see section 1.2 for further discussion of ways to establish loanwords and direction of borrowing). However, while it is rarely attested, complex loanwords can also be borrowed along with their structural properties. Such loanwords give rise to words in the RL that show combinations of non-native affixes with native stems, and native affixes with non-native stems; besides the regular native-native and non-native-non-native combinations. An example of this is Ibatan, which combines non-native prefixes and stems borrowed from Ilokano with native Ibatan affixes and stems (**Gallego**).

1.2 *Methods*

A loanword has a form and a meaning that is identical or similar to the form and meaning of a lexeme in a SL with which plausible contact exists, or existed. For example, contact is plausible when the languages are spoken in adjacent geographical regions, or are known to be (or have been) involved in trade or marriage exchange. If similarities between lexemes are explainable by their common descent, they are not loanwords. Sound imitations and nursery forms are known to be crosslinguistically formed in similar ways without having a shared history, so similarities between such forms cannot be taken to point to contact either.

In some cases, it is not known whether a word is a loanword or a native form in a particular language or language group; then, the form-meaning pair(s) are referred to neutrally as ‘lexemes’, and the investigation of their history considers ‘shared lexicon’ (**Schapper & Huber**) or ‘lexeme sets’, sets of formally similar words that appear across languages (**Fricke; Moro et al.**). Lexeme sets can be distinguished into two types: cognate sets and similarity sets. Cognate sets trace

1 Van Coetsem’s notion of ‘imposition’ corresponds closely to ‘interference through shift’ in Thomason & Kaufman (1988) (see Winford 2020).

back to a reconstructible proto form in a proto language (represented with an asterisk <*> preceding it, e.g. Proto Malayo-Polynesian *pitu ‘seven’), while similarity sets are not known to be reconstructible to a common proto form. They do however show striking form-meaning similarities that suggest some shared history: either common descent, or contact, or a combination of both. If the assumption is that they may share a common ancestor, the possible/hypothetical proto form is preceded by a hashtag <#> to distinguish it from established proto forms (e.g., #kafo ‘eight’, Schapper & Huber Table 6.3; Lamaholot-Kedang #dahe-k ‘near’, Fricke Table 5.2).

In most studies in this volume, loanwords are diagnosed using the results of earlier historical comparative work. For example, one way to argue that a lexeme (set) has been borrowed into Timor-Alor-Pantar languages is to demonstrate that it has a Proto Austronesian (PAN) or Proto Malayo-Polynesian (PMP) reconstructed form with a similar form and meaning, from which it can be regularly derived. Similarly, to argue that a lexeme attested in an Austronesian language is from a non-Austronesian (Papuan) SL, it is useful to show a similar form that has been reconstructed for a non-Austronesian group of languages.

For the etymology of Austronesian lexemes, the database of Austronesian and its subgroups as listed in Blust & Trussel (2016) is used. In addition, several chapters in this volume make use of recent reconstructions of lower-level subgroups within Malayo-Polynesian that have been proposed in recent years: the Flores-Lembata subgroup, and within it, the Lamaholot subgroup (Fricke 2019); the Central Flores subgroup (Elias 2018); the Timor-Babar subgroup and the Central Timor subgroup (Proto Timor-Babar being a sister to Proto Central Timor and Helong, Edwards 2018b; 2018a); the Rote-Meto cluster (Edwards 2021) and the Alorese cluster (Sulistiyono 2022). For the etymology of lexemes from Timor-Alor-Pantar languages, forms from Proto Alor-Pantar (Holton et al. 2012; Holton & Robinson 2017), or Proto Timor-Alor-Pantar (Schapper, Huber & van Engelenhoven 2017) can be compared. With such detailed etymological information available it is possible to establish which forms in a similarity set share an Austronesian or a TAP ancestor, and which forms do not (Klamer; Moro et al.; Schapper & Huber). It also allows us to identify which lexemes are of ‘unknown origin’ or ‘non-Austronesian’ (Fricke; Edwards); forms that can then be hypothesised to have been acquired through language contact.

When loanwords are attested across two or more languages, the next step is to formulate a hypothesis about the SL, or the direction in which the borrowing took place. The chapters of this volume have applied several practical considerations for this, including the following.

1. If similar forms across language family A are demonstrably historically related (e.g., because they are regularly derived from a known proto form,

show regular sound correspondences), while a similar form is only attested in one language of family B, then the direction of borrowing is from A to B.

- II. If similar forms in a language or language family A are more similar to each other and/or show a larger geographical spread than those attested in language (family) B, then the direction of borrowing is from A to B.
- III. If a word is morphologically analyzable in language A, but not in language B, then A is the SL.
- IV. If a word is integrated into the phonological system of language A but not in that of language B, then A is the SL.
- V. If a word is attested in language A, language B, and a sister of B, language C, and language C cannot have been under influence of language A, then B is the SL.

If a word in a particular sub-branch of a language family has no similar forms in the rest of the family, this may be seen as evidence for its status as a loanword. However, this individual word may in fact be an inherited word whose cognates happened to be lost elsewhere in the family, so such instances are not considered to be strong evidence for a contact event (Haspelmath 2009: 44). However, the more words a language has without cognates in the family, the less likely the scenario that all of these words got lost in all the other branches. A large amount of words of unknown ancestry in a particular language or language group is therefore suggestive of a contact event, even if no SL is currently attestable (Fricke; Edwards).

1.3 *Data Sets*

As pointed out above, in Island South East Asia and New Guinea, where most indigenous communities do not have written traditions, it is often impossible to exactly date when certain linguistic changes and language contact events took place. This is reflected in Part I of the volume where the dating of pre-modern contacts often remains vague, placing it between the time of the expansion of Malayo-Polynesian languages into Island SE Asia 4000 Before Present time (BP) and the arrival of the first western colonial powers about 500 BP. The data used in the chapters of Part I are generally from previously unwritten sources, including primary data collected through recent fieldwork and oral histories.

Only a few languages in the region have old written traditions. The two main ones are Malay and Javanese, whose written traditions can be traced back to respectively the 7th Century CE (1300BP), and the 9th Century CE (1500BP) (Hoogervorst). It is the written tradition of Javanese in particular that provides insights into the history of this language and the languages it has been in contact with. At the same time, Malay was the language of the powerful Malay

empire that had its centre in Malacca on the west coast of Malaysia (located between today’s Kuala Lumpur and Singapore). By the end of the 15th C, Malacca exerted its influence on its immediate region with its literature in Malay, its style of government and culture, thus accelerating the spread of the Malay language. At the height of Malacca’s power, the Malay influence even spread to areas beyond their political control, such as the islands of Ternate and Tidore in the Northern Moluccas. Malay thus became the language of literature and the language of court in many parts of the archipelago, and was thoroughly established by the time the European colonizers arrived in the 16th C. It was subsequently taken up by the Portuguese, Dutch and British colonial powers as a tool of centralisation and modernisation (Collins 1997). Malay as the language of trade has retained its role to this day. Malay was (and is) thus the vehicle by which many loanwords from other language families (Dravidian, Indo-Aryan and Indo-European) entered the local languages of Island SE Asia (**Hoogervorst**).

Sometimes, important regional languages were recorded on paper by the colonial powers. This includes for example Tagalog, the current national language of the Philippines, sources of which go back to the time of the Spanish rule in the late 16th C (**Baklanova & Bellamy**). However, in most of the regions discussed in this volume, linguistic documentation only started about fifty years ago, with the bulk of the work taking place during the last twenty years. So, most chapters use synchronic data sets without information on past stages of the languages.

Apart from the fact that they are mostly synchronic in nature, the data sets as used in the studies of this volume are very different in type and size, an overview is given in Table 1.1. Three contributions (**Klamer; Fricke; Moro et al.**) have made use of the data in the online lexical database LexiRumah (Kaiping, Edwards & Klamer 2019). The reader is referred to Lexirumah for the sources of the data.

TABLE 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set

Chapter	Recipient language(s)	Source language(s)	Data type	Data set size
7	Alorese (Moro et al.)	TAP languages	Mainly synchronic lexical data from LexiRumah	Very large: 13 Alorese dialects, 55 Austronesian language varieties, 42 TAP language varieties × ~600 words = more 66,000 lexemes

TABLE 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set (*cont.*)

Chapter	Recipient language(s)	Source language(s)	Data type	Data set size
8	Kilmeri (Border) (Gerstner-Link)	Nimboran / Sentani	Synchronic lexical data from (sketch) grammars, wordlists, dictionaries	Relatively large: 14 Papuan languages (Kilmeri, Waris, Imonda, Amanab, Taikat, Auyi, Nimboran, Sentani, Skou, Wutung, Dumo, Dusur, Tsaka, Barupu), from each language ~100 items
3	TAP languages (Klamer)	Malayo-Polynesian	Synchronic data from wordlists and reconstructed forms in LexiRumah	Large: 54 TAP language varieties and 55 AN language varieties. For each language, 75 concepts were inspected, i.e. 109 lects × 75 lexemes = 8,175 lexemes
4	Proto Rote-Meto (Edwards)	extinct non-AN	Synchronic lexical data; reconstructions based on these forms	Large: 1,173 Proto Rote-Meto reconstructions; the presence of cognates in other languages in the region has also been tracked
5	Lamaholot (Fricke)	extinct non-AN	Synchronic lexical data from wordlists in LexiRumah and from dictionaries, reconstructed forms	Large: 46 Flores-Lembata language varieties, from which over 400 lexeme sets were extracted
9	Tagalog (Baklanova and Bellamy)	Spanish	(a) Historical data from the 19th–early 20th century lexica (b) Contemporary data of the 20th–early 21st century	Large: Older Spanish-Tagalog dictionaries; 34 sample Tagalog texts, 6 pieces of literary texts; modern Tagalog dictionaries, the Tagalog Leipzig Corpus
11	Abui (Saad)	(Alor) Malay	Synchronic data set with utterances	Large: 6 videoclips × 66 speakers = 396 utterances
2	Malay, Javanese and other AN languages (Hoogervorst)	Indo-Aryan (e.g., Sanskrit) and Dravidian (e.g., Tamil)	Written sources, dictionaries, old texts	Unspecified
6	KAWAIMINA languages (Schapper & Huber)	TAP languages	Synchronic data from (sketch) grammars, dictionaries, fieldnotes; reconstructed forms	Unspecified

TABLE 1.1 Data types and data sets used in the chapters of this volume, organised according to size of data set (*cont.*)

Chapter	Recipient language(s)	Source language(s)	Data type	Data set size
10	Ibatan (Gallego)	Ilokano	Synchronic data set including an Ibatan dictionary, and recordings of naturalistic speech during fieldwork in 2018	Unspecified

Intuitively, we might expect that the size of a data set would influence the results: the more lexemes of a language are investigated, the higher the chance of detecting new loanwords. This would particularly be the case when the lexeme sets under investigation are not restricted to basic word lists or non-cultural ‘core vocabulary’ (which are assumed to be more resistant to borrowing than other vocabulary), but also include highly borrowable cultural concepts, such as is the case in the word lists in LexiRumah.

In this respect, it is interesting to note that **Moro et al.** investigated a huge data set of 66,000 forms from LexiRumah, but found that the percentage of Timor Alor Pantar (TAP) loanwords in Alorese is only slightly higher than the (low) percentages found in earlier studies that were conducted on a basic vocabulary Swadesh list. As Moro et al. remark, this suggests that a loanword analysis on the basis of a Swadesh list can give a representative figure of the proportion of loanwords in a language. On the other hand, however, **Edwards** in his contribution shows that in Austronesian Proto Rote-Meto, the basic vocabulary contains fewer non-Austronesian words (31% of 242 items) than the larger lexicon (55% of 1,148 items) (**Edwards**, Table 4.10). Note however, that one third of the basic vocabulary of Proto Rote-Meto was non-Austronesian, a proportion that goes against the generally accepted (but yet unproven) idea that basic vocabulary is immune to borrowing. In general, languages in our region of study appear to be variable in this regard, and core vocabulary items such as body part terms, kinship terms and certain numerals are often borrowed (**Edwards; Schapper & Huber; Moro et al.; Klammer; Gerstner-Link; Hoogervorst**; see also Foley 2010: 799).

2 Contact Settings and Amount of Lexical Borrowing

Generally speaking, when two or more languages are in contact, this means that groups of speakers interact face-to-face to a certain extent. This interaction, as we will see below, can bring about all kind of changes in the structure and the lexicon of the languages involved, usually the more intense the interaction, the more pervasive the changes will be. Linking contact-induced language changes to specific contact settings allows us to make predictions about what will happen in a given scenario, or hypotheses about what has happened in the past. Here is one example (adapted from Aalberse, Backus & Muysken 2019: 13):

Assume that if a prototypical social setting involving language contact *A* (e.g., contact between North Moluccan Malay and Taba, an indigenous language of Indonesia) has been well studied and produces linguistic properties *p* and *q* (i.e., borrowing of grammatical function words from Malay), then a social setting under study *B* (i.e., contact between the local Malay variety and another indigenous language of Indonesia), resembling *A* in crucial ways, will be likely to also have these properties *p* and *q* (i.e., borrowing of approximately the same grammatical function words from Malay), assuming also roughly the same types of languages involved.

So, we can expect that in other indigenous communities of Indonesia dominated by Malay, the local languages will be influenced approximately in the same way as Taba is. This is exactly what we find, as reported for other Austronesian languages, like West Tarangan, Biak, and Central Lembata, and non-Austronesian Abui (e.g., Nivens 1998; van den Heuvel 2006; Fricke & Saad 2017), all of which have incorporated Malay function words like *kalau* ‘if’.

In order to make predictions, like the one above, we need models of language contact, which explain the processes, as well as the psycholinguistic and sociolinguistic mechanisms that underpin outcomes of language contact, and can be used to infer the contact setting that brought about a specific change (Thomason 2001; Kusters 2003; Trudgill 2011; Muysken 2013; Ross 2013).

For example, Thomason (2001: 70–71) proposes the following borrowing scale to predict which types of lexical borrowings can be expected in contact situations.

Intensity of contact correlates with the amount and types of lexical borrowings: under conditions of casual contact only non-basic vocabulary gets

TABLE 1.2 Lexical borrowing in Thomason's borrowing scale

Intensity of contact	Type of speakers	Borrowed elements
1. Casual	Few bilinguals among borrowing-language speakers, borrowers need not be fluent in the source language.	Only non-basic vocabulary. Only content words: most often nouns, verbs, adjectives, and adverbs.
2. Slightly more intense	More fluent bilinguals among borrowing-language speakers, but they are probably still a minority.	Still non-basic vocabulary. Function words (e.g. conjunctions and adverbial particles like 'then') as well as content words.
3. More intense	A conspicuous number of bilinguals among borrowing-language speakers, attitudes and other social factors favoring borrowing.	Basic and non-basic vocabulary. More function words, including closed-class items as pronouns and low numerals; derivational affixes.
4. Intense	Very extensive bilingualism among borrowing-language speakers, social factors strongly favoring borrowing.	Heavy lexical borrowing in all sections of the lexicon.

BASED ON THOMASON 2001: 70–71

borrowed, but as the intensity of contact increases along with the number of fluent bilinguals in the community, then function words, basic vocabulary, and ultimately derivational morphology and all sections of the lexicon can be borrowed as well. Thomason (2001), thus, uses intensity of contact as the main social predictor. The concept of intensity of contact is hard to define, but can be operationalized as a function of the level of fluency of the borrowers, the proportion of borrowing-language speakers who are fully bilingual in the source language, and the speakers' attitudes. Besides intensity of contact, the other major predictor is linguistic: typological similarity between languages enhances the possibility of borrowing, and loose structures are easy to borrow than tightly integrated structures.

Ross (2013) adds a new dimension to the concept of intensity of contact, namely that of age. In his study on shift-induced changes in Melanesia, Ross links life stages of shifting speakers to prototypical linguistic effects: adult second language learning typically leads to the retention of a good amount

of vocabulary from their heritage language into (the version of) the language to which they are shifting (together with phonological transfer, constructional calquing and simplified (morpho-)syntax); while child bilingualism typically leads to lexical calques (together with syntactic copying and complexification).

Taking a cross-linguistic perspective, Tadmor (2009) compares rates of lexical borrowings in the world languages, surveying 41 languages. Tadmor's four levels can be paired with the four types of intensity of contact of Thomason: "low borrowers" (< 10%, casual), "average borrowers" (10–24%, slightly more intense), "high borrowers" (25–50%, more intense), and "very high borrowers" (> 50%, intense). The percentage of lexical borrowing is inevitably linked to specific contact settings, as exemplified by two prototypical cases: Selice Romani (62.7%) and Mandarin Chinese (1.2%). Some of the sociolinguistic circumstances underlying such different borrowing rates are universal multilingualism, minority language status, permissiveness toward borrowings, and donor languages well known in the case of Selice Romani, while we find almost no bilingualism, majority language status, purist attitude and donor languages poorly known in the case of Mandarin Chinese.

We have seen that specific contact settings can predict the amount of lexical borrowing to be found in a given language. However, it is not only the amount of lexical borrowing that varies depending on the sociolinguistic circumstances, but also the meaning of the loanwords, or their semantic fields. Tadmor, Haspelmath & Taylor (2010) investigated the likelihood of borrowing across a list of 22 semantic fields (taken from Buck 1949) in 41 languages. The six fields most likely to be borrowed (> 30%) are: *Religion and belief*, *Clothing and grooming*, *The house*, *Law*, *Social and political relations*, and *Agriculture and vegetation*. Thus, we can expect that in contact situations that involve casual contact, where few speakers are fluent bilinguals in both languages, the loanwords will come from these semantic fields. One example of casual contact is that of Sanskrit loanwords in Malay and Javanese (and in other languages of the region), as discussed in Hoogervorst, that indicate new items or concepts, such as *āgama* 'sacred traditional doctrine or precepts' (*Religion and belief*), or *doṣa* 'transgression', *pañjara* 'prison', *sākṣī* 'witness' (*Law*).

As hinted above, language contact models can be used in two ways (Aalberse et al. 2019: 13):

- I. They could predict, given a specific language contact setting and a specific language pair, what the linguistic outcome is most likely to be.
- II. They could help understand, given a specific linguistic outcome, what would be the most likely contact setting leading to that outcome has been.

In Island SE Asia and New Guinea, a region that lacks archaeological data and historical written sources, the study of language contact mostly serves

purpose (ii). In facts, virtually all contributions in this volume try to understand, on the basis of the amount and type of lexical borrowings, what was the most likely contact scenario that gave rise to that type of lexical influence. The languages discussed in this volume can be divided according to the intensity of contact, the level of borrowing, the contact processes and the borrowed elements (see Table 1.3 on the next page).

In this region, we find possibly all types of contact setting and related outcomes, from casual contact to intense contact. Four studies report low levels of borrowings in the recipient languages: Kilmeri (**Gerstner-Link**), Alorese (**Moro et al.**), TAP languages (**Klamer**), and Kawaimina languages (**Schapper & Huber**). The limited lexical influence can be accounted for by lack of long-term contact, and pressure to maintain identity (**Gerstner-Link**), by asymmetric bilingualism patterns and numerous first languages (L1s) interfering with each other (**Moro et al.**), by superficial contacts between speakers (**Klamer**), and by lack of data from the non-AN donor languages of Timor, especially in crucial domains such as plants and animals (**Schapper & Huber**). The study of **Hoogervorst** on lexical influence from South Asia languages (e.g., Sanskrit and Tamil) on Malay, Javanese and other languages of the region does not discuss percentages for the individual languages, nor does it specify the type of speakers who were involved. The transmission of South Asian loanwords was primarily the result of language contact with Malay, both for Austronesian and non-Austronesian languages, and therefore we can hypothesize that the type of contact was casual and involved only few bilinguals among borrowing-language speakers.

Two studies report high level of borrowing in Tagalog (**Baklanova & Bellamy**), and Ibatan (**Gallego**). In Tagalog and in Ibatan, two cases of relatively intense contact, we find borrowing of derivational morphology, as expected according to Thomason's scale (see Table 1.2 above); the contact process is imposition transfer by Ilokano-dominant bilinguals for Ibatan, and by Chinese mestizos for Tagalog. We find only two cases of very high levels of borrowings: **Edwards** who discusses loanwords from an extinct non-AN language into Proto Rote-Meto, and **Fricke** who discusses loanwords from an extinct non-AN language into Lamaholot. Both studies discuss lexical borrowing from a language(s) for which we no longer have direct evidence (also known as 'reconstructio ex silentio', see Ross 2013: 11). The difference is that in the case of Proto Rote-Meto, the contact process was adult language shift, as evidenced by the fact that loanwords come from specific semantic domains, and that we also find traces of phonological transfer (see Ross 2013), while in the case of Lamaholot code-switching was the more likely process, as all domains of the lexicon are involved.

TABLE 1.3 Contact settings and lexical borrowing in the contributions of this volume

Recipient language(s)	Source language(s)	Intensity of contact	Level of borrowings	Contact process	Borrowed elements
Malay and Javanese (Hoogervorst)	South Asian	(not discussed)	(not discussed)	Not specified in the paper. Malay and Javanese were the carriers of loanwords into other local RLS.	Semantic domains of loanwords: precious minerals, and metals, geography, law, plants, numerals, religion, mythology, governance, toponyms, and royal titles.
Kilmeri (Border) (Gerstner-Link)	Nimboran / Sentani	Casual	Low (2,3%)	Bilingualism in the family and village contexts due to intermarriage. Language is seen as an emblem of group identity (e.g., for Kilmeri).	Loanwords in the semantic domains of nature, animals, kinship, body parts, and motion. <i>Wanderwörter</i> regarding 'water', 'vegetation' and 'arrow' suggestive of trade (bird of paradise).
Alorese (Moro et al.)	TAP languages	Casual	Low (4.7%)	Asymmetric bilingualism, several Lis interfering with each other.	Loanwords especially in the semantic domains of tools, vegetation, and basic actions.
TAP languages (Klamer)	Malayo-Polynesian	Casual	Low (~8%)	No pervasive bilingualism, nor shift; more likely superficial contact.	Loanwords especially in the semantic domains of technology, societal structures, and subsistence and trade.
Kawaimina languages (Schapper & Huber)	TAP languages	Casual	Low (11 items, percentage not given)	(not discussed)	Loanwords especially in the semantic domains of plants and animals, in particular creepy-crawlies.
Tagalog (Baklanova)	Spanish	More intense	High (20–32%)	(not discussed)	Derivational morphology.
Ibatan (Gallego)	Ilokano	More intense	High (40%)	Imposition transfer by Ilokano-dominant bilinguals.	Derivational morphology.
Lamaholot (Fricke)	extinct non-AN	Intense	Very high (50%)	Code-switching.	Basic and non-basic vocabulary, no specific semantic domain(s).
Proto Rote-Meto (Edwards)	extinct non-AN	Intense	Very high (55%)	Adult-language shift.	Basic and non-basic vocabulary, especially in the semantic domains of tools, and vegetation.
Abui (Saad)	(Alor) Malay	Intense	(not discussed)	Transitional bilingualism: (pre)adolescents and young adults dominant in Malay.	Semantic changes in the lexicon: generalization in three verbal domains.

As for the language Abui, **Saad** does not discuss lexical borrowing, but rather the lexical semantic change of ‘generalization’, whereby some specific words fall into disuse and become replaced by more frequent words. This change is more dramatic in those bilingual speakers who are psycholinguistically dominant in Malay ((pre)adolescents and young adults), thus showing that generalization correlates with intense contact.

Interestingly, in two cases of intense contact out of five, namely Tagalog and Abui, the recipient or donor language is a ‘High’ variety: a colonial language (Spanish) or a lingua franca or a national language (Malay/Indonesian for Abui). Thus, it seems that when only indigenous local languages are involved in the contact, high or very high levels of borrowing are unlikely. This is possibly connected to the observation that adult language shift (leading to high level of borrowing) is rare in small-scale societies (Ross 2013: 28), such as the ones discussed in this volume.

Finally, an interesting pattern emerges looking at the semantic fields of the loanwords. In the cases of casual contact of Alorese (**Moro et al.**), and Kawaimina languages (**Schapper & Huber**), but also in the case of Proto Rote-Meto (**Edwards**) characterized by intense contact, the semantic fields of *Tools/Technology, Agriculture and Vegetation, Animals* and *Social and political relations* (including societal structures) are favored. Interestingly, these three case studies discuss possible non-Austronesian lexical influence on Austronesian languages, thus they indicate that non-AN languages of the region mostly contributed with words related to the environment and technology. The case study of **Klamer** on Austronesian influence on TAP languages presents a complementary view, showing that the Austronesian languages contributed with words related to textile technology, societal structures (‘slave’, ‘king/ruler’), subsistence and trade (‘salt’, ‘seed’, ‘maize’, ‘skin’), and marriage (‘bride price’).

3 Introducing the Volume

The volume consists of two parts covering different periods of time. Part I contains five studies of contact that took place in ancient and pre-modern times, and whose contact settings do not exist anymore, or their dynamics have changed dramatically. This is the time between the expansion of Malayo-Polynesian languages into Island SE Asia, which started some 4000 years BP, and the advent of the first western colonial powers about 500 years BP. The contact events in this period cannot be dated with any precision, but must have taken place before the time when western colonial powers produced their written historical records of parts of the region.

The first chapter in Part I is by **Hoogervorst**, who takes the whole of Island SE Asia as region of investigation. His contribution shows traces of ancient East Asian loanwords in the Austronesian and Papuan languages of Island SE Asia, whose dispersal was either direct, or mediated through Malay and Javanese, with Sanskrit mostly a source for cultural borrowings (prestigious concepts), and Tamil for replacive borrowings (every-day items).

The contribution of **Klamer** analyzes Austronesian loanwords attested in TAP languages and shows that the Austronesian influence in pre-modern times involved animals ('pig', 'deer'), textile technology ('needle', 'to weave', 'to sew'); societal structures ('slave', 'king/ruler'), body parts ('breast', 'navel'), subsistence and trade ('salt', 'seed', 'maize', 'skin'), and marriage ('bride price'). She also argues that, while TAP communities have been in contact with Malayo-Polynesian speaking groups since the stage of proto TAP, thousands of years ago, their mutual contacts generally must have remained superficial, being limited to circumscribed domains and individual people.

The chapters by **Edwards** and **Fricke** present a stratigraphic analysis of the lexicon of Rote-Meto and of Lamaholot, respectively. These two languages have undergone a process of relexification, whereby a good amount of pre-existing words have been replaced with words from an (unattested) language. In such cases, lexical borrowings are the only evidence of the existence of an unattested language or scenario of contact (Grant 2015: 13).

Schapper & Huber investigate the lexical entwinement of the (Austronesian) Kawaimina languages and the (TAP) Maka languages in East Timor, and argue for bidirectionality in lexical borrowing between Papuan-Austronesian languages in the Timor area. They show that Papuan etyma found in the Kawaimina languages have not necessarily been borrowed from the Maka languages. At the same time, Makasae, the largest Maka language, is the immediate source for Austronesian etyma in the Kawaimina languages; and some lexicon that is shared between Kawaimina and Maka languages has no clear origin outside of those groups or appears to have been borrowed in parallel into both group's languages from one or more unknown languages.

Part II of this volume covers studies of contact in modern and contemporary times (from 500BP to the present), in contact settings that are to some extent still present today.

The contribution of **Moro, Sulistyono & Kaiping** on Alorese, an Austronesian language surrounded by Papuan TAP languages, display a clear example of a language in which, despite a long history of contact, lexical borrowing is not very significant in quantitative terms, but it can be revealing to understand pattern of interactions and dialect dispersal.

Gerstner-Link investigates lexical borrowing in a complex exchange scenario involving the Papuan families of Border, Nimboran, Sentani, and Skou



FIGURE 1.1 Locations of languages or language areas discussed in the chapters of this volume, by their chapter number

Legend to map

2. *Hoogervorst: Lexical influence from South Asia (map indicates locations of Malay and Old Javanese)*
3. *Klamer: Traces of pre-modern contact between Timor-Alor-Pantar and Austronesian speakers*
4. *Edwards: Phonological innovation and lexical retention in the history of Rote-Meto*
5. *Fricke: The mixed lexicon of Lamaholot (Austronesian): A language with a large lexical component of unknown origin*
6. *Schapper & Huber: Entwined histories: the lexicons of Kawaimina and Maka languages*
7. *Moro, Sulistyono & Kaiping: Detecting Papuan loanwords in Alorese: Combining quantitative and qualitative methods*
8. *Gerstner-Link: Multilateral lexical transfer among four Papuan language families: Border, Nimboran, Sentani, and Sko*
9. *Baklanova & Bellamy: Spanish suffixes in Tagalog nominal derivation: The case of common nouns*
10. *Gallego: The structural consequences of lexical transfer in Ibatan*
11. *Saad: The effects of language contact on lexical semantics: The case of Abui*

located in the island of New Guinea. On the basis of the high number of mutual loans between Border and Nimboran languages, new hypotheses are formulated about the migration routes of the Border people, as well as about the genetic unity of the Border and Nimboran families.

The paper by **Baklanova & Bellamy**, as well as the one by Gallego, both show that loanwords can lead to the transmission and integration of derivational morphemes in the recipient languages. For instance, as shown by Baklanova and Bellamy, Tagalog has absorbed many Spanish words which acted as a conduit for the borrowing of agentive and adjectival suffixes. Similarly, **Gallego** analyses the history and development of the verbal prefix *mag-* in Ibatan, which has been copied from Ilokano as part of complex loanwords.

Saad's is the only contribution that focuses on the outcome of contact-induced change in the semantics of language, by demonstrating that the meaning of certain verbs in Abui (a TAP language) has changed due to the influence of semantically similar verbs in the dominant language Malay (Austronesian).

The linguistic region covered by each of the chapters is indicated on the map in Figure 1.1 on page 17. More detailed maps of these respective areas are provided in the individual chapters.

Bibliography

- Aalberse, Suzanne, Ad Backus & Pieter Muysken. 2019. *Heritage Languages: A language contact approach*. Amsterdam: John Benjamins.
- Andersen, Henning (ed.). 2003. *Language Contacts in Prehistory: Studies in Stratigraphy*. Amsterdam: Benjamins.
- Blust, Robert & Stephen Trussel (eds.). 2016. The Austronesian Comparative Dictionary (web edition). Retrieved from <http://www.trussel2.com/ACD>
- Buck, Carl Darling. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages*. Chicago: University of Chicago Press.
- Collins, James T. 1997. *Malay, world language of the ages. A sketch of its history*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink & Angela Terrill. 2008. Structural Phylogeny in Historical Linguistics: Methodological Explorations Applied in Island Melanesia. *Language* 84(4). 710–759.
- Dutton, Tom & Darrell T. Tryon (eds.). 1994. *Language contact and change in the Austronesian world*. Berlin / New York: Mouton de Gruyter.
- Edwards, Owen. 2018a. Parallel Histories in Rote-Meto. *Oceanic Linguistics* 57(2). 359–409.
- Edwards, Owen. 2018b. Top-down historical phonology of Rote-Meto. *Journal of the*

- Southeast Asian Linguistics Society (JSEALS)* 11(1). 63–90. DOI: <http://hdl.handle.net/10524/52421>
- Edwards, Owen. 2021. *Rote-Meto comparative dictionary*. Canberra: ANU Press. Retrieved from <http://doi.org/10.22459/RMCD.2021>
- Elias, Alexander. 2018. Lio and the Central Flores languages. Research Master thesis Leiden University. <https://hdl.handle.net/1887/69452>
- Ewing, Michael & Marian Klamer (eds.). 2010. *East Nusantara: Typological and areal analyses*. Canberra: Pacific Linguistics.
- Foley, William. 2010. Language Contact in the New Guinea Region. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 795–813. West Sussex, UK: Wiley-Blackwell.
- Fricke, Hanna. 2019. *Traces of language contact: the Flores-Lembata languages in eastern Indonesia*. Amsterdam: LOT Publications.
- Fricke, Hanna & George Saad. 2017. 'If' it exists in one language, why not use it in another? The insertion of kalau 'if' in Abui and Central Lembata, two languages of Eastern Indonesia. Presented at the Ninth Austronesian and Papuan Languages and Linguistics conference (APLL 9), Paris.
- Gasser, Emily. 2019. Borrowed Color and Flora/Fauna Terminology in Northwest New Guinea. *Journal of Language Contact* 12. 609–659.
- Grant, Anthony. 2015. Lexical borrowing. In *The Oxford Handbook of the Word*, 1–23. Oxford: Oxford University Press. Retrieved from DOI 10.1093/oxfordhb/9780199641604.013.029
- Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In Uri Tadmor & Martin Haspelmath (eds.), *Loanwords in the World's Languages: A Comparative Handbook*, 35–54. Walter de Gruyter.
- Holton, Gary & Marian Klamer. 2017. The Papuan languages of East Nusantara and the Bird's Head. In Bill Palmer (ed.), *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*, 569–640. Berlin: Mouton de Gruyter.
- Holton, Gary, Marian Klamer, František Kratochvíl, Laura C. Robinson & Antoinette Schapper. 2012. The historical relations of the Papuan languages of Alor and Pantar. *Oceanic Linguistics* 51(1). 86–122.
- Holton, Gary & Laura C. Robinson. 2017. The linguistic position of the Timor-Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology (2nd ed.)*, Second edition. 155–198. Berlin: Language Science Press.
- Kaiping, Gereon, Owen Edwards & Marian Klamer. 2019. LexiRumah 3.0.0. Leiden University Centre for Linguistics. Retrieved from <https://lexirumah.model-ling.eu/lexirumah/>
- Klamer, Marian, Ger Reesink & Miriam van Staden. 2008. East Nusantara as a linguistic area. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 95–149. Amsterdam: Benjamins.

- Klamer, Marian & George Saad. 2020. Reduplication in Abui: A case of pattern extension. *Morphology* 30(4). 311–346. DOI: <https://doi.org/10.1007/s11525-020-09369-z>.
- Kusters, Wouter. 2003. *Linguistic complexity*. Utrecht: LOT Publications.
- Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(04). 709–730. DOI: <https://doi.org/10.1017/S1366728912000727>
- Nivens, Richard. 1998. *Borrowing vs. code-switching: Malay insertions in the conversations of West Tarangan speakers of the Aru islands of Maluku, eastern Indonesia*. PhD thesis University of Hawaii.
- Reid, Lawrence A. 1994. Possible Non-Austronesian Lexical Elements in Philippine Negrito Languages. *Oceanic Linguistics* 33(1). 37–72.
- Robinson, Laura C. 2015. The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara. In Malcolm Ross & I Wayan Arka (eds.), *Language change in Austronesian languages*, 19–33. Canberra: Asia-Pacific Linguistics.
- Ross, Malcolm. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In Mark Durie & Malcolm Ross (eds.), *The Comparative Method Revisited: Irregularity and Regularity in Language Change*, 180–217. Oxford: Oxford University Press.
- Ross, Malcolm. 2013. Diagnosing Contact Processes from their Outcomes: The Importance of Life Stages. *Journal of Language Contact* 6(1). 5–47. DOI: <https://doi.org/10.1163/19552629-006001002>
- Ross, Malcolm & I Wayan Arka (eds.). 2015. *Language change in Austronesian languages: papers from 12-ICAL, Volume 3*. Canberra: Asia-Pacific Linguistics.
- Saad, George, Marian Klamer & Francesca R. Moro. 2019. Identifying agents of change: Simplification of possessive marking in Abui-Malay bilinguals. 4(1), 57. *Glossa: A Journal of General Linguistics* 4(1). 57.
- Schapper, Antoinette. 2015. Wallacea, a linguistic area. *Archipel. Études Interdisciplinaires Sur Le Monde Insulindien* (90). 99–151.
- Schapper, Antoinette, Juliette Huber & Aone van Engelenhoven. 2017. The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology* (2nd ed.), 99–154. Berlin: Language Science Press.
- Sulistyono, Yunus. 2022. *A history of Alorese*. PhD thesis Leiden University. Utrecht: LOT Publications.
- Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 55–75. Berlin / New York: De Gruyter Mouton.
- Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246. DOI: <https://doi.org/10.1075/dia.27.2.04tad>

- Terrill, Angela. 2003. Linguistic stratigraphy in the central Solomon Islands: Lexical evidence of early Papuan / Austronesian interaction. *The Journal of the Polynesian Society* 112(4). 369–401.
- Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. & Terence S. Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- van Coetsem, Frans. 1988. *Loan Phonology and the Two Transfer Types in Language Contact*. Dordrecht: Foris.
- van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Monographien zur Sprachwissenschaft Vol. 19. Heidelberg: Winter.
- van den Heuvel, Wilco. 2006. *Biak: description of an Austronesian language of Papua*. PhD dissertation Vrije Universiteit Amsterdam. Utrecht: LOT Publications.
- Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York. Reprinted 1986, The Hague: Mouton.
- Winford, Donald. 2020. Theories of Language Contact. In Anthony Grant (ed.), *The Oxford Handbook of Language Contact*, 51–75. Oxford: Oxford University Press.

PART 1

Ancient and Pre-Modern Contact



Lexical Influence from South Asia

Tom G. Hoogervorst

Introduction

South and Southeast Asia have been in contact for millennia. It is therefore no surprise to find traces of lexical borrowing across its languages and language families. In South Asia, the most widespread and expansive language families are Indo-European (specifically Indo-Aryan) and Dravidian (specifically South Dravidian). The former includes classical languages such as Sanskrit (Sk.) and Pali (Pa.), next to present-day mother tongues such as Hindustani (Hi.), Bengali (Be.), Gujarati, Sinhala, and Odia. Sanskrit represents the Old Indo-Aryan (OIA) stage of historical development, whereas Pali and several extinct vernaculars known collectively as “Prakrit” are classified as Middle Indo-Aryan (MIA), and the modern languages as New Indo-Aryan (NIA). The South Dravidian branch includes Tamil (Ta.), Malayalam (Ma.), Kannada, and Tulu. Tamil and Malayalam have been most prominent in language contact with Southeast Asia. While they are now considered separate languages, Tamil and Malayalam formed an undivided dialect continuum during the earliest stage of language contact with Southeast Asia. I will nevertheless treat them as separate entities in this chapter, as a number of phonological differences allow us to determine whether certain words were borrowed from the eastern or western part of this historical continuum.

In Maritime Southeast Asia, Javanese and especially Malay have historically been crucial for the transmission of loanwords from external sources (Sanskrit, Tamil, Arabic, Portuguese, Dutch, English, etc.) to the region's smaller languages. Javanese is furthermore important on account of its extensive record of inscriptions and other texts, starting from the ninth century CE, which provide valuable insights into language development.¹ What are commonly referred to as “Papuan” languages consist of a number of separate families spoken in the

1 Old Javanese was written in an Indic syllabary and is transliterated according to the ISO 15919 standard by an increasing number of scholars, including elsewhere by the present author. For comparative purposes, I have chosen in this chapter to homogenize the transcription of Old Javanese with that of the other Austronesian languages. Concretely, this means I have not indicated orthographic details that are not based on (reconstructed) phonological realities.

eastern parts of Maritime Southeast Asia and the western parts of Oceania. There is no irrefutable evidence for direct contact between any “Papuan” language and any “Indian” language. With Malay as the chief vector of secondary borrowing, loanwords from Sanskrit and other South Asian languages are chiefly attested in Papuan languages families of the Wallacea region, such as the Timor-Alor-Pantar languages and North Halmahera languages. Further to the east, the presence of South Asian vocabulary is either minimal or very recent, and has always passed through Indonesian.²

This chapter investigates lexical traces from Indo-Aryan and South Dravidian languages in Austronesian and, to a lesser extent, Papuan languages. It does not attempt to be complete.³ Examples have been selected on account of their ability to illustrate the main tendencies underlying early lexical borrowing from South to Southeast Asia.⁴ As far as Austronesian languages are concerned, the geographical distribution of these loanwords is limited to Maritime Southeast Asia (including East Timor and the Philippines), Madagascar, small pockets of Mainland Southeast Asia (in particular the Chamic and Mokenic languages), and—to a minimal degree—Taiwan. There is no evidence of early language contact between South Asia and the Pacific. The role of Austroasiatic languages, which are spoken in South Asia as well as Southeast Asia, is too extensive a topic to be discussed here. I will also not look at Arabic and Persian words that entered Southeast Asia through the springboard of South Asia, nor at “Indian” loanwords recently introduced through European languages.

One of the most underestimated tasks of a historical linguist is to reconstruct regular sound correspondences, both in inherited and borrowed vocabulary. Borrowings between languages with vastly different phonological inventories are often unrecognizable as such. Consider, for example, the Hawaiian words *kalikimaka*, *kanauika*, and *manakuke*, which regularly go back to English ‘Christmas’, ‘sandwich’, and ‘mongoose’. Conversely, words that look similar may prove to be unrelated after the historical phonology of both languages is taken

2 Klammer (this volume) finds no South Asian loanwords in Timor-Alor-Pantar languages that display signs of early acquisition.

3 Extensive overviews of Sanskrit and other South Asian loanwords in the languages of Southeast Asia include Gonda (1973) and Jones (2007). Middle Indo Aryan influence is investigated in De Casparis (1986) and Hoogervorst (2017a, 2017b), whereas South Dravidian influence is investigated in Van Ronkel (1902, 1903) and Hoogervorst (2015).

4 Due to the higher sociolinguistic status of Sanskrit, lexical borrowing has predominantly taken place in the eastward direction. However, see Hoogervorst (2013:106–116) on Malayo-Polynesian loanwords in South Asian languages. There has also been a long tradition among Indologists of detecting purported “Austic” influence in Indo-Aryan languages, although this would-be language family is no longer supported by academic research.

TABLE 2.1 Rejected South Asian borrowings in Malay

Malay	South Asian faux-etymon	Genuine etymology
barat ‘west’	bhārata (Sk.) ‘India’	PMP *habaRat ‘southwest monsoon’
bəli ‘to purchase’	vil- (Ta.) ‘to sell’	PAN *beli ‘to buy’
dara ‘girl’	dāra (Sk.) ‘wife’	PMP *daRa ‘maiden, virgin, unmarried girl’
dua ‘two’	dva (Sk.)	PAN *duSa
hari ‘day’	hari (Sk.) ‘the sun’	PAN *waRi ‘day; sun; dry in the sun’
kəlam ‘dark, obscure’	kālam (Ta.) ‘blackness, darkness’	PAN *kelem ‘night, darkness’
mabuk ‘drunk’	mappu (Ta.) ‘beclouded state of the intellect, as by intoxication’	PAN *ma-buSuk ‘drunk, intoxicated’
patah ‘broken’	phaṭā (NIA) ‘torn, split, broken’	PMP *pataq ‘break, broken, cut through’
saruṅ ‘sarong (hip-wrapper)’	sāraṅga (Sk.) ‘of a variegated colour’	PAN *duṅ ‘shelter’ ^a

a “Attested mainly as doubled or with a petrified prefix” (Wolff 2010:825).

into account.⁵ Table 2.1 lists some widespread faulty etymologies—displaying different levels of credibility—in Malay.

This chapter examines South Asian lexical influence along three lines of inquiry: the integration of loanwords, the timeframe of borrowing, and the trajectories of borrowing. Lexical borrowing from South to Southeast Asia is complicated by vastly different phonological systems, especially in scenarios of secondary and tertiary transmission. Most lexical borrowing furthermore features semantic shift. The timeframe of acquisition is difficult to determine precisely. Textual attestations only provide a “not-after date” of transmission, whereas historical phonology allows for relative dating. Loanwords that exhibit the same phonological innovations as inherited vocabulary, for example, tend to be relatively early introductions. Loanwords for which high-level Austronesian protoforms can be reconstructed tend to be relatively ancient as well. The

5 Common reasons to reject superficially attractive borrowing hypotheses include fortuity, transmission in the opposite direction, and similarities of a universal nature, such as onomatopoeia and kinship terms.

geographical distribution of South Asian loans is another tool to gauge their antiquity. Loanwords only found in the western parts of Maritime Southeast Asia, which had more intensive contact with South Asia, were arguably less prominent than loanwords that spread further eastwards.

1 Integration

As most South Asian loanwords spread across Maritime Southeast Asia through Malay and—to a lesser extent—Javanese, the phonological systems of both languages merit some further comment. Unlike some other Austronesian languages, modern Malay and Javanese normally lack long vowels and gemination. The three-way distinction of sibilants found in Sanskrit (<ś> /s/; <ṣ> /ʃ/; <ṣ̥> /ʃ̥/) is alien to Austronesian languages. Javanese has contrastive dental and retroflex stops—/t̪/ and /d̪/ versus /t/ and /d/—while Malay only has /t̪/ and /d/. As a result, direct South Asian borrowings into Javanese often retain their retroflex /t̪/ (e.g. *cəməʃi* ‘whip’ < MIA **cammat̪hi*, *kaʃil* ‘bedstead’ < Ta. *kaʃtil*, *paʃi* ‘box’ < NIA **peʃi*), whereas borrowings acquired through Malay tend to display their dental counterparts (e.g. *kati* ‘a weight unit’⁶ < Ta. *kaʃti*, *roti* ‘bread’ < NIA **roʃi*, *topi* ‘hat’ < NIA **ʃopi*). Malay historically substituted /w/ by /b/ and /y/ by /j/, except in Arabic loanwords (Hoogervorst 2017b: 295). It furthermore exhibits a tendency to voice the historically voiceless /k/ and /c/ to respectively /g/ and /j/ (Hoogervorst 2015:84–86, 2017b: 296–297). The first syllable of loanwords originally consisting of three or four syllables is often clipped in Malayo-Polynesian languages, e.g. Malay *biasa* ‘usual’ and *puasa* ‘to fast’ respectively from Sanskrit *abhyāsa* ‘repetition; habit’ and *upavāsa*. Modern Malay and Javanese lack aspirated consonants, yet secondary borrowings in Tagalog reveal that aspirate consonants were historically retained by at least some speakers (Adelaar 1994:63). In Malagasy, the aspirated velar stops /kh/ and /gh/ both became /k/, whereas their non-aspirated counterparts /k/ and /g/ became /h/ (Adelaar 1994:64). In Malay, such Sanskrit loans as *bahagia* ‘fortunate’ (< *bhāgya*), *bahasa* ‘language’ (< *bhāṣā*), and *pahala* ‘reward’ (< *phala*) also reflect historical aspiration. In Toba Batak, the historical presence of aspirated consonants is revealed by an epenthetic /a/ (historically preceding a /h/) in words like *baima* ‘a name’ (< Sk. *bhīma*), *bauta* ‘a kind of spirit’ (< Sk. *bhūta*), and *daupa* ‘incense’ (< Sk. *dhūpa*) (van der Tuuk 1971:69).

6 In Old Javanese inscriptions, we find <*kāti*> (Kurungan, 885 CE), <*kati*> (Salingsingan, 880/905 CE), or the abbreviation <*kā*>, whereas later sources mostly feature *kati* or *kāti* (Clavé & Griffiths 2022:228, n.76).

Equal attention should be given to phonological innovations that took place in South Asia. Sanskrit was no longer a spoken language during the first centuries CE, when it exerted lexical influence on parts of Southeast Asia. The vernacular languages of North India around that period presumably constituted an intermediate stage between Middle- and New Indo-Aryan. MIA phonology displays lenition of intervocalic consonants and various assimilation processes of consonant clusters. Loanwords in Austronesian languages that also display these features, hence must have been acquired from MIA rather than Sanskrit. Most (but not all) NIA languages exhibit a further development: the elision of unstressed word-final vowels. Such forms are already attested in Old Javanese literature (Hoogervorst 2017a: 423–431). Other Indo-Aryan loanwords in Austronesian languages retain the unstressed word-final /a/, suggesting an even earlier transmission. Table 2.2 on the next page gives some examples, in which the likeliest stage of transmission is marked grey. Etyma marked with an asterisk (*) are my own reconstructions.

Sanskrit and other Indo-Aryan languages have lexical gender, in which female forms, derived adjectives, and other derivations are marked with a word-final /ī/. One noteworthy phenomenon in a number of West-Malayo-Polynesian languages is a preference for the i-forms of Sanskrit loanwords, even when these are rare or unattested in South Asia (Hoogervorst 2017b: 302–313).⁷ A number of common examples are given in Table 2.3.

The same process was active for MIA loanwords, whose reconstructions are chiefly based on historical phonology rather than textual attestation. Table 2.4 lists some of my postulations. This observation has some far-reaching implications for Austronesian historical linguistics. If a borrowed form **jadi* in the meaning of ‘to be born; to become; to come about’ has indeed made its way into Southeast Asia through a MIA source, and we are not dealing with a case of chance resemblance, this transmission must have taken place at a remarkably early stage. Tentative reflexes such as Javanese *dadi*, Malagasy *zary* and Makassar *jari* all display the expected sound changes of inherited vocabulary (but Tagalog *yari* does not). This would imply that language contact took place when the innovation *j > d in Javanese was still ongoing. Reflexes of a hypothetical **kosali* ‘village hall’ display an equally vast distribution, from Sumatra and the Philippines to Maluku (Lafeber 1922:135–136). The protoform **suligiq* ‘kind of lance’, with attestations in the Philippines and western Indonesia and earlier reconstructed for the somewhat controversial entity of “proto

7 Tetun displays a similar preference for feminine forms of Portuguese loanwords (Hajek & Williams-van Klinken 2019).

TABLE 2.2 MIA and NIA loans in Malay and Javanese

OIA	MIA	NIA	Malay	Javanese
caurikā ‘theft’	cōriā	corī	curi	–
cukra ‘vinegar’	*cukka	cūk	cuka	coka?
dāḍima ‘pomegranate’	dālīma	dāḍim	dəlīma	dlima
gōpāla ‘cowherd’	gōvāla	goāl	gəmbala	–
guḍa ‘sugar’	gula	guḥ	gula	gula
jāgrat ‘to be awake’	jagga	jāg	jaga	jaga
karpāsa ‘cotton’	kappāsa	kapās	kapas	kapas
kuñcīkā ‘a key’	kuñciyā	kuñci	kunci	kunci
kustumbarī ‘coriander’	kutthumbharī	*kuthumbār	kətumbar	kətumbar
*mukhadvāra ‘a river mouth’	*muhavāra	–	muara	muwara
*pragaḍḍa ‘enclosure; fence’	*pagaḍḍa	pagār	pagar	pagər
rājñī ‘queen’	raṇṇī	rāṇī	rani	–
sakala ‘entire, all’	sagala	–	səgala	–
śmaśāna ‘cemetery’	masāṇa	masān	mesan ^a	maesan
śṅkhala ‘chain’	saṅkala	sāṅkal	səṅkəla	–
*sukaṁsa ‘pinchbeck’	*suhamsa	suāsā	suasa	suwasa
taḍāga ‘a pond’	talāga	talāu	təlaḡa	tlaga
tāmraka ‘copper’	tambaga	–	təmbaga	təmbaga
uṣṭra ‘a camel’	uṭṭa	ūṇṭ	unta	unta
vajra ‘steel’	vajja	bāj	baja	waja

a The first vowel in Malay and Javanese is irregular and presumably reflects confusion with the word *nisan* or *nesan* ‘gravestone’ (< Persian *nišān* ‘sign; mark’).

TABLE 2.3 The preference for i-forms in Sanskrit loanwords

Sanskrit	Malay	Javanese	Toba Batak	Tagalog
artha ‘meaning’	arti	ər̥ti	arti	–
bhāga ‘part’	bahagi, bagi	bage	bagi	bahagi
bīja ‘seed’	biji ⁸	wiji	–	–
kacchapa ‘a lute’	kəcapi	kəcapi	hasapi	kudya’pī?

8 Also compare *vijaiḥ* in Old Cham (cf. Lepoutre 2013:234–235; Griffiths & Lepoutre 2016:216, 223–224, 264). The diphthong /ai/ corresponds to /i/ in Malay but the word-final /h/ remains unexplained.

TABLE 2.3 The preference for i-forms in Sanskrit loanwords (*cont.*)

Sanskrit	Malay	Javanese	Toba Batak	Tagalog
krakaca ‘a saw’	gərgaji	graji	garagaji	lagari?
kuṇḍa ‘a vessel’	kəndi	kəṇḍi	hondi	–
pārāpata ‘pigeon’	mərpati	–	darapati	kalapati
roga ‘infirmity; disease’	rugi ‘to suffer financial loss’	rugi	rugi	lugi

TABLE 2.4 Possible i-forms of MIA loanwords

Sanskrit	MIA	Postulated i- protoform	Old Javanese	Malay
dyūta ‘to gamble’	*jūda	*judi	judi, juḍi ^a	judi
jāta ‘born; to come into existence’ (or: jāti ‘birth’) ^b	*jāda (or *jādi)	*jadi	dadi ‘coming into existence; being done’	jadi
kauśalya ‘a kind of pavilion’	*kosalla	*kosali	gusali, gosali ‘smithy’ ^c	–
saraka ‘a drinking vessel’	*saraga	*saragi	saragi ‘a copper kettle or pot’	–
śūlikā ‘a sharp instrument’	*sūligā	*suligi	suligi ‘a kind of spear, javelin’	səligi

- a While absent in Zoetmulder (1982), <juḍi> appears to be the more common spelling (Arlo Griffiths, pers. comm. 2020). Also compare modern Javanese *juḍi* ‘gambling’.
- b The possible connection between Sanskrit *jāti* and Old Malay *jādi* has been pointed out independently by Clavé & Griffiths (2022:224, n.46).
- c Presumably with a broader meaning historically, as reflexes of *gosali* denote a sort of social space in other Austronesian languages. The etymologically related form *gohāli*—found in the Prakrit of North Bengal around the turn of the sixth century—has been interpreted as ‘hamlet’ (Griffiths 2018:40–42).

Western-Malayo-Polynesian”, represents a similar instance of early borrowing from South Asia (Hoogervorst 2016:567–568).

Some loanwords already exhibited i-forms in South Asia. In Hindustani and other NIA languages, ī-suffixation became a productive process to form diminutives and derive abstract nouns (Hoogervorst 2017b: 313–316). Table 2.5 lists a number of common examples.

In addition to direct contact with Indo-Aryan languages, a number of loanwords were evidently transmitted through Tamil or a closely related South

TABLE 2.5 NIA loanwords displaying the suffix -ī

OIA	NIA	Malay	Javanese
bhēdra 'ram'	bheṛī 'sheep'	biri-biri ^a	–
pēṭṭa 'lower belly'	peṭī 'box'	pəti	pəṭi
rōṭṭa 'bread'	roṭī	roti	roti
sthāna 'place'	thānī 'a permanent cultivator'	tani	tani
ṭōppa 'hat'	ṭopī	topi	topi

a I am uncertain how this word relates to Old Javanese *wīwi* 'goat' (attested from the ninth century CE) and Proto Rote-Meto **bibi*.

TABLE 2.6 Indo-Aryan loanwords introduced by speakers of Tamil

Indo-Aryan etymon	Tamil pronunciation	Malay
ghoṭa (Sk.), ghoḍa (MIA) 'horse'	*ko:ḍa	kuda
*joṛo (NIA) 'couple'	*jo:ḍu <cōṭu>	jodoh
loha (Sk., MIA) 'metal'	*lo:ḥam, lo:gəm <lōkam>	logam
parikhā (Sk.), parihā (MIA) 'moat, ditch'	*pəriḡe <parikai>	pəriḡi
rāga (Sk.) 'melody'	*ra:gəm <rākam>	ragam

Dravidian language. Sanskrit loans ending in a short /a/ occasionally obtain the Tamil ending /am/, while those ending in a long /ā/ obtain /ai/. In addition, postnasal or intervocalic stops in Tamil tend to be voiced, whereas word-initial stops tend to be devoiced.⁹ Table 2.6 lists some common Indo-Aryan loanwords in Malay that were presumably introduced by speakers of Tamil.

In some instances, the precise trajectories of borrowing are uncertain. For example, MIA **cammatṭhi* 'whip' and its Tamil equivalent *cammatṭi* would yield the exact same form in Austronesian languages, as would NIA *peṭī* 'box' and Ta. *peṭṭi*. Old Javanese *calana* and Malay *calana* 'trousers' resemble Hindustani *colnā* 'short breeches', yet both may ultimately reflect a South Dravidian form.¹⁰

9 Phonemically there is no opposition between voiced and unvoiced consonants in Tamil, nor in the script.

10 While Tamil and Malayalam have *callaṭam*, Kannada and Tulu exhibit *callaṇa* (Burrow & Emeneau 1984:209).

Old Javanese *joli* and Malay *juli* ‘palanquin’ resemble NIA *ḍolī* or Tamil *ṭōli*, yet the word-initial consonant remains unexplained. The Malay word *bəlanja* ‘expenditure’ is another etymological puzzle. It has been identified as an isolated instance of Sinhala influence, going back to *valaṅḍa-navā* ‘to consume (of important people)’ through the locally created Pali form *valaṅja* (Gonda 1973:80–81). In other instances, the intervocalic /c/ was voiced in Austronesian languages under unclear circumstances, e.g. Old Javanese and Malay *ajar* ‘teaching’ and *ujar* ‘speech’, ultimately from Sk. *ācārya* ‘teacher’ and *uccāra* ‘pronunciation’.

The semantic integration of South Asian vocabulary languages forms an equally important point of attention. For literary languages, we can spot changes in meaning over time. In Malay, for example, the word *desa* (< Sk. *deśa* ‘province; country’) historically referred to any land and later to a rural settlement, while *sastra* (< Sk. *śāstra* ‘teaching; book or treatise’) initially denoted sacred books and astrological tables and later literature in general. The Old Javanese literature is of even greater value in the semantic domain, as it tends to reveal intermediate stages between original etyma and their contemporary derivations. As shown in Table 2.7, Sanskrit loanwords in Old Javanese can often be regarded as a missing link. Note that the Old Javanese examples on the next page are represented in their (reconstructed) phonological rather than orthographic forms.¹¹

With Malay being the chief vector of transmission, we often see multiple semantic shifts; one upon acquisition into Malay and another into the second recipient language. Table 2.8 lists multiple semantic shifts seen in loanwords adopted into Malay and subsequently into Yakan, a language of the southern Philippines.

In some cases, loanwords are difficult to recognize as such due to their phonological integration in the recipient language. The examples in Table 2.9 are from Leti and Rote, two languages spoken, respectively, on the islands east and west of Timor. The relative time depth of borrowing can occasionally be deduced from phonological evidence. Rote *kapa* ‘ship’, for example, is more recently acquired than *aba* ‘cotton’, as the latter exhibits the innovation *k > Ø/#_ also attested in inherited vocabulary. Also note that Leti exhibits a specific type of metathesis yielding vowel-final stems.

11 I provisionally regard aspirated stops in Old Javanese as distinct phonemes on account of the realization of possible Old Javanese loans in Tagalog and Malagasy (see Table 2.21 and 2.22).

TABLE 2.7 Diachronic shifts in meaning

Sanskrit	Old Javanese	Malay
āgama 'approaching; acquisition of knowledge; a traditional doctrine'	agama 'sacred traditional doctrine or precepts'	agama 'religion'
bhaṅga 'breaking; disturbance; rejection'	bhaṅga 'breaking or destroying the laws of <i>dharmā</i> '	banṅa 'proud'
bheda 'breaking; disuniting'	bheda 'separation; disuniting; different'	beda 'different'
cāmara 'a fly-whisk; a plume on the heads of horses'	camara 'a fly-whisk; plume, tuft (on shields)'	cāmara 'ornamental tuft'
carita 'gone; moving; deeds'	carita 'events; story'	cārita 'story'
kapāla 'skull'	kapala 'skull, upper part of the head'	kāpala 'head'
padāti 'going or being on foot; a pedestrian'	padati 'pedestrian; cart'	pādati 'cart'
paḥṣa 'wing; position; a point or matter under discussion'	paḥsa 'fixed intention; firmly decided to'	paḥsa 'compulsion; favourable opportunity'
parihāra 'avoiding; seizing; concealment'	parihara 'to refute; to restrain'	pālihara 'to domesticate (animals); to look after'
saṃyatta 'come into conflict; being on one's guard'	sanjata 'weapon; armed forces'	sānjata 'weapon'
vaca 'speaking; talking'	waca 'to read, sing (a text)'	baca 'to read'
vaṃśa 'a cane; the line of a pedigree'	waṅsa 'lineage, dynasty, posterity'	baṅsa 'race; descent'
viṣa 'a servant; anything active; poison'	bisa 'venomous; highly effective; skilled'	bisa 'venom; ability'

TABLE 2.8 Semantic shifts in loanwords transmitted through Malay

South Asian etymon	Malay	Yakan
ḍāhaga (MIA) 'a burning sensation'	dahaga 'thirst'	dahaga? 'to be greedy (for food only)'
*drohaka (Sk.) 'mischievous; treachery'	dārḥaka 'insurgent; rebellious'	dahulaka? 'destructive'
guliga (MIA) 'kernel'	guliga 'a bezoar-stone'	buliga? 'charm (consisting of stones of beautiful colours or petrified item)'

TABLE 2.8 Semantic shifts in loanwords transmitted through Malay (*cont.*)

South Asian etymon	Malay	Yakan
tenggara (Ma.) 'southeast'	təŋgara 'southeast; south-eastern wind'	tuŋgara? 'a dry spell'
uttara (Sk.) 'upper; northern'	utara 'north; northerly'	uttala? 'dry season'
vañcana (Sk.) 'deception'	bəncana 'affliction'	binsana? 'to be in a state of suffering'
vicāra (Sk.) 'deliberation; discussion'	bicara 'discussion; to speak'	bissa: 'word, language'
vidyādhari (Sk.) 'a female supernatural being'	bidadari 'nymph'	birarali 'rainbow (sky maiden)'
vināśa (Sk.) 'destruction'	binasa 'destruction, ruin'	binasa 'to kill; having intention to kill, to inflict pain'

TABLE 2.9 Phonological integration in Leti and Rote

South Asian etymon	Malay	Rote	Leti
chalaka (Sk.) 'fraud, deceit'	cəlaka 'misfortune'	silaka	slaka
jāla (Sk.) 'casting net'	jala	dala	diala
kapās (NIA) 'cotton'	kapas	aba	kawsa
kappal (Ta.) 'ship'	kapal	kapa	kapla
laśuna (Sk.) 'onion'	— ^a	laisona	lasoa
vajja (MIA) 'steel'	baja	bai	wai

a Cf. Makassar, Bugis *lasuna*.

2 Timeframe

As mentioned previously, literary and epigraphic attestations can provide some information on the approximate time depth of borrowing. The writing traditions of Cham, Malay, and Javanese can be traced back to respectively the fifth, seventh, and ninth century CE. These classical languages constitute high-prestige registers, in which the amount of Sanskrit loans was presumably higher than in the spoken language. Yet the quantity of Sanskrit vocabulary is still vast in many vernaculars. In terms of tangible items, many names for

TABLE 2.10 Sanskrit words often found as loanwords in Maritime Southeast Asia

Category	Examples
geography	bhūmi 'earth', guha 'cave', koṭa 'fort'
law	doṣa 'transgression', pañjara 'prison', sākṣī 'witness'
materials	kāca 'glass', saindhava 'saltpetre'
numerals	ayuta 'ten thousand', koṭī 'ten million', lakṣa 'hundred thousand' ^a
plants	jambu 'rose apple', kusumbha 'safflower', paṭola 'pointed gourd', tulasī 'holy basil'
products	ghaṇṭā 'bell', jāla 'casting net', madhu 'honey'
religion	dhūpa 'incense', jīva 'life', naraka 'hell', svarga 'heaven'
scholarship	akṣara 'letter', bhāṣā 'language', guru 'teacher', kathā 'speech', paṇḍita 'scholar'
social life	duḥkha 'sorrow', manuṣya 'human', sahodara 'uterine brother', sukha 'happy'
time	kāla 'time', māsa 'month' ^b

a Across Malayo-Polynesian languages, these numerical values have shifted to respectively 'one million', 'hundred thousand', and 'ten thousand'.

b Typically borrowed in the meaning of 'season' or 'period'.

precious minerals, jewels, and metals in the Malayo-Polynesian languages of Maritime Southeast Asia have Indo-Aryan and/or South Dravidian etymologies (Hoogervorst 2013:116–121, 2016:562–568). South Asian loanwords also occur in several other domains. Table 2.10 above lists some widespread Sanskrit loans in the languages of western Indonesia and, to a lesser extent, the Philippines.

The lexical influence of Pali, the liturgical and intellectual language of Theravāda Buddhism, has been considerable in Mainland Southeast Asia. By contrast, very few loanwords in Austronesian languages can be identified as originating from Pali, and those that look phonologically similar are better explained as MIA borrowings (Hoogervorst 2017a). Whenever we do find Pali influence, it is invariably transmitted through a non-Austronesian language. In the case of Moklenic languages, a low-order branch found around the Mergui Archipelago, such vectors include Old Mon and Thai, as will be discussed below (see Table 2.13). For Cham, the situation is more complex. Old Cham borrowed directly from Sanskrit but in modern Cham, spoken in different varieties on the Southeast Asian mainland, we find a number of South Asian (re)borrowings that appear to have entered the language through Khmer on account of their phonological shape. Among other things, this can be seen from the elision of the word-final short /a/ (Table 2.11).

TABLE 2.11 South Asian loans in Cham borrowed through Khmer

South Asian etymon	Old Khmer	Cham	Malay
āditya (Sk.) 'the sun'	ādity	adit	–
anyāya (Sk., Pa.) 'injustice'	anyāy	iniai 'to be bewitched'	aniaya
āyus (Sk.), āyu (Pa.) 'life'	āyuḥ	āyuḥ	–
bala (Sk., Pa.) 'forces'	bal	bal	bala
budha (Sk.) 'Mercury, Wednesday'	budh	but	–
campaka (Sk., Pa.) 'champak flower'	cāṃpā	caṃpā	cəmpaka
guru (Sk., Pa.) 'teacher (spiritual)'	grū	grū, gru	guru
kāla (Sk., Pa.) 'time'	kāl	kal 'when; time'	kala
lābha (Sk., Pa.) 'receiving; gain'	lābh	lap	laba
pāpa (Sk., Pa.) 'sin'	pāp	pap	–
puṇya (Sk.) 'merit'	puṇ ^a	ḃon	–
rūpa (Sk., Pa.) 'form'	rūp	rūp	rupa
sukha (Sk., Pa.) 'happiness'	sukh	thuk /θuk/	suka
varṇa (Sk.) 'colour'	bār	bar	warna
yakkha (Pa.) 'ogre'	yakkh	yak	–

a Pronounced as /bon/ in contemporary Khmer.

The cultural domains of lexical borrowing speak volumes about the nature of historical contact. In addition to the practical items and concepts mentioned previously, Sanskrit words prevail in the domains of religion, mythology, governance, toponyms, and royal titles (Gonda 1973: 216–353). In addition, a number of common words in the languages of Java and Sumatra consist of Sanskrit elements yet appear to have been formed locally.¹² These include numerous plant names and words like Old Javanese *gajamina* 'a mythological whale' (Sk. *gaja* 'elephant' + *mīna* 'fish') and *mutyahara* 'pearl' (Sk. *mutya* 'pearl' + *hāra* 'garland'), corresponding to *gajah mina* and *mutiara* in Malay.¹³ Many South Asian borrowings pertain to concepts already available in the recipient language. A well-known example in Malay is the substitution of **təlu* 'three' for *tiga*,

12 These also include numerous Indonesian neologisms, such as *basantara* 'lingua franca' (Sk. *bhāṣā* 'language' + *antara* 'in the interior') and *mitra bastari* 'peer reviewer' (Sk. *mitra* 'friend' + Sk. *vistāri* 'great'). See Gonda (1973:626–634) for several older examples.

13 The former appears to be a calque of Old Javanese *iwak liman*, whereas the latter corresponds to *muktāhāra* in Sanskrit. Both *mutya* and *muktā* are back-formations (cf. MIA *muttā*, *mottā*) ultimately reflecting a Dravidian precursor (Turner 1966:584).

TABLE 2.12 Sanskrit loans substituting inherited vocabulary in Malay

Sanskrit	Malay	
	Sanskrit loan	Inherited equivalent
bāhu ‘upper arm’	bahu ‘shoulder’	(PAN *qabaRa)
gaja ‘elephant’	gajah	liman
kapāla ‘skull’	kāpala ‘head’	hulu
mālatī ‘jasmine’	məlati	məlur
mukha ‘mouth, face’	muka ‘face’	(PAN *daqis ‘forehead; face’)
nīra ‘water, juice’	nira ‘palm juice’	lahaj
phala ‘nutmeg’	pala	?
samudra ‘ocean’	samudra	lautan
sūrya ‘sun’	surya	matahari

presumably borrowed from MIA **tiga* ‘triple’ (Dyen 1946; Hoogervorst 2017a: 414–415). Equally illustrative is the co-existence in Malay and several other Austronesian languages of *manḡa* (< Ta. *māṅkāy*), *məmpəlam* (< Ta. *māmpalam*), and *pauh* (< PMP *pahuq) for ‘mango’ (Mahdi 2007:46–47). Conceivably, these forms originally denoted different cultivars or ripening stages of the same fruit. Additional examples of Sanskrit “luxury loans” in Malay are given in Table 2.12.

As mentioned previously, the antiquity of South Asian loanwords can at times be gauged from the phonological regularity of their tentative reconstructions. A number of Indo-Aryan and South Dravidian loans regularly reconstruct back to a proto Malayo-Polynesian level, while others have previously been assigned a “proto Western-Malayo-Polynesian” pedigree (Hoogervorst 2016). This number increases for low-order branches of the Austronesian language family. Table 2.13 lists some regular proto Moklenic reconstructions, which I postulate go back to Indo-Aryan etyma through intermediate languages such as Malay, Old Mon, and Thai.

As mentioned previously, the Old Javanese literature provides rough insights into the timeframe of lexical borrowing. Accordingly, the influence of Tamil turns out to be of considerable antiquity. A number of Tamil loanwords are found in Old Javanese inscriptions and literary texts predating the thirteenth century (Hoogervorst 2015). Some examples are listed in Table 2.14.

Absences have analytical value as well. The non-attestation in the vast Old Javanese textual record of some widespread Tamil loans in Malay, modern Javanese, and other Austronesian languages presumably indicates a more re-

TABLE 2.13 Proto Moklenic reconstructions borrowed from Indo-Aryan languages

Indo-Aryan etymon	Intermediate source	Proto Moklenic
gaja (Sk., Pa., MIA) ‘elephant’	gajah (Malay)	*gajah
hattha (Pa., MIA) ‘cubit’	hat (Old Mon, Thai)	*hat
jāla (Sk., Pa., MIA) ‘casting net’	jan (Thai)	*pə-ja:n
kācaka (Sk.) ‘glass’	krajok (Thai)	*kecək
manuṣya (Sk.), manussa (Pa.) ‘human being’	manut (Thai)	*manut
marīca (Sk.), marica (Pa.) ‘pepper’	mrek (Old Mon)	*melek
panasa (Sk., Pa.) ‘jackfruit’	panah (Old Mon) or panaih (Acehnese)	*paneh

TABLE 2.14 Tamil loanwords in Old Javanese

Tamil	Old Javanese
ceppu ‘small box’	cupu
kaṭai ‘shop’	gaḍay, gaḍe ‘pawning’
kaṭṭi ‘a weight unit’	kaṭi, kati
kayappū ‘an aquatic flower’	kayapu
koṅṭi ‘prostitute; concubine’	guṅḍik ‘female attendant’
pānai ‘earthen pot’	panay, pane
paricai ‘shield’	parisya, parise, parsi
uṅkal ‘limestone’	wuṅkal ‘boulder’
uṅṭai ‘ball’	uṅḍi
viricu ‘a kind of rocket’	mərəcu, mərču ‘fireball (from the sky)’

cent transmission. Examples in this category are given in Table 2.15, which juxtaposes Tamil loans in Malay, modern Javanese, and Tausug, a language of the southern Philippines.

In a relatively small number of cases, loanwords in the above category reveal clear South Dravidian origins but cannot be derived from a Tamil etymon. Table 2.16 lists some examples of borrowings that presumably spread eastwards through Malayalam.

TABLE 2.15 Tamil loanwords in Austronesian languages

Tamil	Malay	Javanese	Tausug
appam ‘a round rice flour cake’	apam	apəm	apam
cauttu ‘pattern, sample, model’	contoh	conto	suntu-an
cukkai ‘passage money’	cukai	–	sukay
kaḷutai ‘donkey’	kələdai	kuldi	–
kappal ‘ship’	kapal	kapal	kappal
kaṭṭil ‘cot; bedstead’	katil	kaṭṭil, kaṅṭṭil	kantil
kāval ‘guard’	kawal	kawal	
māḷikai ‘palace’ ^a	maligai	malige	ma:ligay
mīcai ‘moustache’	misai	–	misay
mutal ‘capital’	modal	moḍal	muddal
pāvāṭai ‘a cloth used as a seat for important people’	puadai	puwaḍe	–
puṭṭu ‘a steamed snack of rice flour’	putu	puṭu	putu
taṅṭu ‘palanquin’	tandu	taṅḍu	–
veṭṭil ‘explosion’	bəḍil ‘rifle’	bəḍṭil	–
vilāṅku ‘fetters’	bələṅgu	bləṅgu	biləṅgu?

- a The meaning has shifted in Javanese to ‘throne’ and in Tausug to ‘a (small) house-shaped receptacle containing confections and money (which is carried on the shoulders of two men in an Islamic Studies graduation procession or a wedding procession); miniature ceremonial palace’.

TABLE 2.16 Malayalam loanwords in Austronesian languages

Malayalam	Malay	Javanese	Tausug
kiḷikkatti ‘areca nut slicer’	kələkati	– ^a	kakati
paṅṅikkar ‘martial arts expert’	pəndekar	paṅḍekar	pandikal ‘wise’
paravadāni ‘a carpet’	pərmadani	praṅwədani	palmaddani?
saṅṅambi ‘a structure near the outside of a building’	sərambi	srambi	–
tenkara ‘southeast’	təṅgara	tunḡara	tunḡara?

- a Compare Old Sundanese *kalakatri*, which should probably be read as *kalakaṭi* since ⟨tr⟩ and ⟨ṭ⟩ are spelled identically in the Indic writing system of this language (Balogh & Griffiths 2020:21).

3 Trajectories

As mentioned in the previous two sections, Javanese and especially Malay were the chief vectors of lexical transmission into Maritime Southeast Asia and beyond. On limited occasions, European languages played a comparable role in later stages of history. We may think of Portuguese in the case of Timorese languages, as shown below, or English in British Malaya. These recent borrowings lack the wide geographical distribution of earlier loans acquired through Malay and Javanese. A common shibboleth of European intermediacy is the addition of a “plural” /s/ to certain product names. We may, for example, assume that such words as Malay *durias* ‘course muslin’ (< Hi. *ḍoriyā*), *gauris* ‘cowry shell’ (< Hi. *kaurī*), and *giras* ‘a coarse cloth’ (< Hi. *gārhā*) entered Southeast Asia through Dutch, English, or Portuguese. The early-modern period also saw European loanwords transmitted by South Asians, especially in British Malaya. Some Malay words, in turn, spread to South Asia in this period (Hoogervorst 2013:32, 33, 35).

In some cases, lexical borrowing from Indo-Aryan and/or South Dravidian languages took place directly, rather than through Malay or Javanese. This was particularly the case in Sumatra, the Indonesian island closest to the Indian Subcontinent. In Acehnese, spoken on Sumatra’s westernmost tip, we find several loanwords not attested in other Austronesian languages (Table 2.17). The fact that these loanwords can be traced to relatively modern languages and did not find their way into Malay indicates that their transmission is of no great antiquity.

Another Sumatran speech community that has been in direct contact with South Asia are the Karo Batak. The presence in North Sumatra of medieval trading guilds from South India is well documented archaeologically and epigraphically. A small number of Karo Batak family names (*marga*) have been identified as South Dravidian in origin (Kern 1903; van Ronkel 1918), whereas lexical influence has been observed in the medieval Tamil word *ūrōm* ‘village assembly’, which reportedly gave rise to *uruṅ* ‘alliance; federation of different villages’ in Karo Batak and some closely related languages (Edwards McKinnon 1996:93).¹⁴ Additional Tamil loans in Karo Batak are listed in Table 2.18.

North Sumatra’s Batak languages have also undergone lexical influence from Sanskrit, including in the names of the wind directions, months, days of the week, and zodiac (Voorhoeve 1972; Gonda 1973:119–130; Parkin 1974). Interest-

14 If so, the innovation from *m > ŋ in word-final position needs further explanation.

TABLE 2.17 South Asian loanwords found in Acehese but not in other Austronesian languages

South Asian etymon	Acehnese
bel (Hi., Be.) ‘wood-apple tree’	bi
bhaṅgī (Hi.) ‘person addicted to drinking <i>bhaṅg</i> ’	baṅgi ‘opium addict’
cansur (Hi.) ‘cress (plant)’	camcuruih
daṣi (Be.) ‘wick of a lamp’	daih
kuṛaṭu (Ta.) ‘pincers’	gurudu
pacīsī (Hi.) ‘a kind of game’	pacih
panas (Hi., Be.) ‘jackfruit’	panaih
pīr (Ta.) ‘luffa’	pi?
pukaiyilai (Ta.) ‘tobacco’	pa?ele?
uḷi (Ta.) ‘chisel; engraver’s tool’	uli ‘spanner’

TABLE 2.18 Tamil loanwords in Karo Batak

Tamil	Karo Batak
ciṛṭālī ‘a kind of small <i>tālī</i> given by a paramour to his concubine’	sərtali ‘big, golden necklace worn during ceremonies’ ^a
curai ‘head of an arrow’	sore ‘an old fashioned arrow’
kaṇam ‘trifle, triviality’	kanam ‘fond of jokes, witty, fanciful’
keṭṭam ‘beard’	guram
māttu ‘checkmate’	mətu
oppam ‘ornamentation’	umpam ‘array, finery’
paṭṭam ‘an ornament worn on the forehead by women’	patam ‘a mark on the forehead made with betel saliva’
tukkam ‘sorrow, distress, affliction’	tukam ‘to pay respect during the <i>ngombak</i> ritual’

a I thank Edmund Edwards McKinnon (pers. comm. 2011) for pointing out this etymology.

ingly, many of these borrowings are unattested in Malay but do occur in Old Javanese. Toba Batak furthermore has a number of seemingly unique Sanskrit loanwords, as listed in Table 2.19.

In some cases, a language other than Malay served as the vector of lexical borrowing. In the language of Nias, an island off Sumatra’s west coast, the Min-

TABLE 2.19 Sanskrit loanwords found in Toba Batak but not in other Austronesian languages

Sanskrit	Toba Batak
aṅgāra ‘the planet Mars’	aṅgara ‘third day of the month’
jñāpita ‘made known; taught’	jamita ‘sermon’
pad ‘foot’	pat
pāśa ‘a snare; cord’	pasa ‘rope’
phaṇi ‘serpent’	pane ‘the god of the underworld <i>Pane na Bolon</i> ’
vāda ‘speaking about; discussion; quarrel’	bada ‘quarrel or dispute’

angkabau language appears to have been of greater significance. Table 2.20 lists some South Asian loanwords in Nias and their presumed Minangkabau precursors.

Parts of the Philippines have been in precolonial contact with Borneo, Java, and Sumatra. The presence of Sanskrit words in Tagalog, a language from Luzon, is well known (Kern 1880; Wolff 1976). Many of these words did not make their way into (modern) Malay but can be found in Old Javanese. It is not impossible, however, that they were also once part of the Old Malay vocabulary and simply happen not to occur in the very small corpus of Old Malay texts preserved to us (Adelaar 2009:725). Some examples are listed in Table 2.21.

In Taiwan, the northernmost home of Austronesian languages, early South Asian loanwords are rare. For example, we find one isolated borrowing in Siraya, a now extinct language of Taiwan’s southwestern coast. The word in question is *tabe* ‘a greeting’, presumably from the now obsolete Malay or Javanese *tabik* (Adelaar 1994:57). This word reflects Old Javanese *santabya*, *santawya* ‘may (I) be pardoned, pardon (me)’, Toba Batak *santabi*, Makassar *tabea*, and ultimately Sanskrit *kṣantavya*. It presumably entered Siraya in the seventeenth century, given that many people in service of the Dutch East India Company came from the Indonesian Archipelago. Puyuma *dawa* ‘foxtail millet’ appears to be a borrowing from Maritime Southeast Asia, where the word may have originally denoted ‘sorghum’ (Mahdi 1994: 431–441). It ultimately reflects Sanskrit *yava* ‘barley’.

In Malagasy, a South East Barito language spoken on the island Madagascar, several Sanskrit loanwords have been identified (Dahl 1951:96–119; Adelaar 1994:55–56). Here, the transmission was certainly precolonial. Archaeological evidence points to roughly the seventh to eighth centuries CE as a likely time-

TABLE 2.20 South Asian loanwords in Nias

South Asian etymon	Minangkabau	Nias
āgama (Sk.) ‘a traditional doctrine’	ugamo ‘religion’	ugamo
gōvāla (MIA) ‘cowherd’	gumbalo	kubalo
gula (MIA) ‘sugar’	gulo	gulo
jagga (MIA) ‘to be awake’	jago ‘to guard’	zago
kuñci (NIA) ‘key’	kunci ‘key; to lock’	kusi
kusumbha (Sk.) ‘safflower’	kasumbo ‘red’	kasumbo ‘a citrus fruit’
lasun (Hi.), rasun (Be.) ‘garlic’	dasun	dasu
māmpaḷam (Ta.) ‘mango’	marapalam	marafala
parīkṣā (Sk.) ‘examination’	pareso ‘to examine’	fareso
phala (Sk.) ‘nutmeg’	palo	falo
rāja (Sk.) ‘king’	rajo	razo
rasa (Sk.) ‘essence; taste; love’	raso ‘to feel’	raso
sūtra (Sk.) ‘thread’ ^a	suto ‘silk’	suto
siṁha (Sk.) ‘lion’	siṅo	siṅo
upavāsa (Sk.) ‘to fast’	puaso	fuaso

a But already denoting ‘silk’ in the forms *paṭṭasūtra* and *rāgasūtra* ‘a silk thread’. Hence Old Khmer *sūtra* and Old Javanese *sutra* ‘silk’.

frame for the settlement of the Malagasy speech community from southern Borneo to Madagascar. This is a period in which Malay was already heavily Sanskritized, substantiating the theory that South Asian influence on Malagasy was not direct (Adelaar 1989:32–33). As in the case of Tagalog (Table 2.21), some Sanskrit loanwords in Malagasy are not attested in (modern) Malay but we do find them in Old Javanese. Some examples of these indirect Sanskrit loans in Malagasy are listed in Table 2.22.

On Borneo itself, little evidence has been provided so far of direct contact between South Asian and local languages other than Malay. On the surface, it appears that few of the loanwords found in the languages of Borneo display a great time depth and most are found in Malay as well. However, this may simply reveal a lack of scholarly attention. Table 2.23 lists some examples taken from Smith (2017).

Further to the east, the transmission of South Asian loanwords was primarily the result of language contact with Malay, both for Austronesian and non-Austronesian languages. The North Maluku archipelago—a historical centre of the lucrative spice trade—is home to several “Papuan” languages belonging

TABLE 2.21 Sanskrit loanwords in Tagalog

Sanskrit	Old Javanese	Tagalog
cheda ‘cutting off’	cheda ‘injured, hurt, with a defect’	si’ra ‘a break; damage’
maṇḍala ‘a circle; anything round’	maṇḍala ‘circle; abode of a religious community’	mad’la? ‘the people; the public’
mokṣa ‘liberation; death’	muksa ‘to vanish, disappear’	puksa ‘exterminated; annihilated’ ^a
paribhoga ‘enjoyments’	paribhoga	alibu’gha? ‘irresponsible’
pramāda ‘intoxicated, negligent, careless’	pramada	palamara ‘traitor’
rekhā ‘a line’	rekha ‘line; outward appearance; to give shape to’	li’kha? ‘creation’
rūksa ‘rough; unpleasant’	ruksa ‘dreary, dismal’	luksa ‘in mourning’

a In a number of Malayo-Polynesian languages of Maritime Southeast Asia, the substitution of /m/ for /p/ is common in loanwords that have been interpreted as prenasalised verbs (Hoogervorst 2015:48, 2017a:396).

TABLE 2.22 Sanskrit loanwords in Malagasy

Sanskrit	Old Javanese	Malagasy
āṣāḍha ‘a month (June–July)’	asadha	asara ‘the rainy season’
bhādrapada ‘a month (August–September)’	bhadrawada	vatravatra ‘one of the months’
kārttika ‘a month (October–November)’	kartika	hatsiha ‘the name of a month’
kṣetra ‘field’	setra	hetra ‘feudal land; tax’
māgha ‘a month (January–February)’	magha	maka ‘one of the months’
maṇḍapa ‘open hall’	maṇḍapa	lapa ‘a place of assembly’
mrgaśīrṣa ‘a month (November–December)’	margasira	valasira ‘the harvest season’
tantra ‘the leading or principal or essential part’	tantra ‘illustrative stories (of the <i>nītiśāstra</i>)’	tantara ‘a history; a tale’
yaśa ‘worth; honour’	yasa ‘a meritorious deed’	asa ‘labour, work’

TABLE 2.23 South Asian loanwords in languages of Borneo

South Asian etymon	Malay	Attestation in Borneo
guha (Sk.) 'cave'	goa	goa (Benyadu, Golik, Jangkang, Kendayan, Paser), goha? (Bakumpai, Dusun Witu, Kapuas), gua (Bekati, Dalat, Hliboi Bidayuh, Kanowit, Keninjal, Kereho, Sanggau), guá (Mualang, Ribun), gua: (Iban), guan̄ (Gaii), guha? (Maanyan), guho (Ketapang)
*jāda~*jādi (MIA) 'born; to come into existence'	jadi	jadəy (Kejaman, Seberuang, Sekapan), jadi (Lahanan), jadi? (Kadorih, Ngaju), jadih (Busang, Data Dian), jadiña (Kendayan), jari? (Dusun Witu, Maanyan), mənjadi (Ketapang), mənjadi (Mualang), mənjadi? (Benuaq, Tunjung), ñadi (Keninjal, Iban), ñadin (Dalat) 'to become'
kāraṇa (Sk.) 'cause'	karəna	kaŋɲná (Mualang), karena (Kadorih), karəna? (Gaii), karna? (Kendayan), kaɣəna (Keninjal), kɣəna (Seberuang) 'because'
paricai (Ta.) 'shield'	pərisai	pərisay (Sungkung, Ribun, Paser), pərisay (Golik), pəyisay (Sanggau, Keninjal, Seberuang, Mualang)

to the North Halmahera branch. We find several South Asian words in the local languages, all of which appear to have been transmitted via Malay. By way of illustration, Table 2.24 lists several examples in Ternate and Galela.

The easternmost point of lexical influence from South Asia can be identified as northwest New Guinea. Here, too, Malay played a key role in the transmission of these words. Table 2.25 lists several examples of South Asian loans in the Numfor-Dore dialect of Biak, a language from the Cenderawasih Bay north of New Guinea. A phonological analysis of the Biak data reveals different layers of borrowing. The word *sarak* 'silver', for example, displays both the innovation *l>r and the elision of the historical word-final /a/, precisely as in inherited vocabulary.¹⁵ The words exhibiting a word-final /a/ are more recent acquisitions. Along similar lines, we may assume that *cap* 'to sign' is a relatively new loan on account of its /c/, *fonto* 'similarity' represents an earlier stage of phonological integration, whereas *samara* 'a kind of large machete' is even older, yet still not as old as *sarak*.

15 For example Biak *rim* 'five' from PMP **lima*.

TABLE 2.24 South Asian loanwords in Ternate and Galela

South Asian etymon	Malay	Galela	Ternate
ācāra (Sk.) 'conduct; custom'	cara 'method'	cara	cara
bheda (Sk.) 'breaking; disuniting'	beda 'different'	beida 'not on good terms'	beda 'difference'
buddhi (Sk.) 'intelligence; reason'	budi 'kindness'	budi	budi
doṣa (Sk.) 'transgression'	dosa	dosa	dosa
guru (Sk.) 'teacher'	guru	guru	guru
hasta (Sk.) 'cubit'	hasta	hɑ:sita	hasta
jagga (MIA) 'to be awake'	jaga 'to guard'	jaga	jaga
kappal (Ta.) 'ship'	kapal	kɑ:pali	kapal
kaṭṭi (Ta.) 'a weight unit'	kati	kati	kati
kuñci (NIA) 'key'	kunci	kuci	kuci
kusumbha (Sk.) 'safflower'	kasumba 'a red dye; safflower'	kasuba 'red cotton'	kasuba 'violet'
marīca (Sk.) 'black pepper'	mərīca	rica 'Spanish pepper'	rica
tambaga (MIA) 'copper'	təmbaga	tabaga	tambaga
vaṃśa (Sk.) 'a cane; the line of a pedigree'	baṅsa 'race'	baṅsa	baṅsa 'a nobleman'
vicāra (Sk.) 'deliberation; discussion'	bicara 'to discuss'	bicara	bicara

TABLE 2.25 South Asian loanwords in Biak

South Asian etymon	Malay	Biak
cāmara (Sk., MIA) 'a fly-whisk; a plume on the heads of horses'	cəmara 'ornamental tuft'	samara 'a kind of large machete'
cauttu (Ta.) 'pattern, sample, model'	contoh	fonto 'similarity'
chāp (NIA) 'seal'	cap	cap 'to sign'
gula (MIA) 'sugar'	gula	gura
kaṃsa (Sk.) 'copper'	kaṅsa	kansa
kappal (Ta.) 'ship'	kapal	kapar
kuñci (NIA) 'key'	kunci	kudsi
marīca (Sk.) 'black pepper'	mərīca	marisan 'chili pepper'
śalākā (Sk.) 'a kind of coin'	səlaka 'silver'	sarak

TABLE 2.26 South Asian loanwords in Tetun

South Asian etymon	Malay	Portuguese	Tetun
bhansāl (Hi.) 'shed'	baṅsal	baṅgaçal	baṅga'sal
cakkara (Ma.) 'palm sugar'	–	jagra	jarga
cāmara (Sk., MIA) 'a fly-whisk; a plume on the heads of horses'	cāmara 'ornamental tuft'	–	samara 'plume of dyed animal hair'
kapās (NIA) 'cotton'	kapas	–	kabas
kaṭṭi (Ta.) 'a weight unit'	kati	cate, cates	kati, katis
kauṛī (NIA) 'cowrie shell'	–	caurim	kau'riṅ
kuḷam (Ta.) 'pond'	kolam	–	kolaṅ '(saltwater) swamp; lagoon'
maināttu (Ma.) 'laundry(wo)man'	mānātu	mainato	mainatu
mūṅg (NIA) 'mung bean'	–	mungo	mūṅgu
murūṅkai (Ta.) 'horseradish tree'	māruṅgai	–	maruṅgi
nāma (Sk.) 'name'	nama	–	nama 'namesake'
nel (Ta.) 'harvested rice'	–	néle	neli
paṭola (Sk.) 'pointed gourd'	pəṭola 'loofah; rag gourd'	–	patola
ṭaṅḍel (Hi.) 'coxswain'	tandil	tandel	tan'del

In the south-eastern parts of Maritime Southeast Asia, we find a rather complex history of contact. South Asian loanwords, transmitted through Malay and/or Javanese, are relatively limited in number but can be found in Austronesian and Timor-Alor-Pantar languages alike. In East Timor, a former Portuguese colony, we also find a number of South Asian loanwords that found their way to the island through Portuguese. These loans are not found in Austronesian languages outside the island. Even within East Timor, many appear to be restricted to Tetun, which has received the greatest impact from Portuguese. Table 2.26 above lists a number of South Asian loanwords in Tetun, indicating on the basis of their phonological shape whether they were borrowed through Malay or Portuguese.

A number of South Asian borrowings spread across Maritime Southeast Asia in a morphologically complex form. This seems to be the case, for example, with the Malay word *malas* 'lazy' and its reflexes, which consists of the stative/attributive prefix *ma-* and the base *alas* 'laziness' borrowed from some NIA source (Hoogervorst 2016:580). In other cases, the presence of the prefix *sə-* 'one; the same' reveals Malay as the immediate donor. Some examples are given in Table 2.27.

TABLE 2.27 South Asian loanwords featuring Malay sə-

South Asian etymon	Malay	Example of secondary borrowing
jaitra (Sk.) ‘victorious’	sə-jahtəra ‘tranquility’ ^a	sajahitra? (Tausug)
kāla (Sk.) ‘time’	sə-kali ‘once’	sakayi? (Yakan) ‘when’
kūṭṭu (Ta.) ‘companionship’	sə-kutu ‘cooperative association’	səkuṭu (Javanese) ‘allied’
nityaśa (Sk.) ‘always’	sə-nəntiasa ‘everlasting’	sinittiyasa (Yakan) ‘to worship at all prescribed times’
prati (Sk.) ‘towards’	sə-pərti ‘like; resembling’	saparti (Maranao)

a This form has alternatively been explained as a reflex of Sanskrit *sac-chattrā* ‘with an umbrella’ and hence ‘under government protection’ (Poerbatjaraka 1953:41; Hoogervorst 2015:85). With regard to the phonological shape, compare Malay *bahtəra* ‘ship’ from Sanskrit *vahitra*.

In other instances, morphology reveals a Javanese transmission. Reflexes of Old Javanese *panjyut* ‘lamp; torch’, reflecting Sanskrit *jyut* ‘to shine’ combined with the substantive prefix *paN-*, can be found from Sumatra to Maluku (Lafeber 1922:147–148; Mills 1981:69). The words Sk. *vaśa* ‘power’ and Ta. *viricu* ‘a kind of rocket’ yielded Old Javanese *ka-wasa* ‘overpowered; in the power of’ and modern Javanese *mərco-n* ‘fireworks’, which were in turn adopted by other Austronesian languages (e.g. Malay *kuasa*, *mərcun* and Makassar *koasa*, *baraccuy*). Old Javanese *ajar-an* ‘horse’ (Javanese *jaran*) is derived from the aforementioned base *ajar* ‘teaching’ and has been adopted in languages of Borneo, Sulawesi, and Nusa Tenggara, e.g. Ngaju Dayak *hajaran*, Banggai *ajalan*, Tae’ *daraj*, Makassar *jaraj*, Bimanese *jara*, Komodo *jaraj*, Manggarai *jaraj*, Ngadha *dzara*, and Kambara *njara*.¹⁶

We find several more examples of South Asian loanwords unattested in Malay yet found in languages from Java, North Sumatra and South Sulawesi. They may have existed in an earlier stage of Malay but might also reflect direct contact with Indo-Aryan languages. Table 2.28 lists some examples.

16 Data taken from Blust & Trussel (ongoing).

TABLE 2.28 Indo-Aryan loanwords not found in Malay

Indo-Aryan etymon	Austronesian attestations
āyoga (Sk.) 'a yoke for draft animals'	auga (Toba Batak), ioga (Karo Batak), ayoka (Makassar), ajoa (Bugis)
dravya (Sk.) 'object of possession, wealth, goods, money'	drawya, drabya (Old Javanese) 'property, what belongs to', duwe (Javanese) 'to own', rubia (Karo Batak) 'animal', dorbia (Toba Batak) 'domestic animals', Gayo durubiu
lašuna (Sk.) 'garlic', lasun (Hi.), rasun (Be.)	lasuna (Makassar, Bugis, Toba Batak, Karo Batak), lasona? (Maranao), jasun (Old Javanese), dasun (Minangkabau), lasun (Gayo)
nāyaka (Sk.) 'chief, leader'	nayaka (Old Javanese, Javanese), layaka (Makassar, Bugis)
panasa (Sk.) 'breadfruit', panas (Hi., Be.)	panasa (Old Javanese, Bugis), panasa? (Makassar), pinasa (Toba Batak), panaih (Acehnese)

4 Concluding Remarks

The lexical data examined here afford a number of observations. Firstly, as the Old Javanese corpus reveals, Sanskrit, Tamil, MIA, and even NIA words show up in Maritime Southeast Asia at roughly the same time. This indicates that different parts of South Asia were in contact with different parts of Southeast Asia. Through Javanese, Malay, or both, some South Asian loanwords travelled north through the Philippines, east to Nusa Tenggara, and possibly west through Madagascar. The semantics observed in Old Javanese furthermore reveal how the meanings of ancient South Asian words changed over time, offering in many cases a missing link to contemporary reflexes.

An even greater role was played by people who spoke and/or wrote Malay. The amount of early texts in this language is much smaller compared to Javanese, leaving us relatively ignorant about the Old Malay lexicon. Its geographical influence appears to have surpassed that of Javanese. Lexical influence from Malay is found across Maritime Southeast Asia, Madagascar, the western parts of New Guinea, the Southeast Asian mainland, and Taiwan. This includes inherited vocabulary, South Asian loans, and Arabic loans, which in some areas seem to have travelled as a package. In certain languages, such as Biak and Rote, multiple layers of loanwords can be identified on the basis of historical phonology. As standard Indonesian continues to influence all languages from Sumatra to New Guinea, this process is arguably still ongoing. Such

relatively modern borrowings tend not to undergo high levels of phonological alteration.

In a small number of cases, particularly in Sumatra, direct contact with the Indian Subcontinent—that is, without Malay or Javanese as intermediary languages—is in evidence. In the Batak speech communities, contact appears to have taken place from the eleventh to the fourteenth century. In Acehnese, this contact continued into colonial times. In both cases, we find Indo-Aryan as well as South Dravidian loanwords. Another contact scenario is presented by the Chamic and Moklenic languages of the Southeast Asian mainland. Here, Khmer (in the case of Chamic) and Old Mon and Thai (in the case of Moklenic) played a role in the transmission of South Asian vocabulary, although Old Cham also borrowed directly from Sanskrit. Only in these two subgroups do we find some plausible evidence of loanwords from Pali, as opposed to Sanskrit or MIA. In general, Austronesian languages show greater quantities of South Asian loanwords than “Papuan” languages.

Acknowledgements

I am grateful to Arlo Griffiths and to an anonymous peer-reviewer for their valuable suggestions to improve this chapter.

Language Sources

Acehnese	Djajadiningrat 1934
Bengali	Biswas 2000
Biak	van Hasselt 1876
Bugis	Matthes 1874
Cham	Aymonier & Cabaton 1906
Galela	van Baarda 1895
Gayo	Hazeu 1907
Hindustani	Platts 1884
Javanese	Robson & Wibisono 2002
Karo Batak	Prinst 2002
Makassar	Cense 1979
Malagasy	Richardson 1885
Malay	Wilkinson 1932
Malayalam	Gundert 1962
Maranao	McKaughan & Al-Macaraya 1996

MIA (Middle-Indo-Aryan)	Turner 1966
Minangkabau	Moussay 1995
NIA (New-Indo-Aryan)	Turner 1966
Nias	Laiya et al. 1985
OIA (Old-Indo-Aryan)	Turner 1966
Old Javanese	Zoetmulder 1982
Old Khmer	Jenner 2009
Old Mon	Shorto 1971
Pali	Rhys Davids & Stede 1966
Portuguese	Dalgado 1919
PAN (Proto Austronesian)	Blust & Trussell ongoing
PMP (Proto Malayo-Polynesian)	Blust & Trussell ongoing
Proto Moklenic	Larish 1999
Sanskrit	Monier-Williams 1899
Siraya	Adelaar 2011
Tagalog	Ferrer 2003
Tamil	<i>Tamil</i> 1924–1936
Tausug	Hassan et al. 1994
Ternate	de Clercq 1890
Tetun	Hull 1999
Toba Batak	Warneck 1977
Yakan	Behrens 2002

Bibliography

- Adelaar, K. Alexander. 1989. 'Malay Influence on Malagasy. Linguistic and Culture-Historical Implications.' *Oceanic Linguistics* 28(1):1–46.
- Adelaar, K. Alexander. 1994. 'Malay and Javanese loanwords in Malagasy, Tagalog and Siraya (Formosa).' *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 150(1):50–65.
- Adelaar, K. Alexander. 2009. 'Loanwords in Malagasy.' In: Martin Haspelmath and Uri Tadmor (eds), *Loanwords in the World's Languages. A Comparative Handbook of Lexical Borrowing*. Berlin: Mouton de Gruyter, pp. 717–746.
- Adelaar, K. Alexander. 2011. *Siraya. Retrieving the Phonology, Grammar and Lexicon of a Dormant Formosan Language*. Berlin: Walter de Gruyter.
- Aymonier, Étienne and Antoine Cabaton. 1906. *Dictionnaire Čam-Français*. Paris: Ernest Leroux.
- Baarda, M.J. van. 1895. *Woordenlijst Galèlareesch-Hollandsch*. The Hague: Martinus Nijhoff.

- Balogh, Dániel and Arlo Griffiths. 2020. 'DHARMA Transliteration Guide.' halshs-02272407v3.
- Behrens, Dietlinde. 2002. *Yakan-English Dictionary*. Manila: Linguistic Society of the Philippines.
- Biswas, Sailendra. 2000. *Samsad Bengali-English dictionary*. Calcutta: Sahitya Samsad. Third edition.
- Blust & Trussell. Ongoing. *The Austronesian Comparative Dictionary* (www.trussellz.com/ACD).
- Burrow, T. and M.B. Emeneau. 1984. *A Dravidian Etymological Dictionary*. Oxford: Clarendon Press. Second edition.
- Casparis, J.G. de. 1986. 'Some Notes on Words of "Middle-Indian" Origin in Indonesian Languages (Especially Old Javanese).' In: Luigi Santa Maria, Faizah Soenoto Rivai and Antonio Sorrentino (eds), *Papers from the 111 European Colloquium on Malay and Indonesian Studies (Naples, 2-4 June, 1981)*, pp. 51-69. Naples: Instituto Universitario Orientale.
- Cense, A.A. 1979. *Makassaars-Nederlands woordenboek*. The Hague: Martinus Nijhoff.
- Clavé, Elsa and Arlo Griffiths. 2022. 'The Laguna Copperplate Inscription: Tenth-Century Luzon, Java, and the Malay World.' *Philippine Studies: Historical and Ethnographic Viewpoints* 70(2):167-242.
- Clercq, F.S.A. de. 1890. *Bijdragen tot de kennis der Residentie Ternate*. Leiden: E.J. Brill.
- Dahl, Otto Christian. 1951. *Malgache et Maanjan*. Oslo: Egede-Instituttet.
- Dalgado, Sebatião Rodolfo. 1919. *Glossário Luso-Asiático*. Coimbra: Imprensa da Universidade.
- Djajadiningrat, Hoesein. 1934. *Atjèhsch-Nederlandsch woordenboek*. Batavia: Landsdrukkerij.
- Dyen, Isidore. 1946. 'Malay tiga "three".' *Language* 22(2):131-137.
- Edwards McKinnon, E. 1996. 'Mediaeval Tamil Involvement in Northern Sumatra, c11-c14 (The Gold and Resin Trade).' *Journal of the Malaysian Branch of the Royal Asiatic Society* 69(1):85-99.
- Ferrer, Alicia S. 2003. *Diksyunaryo Filipino—English*. Manila: MECS Publishing House.
- Gonda, J. 1973. *Sanskrit in Indonesia*. New Dehli: International Academy of Indian Culture. Second edition.
- Griffiths, Arlo. 2018. 'Four More Gupta-period Copperplate Grants from Bengal.' *Pratna Samiksha* 9:15-57.
- Griffiths, Arlo and Amandine Lepoutre. 2016. 'Études du corpus des inscriptions du Campā, VIII. Les inscriptions des piédroits des temples de Po Klaong Girai (C. 8-11), de Linh Thái (C. 109-110) et de Yang Prong (C. 116).' *Bulletin de l'Ecole française d'Extrême-Orient* 102:195-296.
- Gundert, H. 1962. *A Malayalam and English dictionary*. Kottayam: Sahitya Pravarthaka. Second edition.

- Hajek, John and Catherina Williams-van Klinken. 2019. 'Language Contact and Gender in Tetun Dili. What Happens When Austronesian Meets Romance?' *Oceanic Linguistics* 58(1):59–91.
- Hasselt, J.L. van. 1876. *Hollandsch-Noemfoorsch en Noemfoorsch-Hollandsch Woordenboek*. Utrecht: Kemink & Zoon.
- Hazeu, G.A.J. 1907. *Gajösch-Nederlandsch woordenboek met Nederlandsch-Gajösch register*. Batavia: Landsdrukkerij.
- Hoogervorst, Tom G. 2013. *Southeast Asia in the Ancient Indian Ocean World*. Oxford: Archaeopress.
- Hoogervorst, Tom G. 2015. 'Detecting Pre-Modern Lexical Influence from South India in Maritime Southeast Asia.' *Archipel* 89:63–93.
- Hoogervorst, Tom G. 2016. 'Problematic Protoforms. Some "Hidden" Indic Loans in Western Malayo-Polynesian Languages.' *Oceanic Linguistics* 55(2):561–587.
- Hoogervorst, Tom G. 2017a. 'The Role of "Prakrit" in Maritime Southeast Asia through 101 Etymologies.' In: Andrea Acri, Roger Blench and Alexandra Landmann (eds), *Spirits and Ships. Cultural Transfers in Early Monsoon Asia*, pp. 375–440. Singapore. ISEAS.
- Hoogervorst, Tom G. 2017b. 'Lexical Influence from North India to Maritime Southeast Asia. Some New Directions.' *Man in India* 97(1):293–334.
- Hull, Geoffrey. 1999. *Standard Tetum-English Dictionary*. Winston Hills, NWS: Sebastião Aparício da Silva Project.
- Irene U., Nurhadan Halud, Seymour A. Ashley and Mary L. Ashley. 1994. *Tausug-English Dictionary. Kabtangan Iban Maana*. Manila: Summer Institute of Linguistics.
- Jenner, Phillip. 2009. *Dictionary of pre-Angkorian Khmer and Dictionary of Angkorian Khmer*. Canberra: Pacific Linguistics Series 597. Accessed through <http://sealang.net/> (2020)
- Kern, H. 1880. 'Sanskritsche Woorden in het Tagala.' *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 28(3):535–564.
- Kern, H. 1903. 'Drawidische Volksnamen op Sumatra.' *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 55(3):358–362.
- Lafeber, Abraham 1922. *Vergelijkende klankleer van het Niassisch*. The Hague: Hadi Poes-taka.
- Larish, Michael D. 1999. *The Position of Moken and Moklen within the Austronesian Language Family*. University of Hawai'i, PhD dissertation.
- Laiya, Sitasi Z., Siswanto Zagoto, Happy Laiya, Selamat Zagoto and Amita Zagoto. 1985. *Kamus Nias-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Lepoutre, Amandine. 2013. 'Études du corpus des inscriptions du Campā IV. Les inscriptions du temple de Svayamutpanna. Contribution à l'histoire des relations entre les pouvoirs cam et khmer (de la fin du XI^e siècle au début du XIII^e siècle).' *Journal Asiatique* 301(1):205–278.

- Mahdi, Waruno. 1994. 'Some Austronesian Maverick Protoforms with Culture-Historical Implications—II.' *Oceanic Linguistics* 33(2):431–490.
- Mahdi, Waruno. 2007. *Malay Words and Malay Things. Lexical Souvenirs from an Exotic Archipelago in German Publications before 1700*. Wiesbaden: Harrassowitz.
- Matthes, B.F. 1874. *Boegineesch-Hollandsch woordenboek met Hollandsch-Boegineesche woordenlijst en verklaring van een tot opheldering bijgevoegden ethnographischen atlas*. The Hague: M. Nijhoff.
- McKaughan, Howard and Batua Al-Macaraya. 1996. *A Maranao dictionary*. Manila: De La Salle University & the Summer Institute of Linguistics.
- Mills, Roger F. 1981. 'Additional Addenda.' In: Robert Blust (ed.), *Historical Linguistics in Indonesia I*, pp. 59–82. Jakarta: NUSA.
- Monier-Williams, Monier. 1899. *A Sanskrit-English Dictionary. Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*. Oxford: Clarendon Press.
- Moussay, Gérard. 1995. *Dictionnaire Minangabau—Indonesien—Français*. Paris: l'Harmattan.
- Parkin, H. 1978. *Batak Fruit of Hindu Thought*. Madras: Christian Literature Society.
- Platts, John T. 1884. *A dictionary of Urdu, classical Hindi, and English*. London: W.H. Allen.
- Poerbatjaraka. 1953. 'Beberapa Kata.' *Bahasa dan Budaya* 1(3):42–44.
- Prinst, Darwin. 2002. *Kamus Karo-Indonesia*. Medan: Penerbit Bina Media.
- Rhys Davids, T.W. and William Stede (eds). 1966. *The Pali Text Society's Pali-English dictionary*. London: Luzac, 1966. Reprint.
- Richardson, James. 1885. *A New Malagasy-English Dictionary*. Antananarivo: The London Missionary Society.
- Robson, Stuart and Singgih Wibisono. 2002. *Javanese English Dictionary*. Singapore: Periplus.
- Ronkel, Ph.S. van. 1902. 'Het Tamil-element in het Maleisch.' *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 45:97–119.
- Ronkel, Ph.S. van. 1903. 'Tamilwoorden in Maleisch gewaad.' *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 46:532–557.
- Ronkel, Ph.S. van. 1918. 'Drawidische Volksnamen op Sumatra.' *Bijdragen tot de Taal-, Land- en Volkenkunde* 55(1):358–362.
- Shorto, Harry L. 1971. *A Dictionary of the Mon Inscriptions from the Sixth to the Sixteenth Centuries*. London: Oxford University Press.
- Smith, Alexander D. 2017. *The Languages of Borneo. A Comprehensive Classification*. University of Hawai'i, PhD dissertation.
- Tamil*, 1924–1936. *Tamil Lexicon*. [Madras:] University of Madras. Six volumes.
- Turner, Ralph Lilley. 1966. *A Comparative Dictionary of the Indo-Aryan Languages*. London: Oxford University Press.

- Tuuk, H.N. van der. 1971. *A Grammar of Toba Batak*. The Hague: Martinus Nijhoff. Reprint [1864].
- Voorhoeve, P. 1972. 'Sanskrit maandnamen in het Bataks.' *Bijdragen tot de Taal-, Land- en Volkenkunde* 128(4):494–496.
- Warneck, Joh. 1977. *Toba-Batak—Deutsches Wörterbuch*. The Hague: Martinus Nijhoff.
- Wilkinson, R.J. 1932. *A Malay-English Dictionary (Romanised)*. Mytilene: Salavopoulos and Kinderlis.
- Wolff, John U. 1976. 'Malay Borrowings in Tagalog' In: C.D. Coutan and O.W. Wolters (eds), *Southeast Asian History and Historiography. Essays Presented to D.G.E. Hall*, pp. 345–367. Ithaca, NY: Cornell University Press.
- Wolff, John U. 2010. *Proto-Austronesian Phonology with Glossary*. Ithaca: Cornell Southeast Asia Program.
- Zoetmulder, P.J. 1982. *Old Javanese-English Dictionary*. The Hague: Martinus Nijhoff.

Traces of Pre-modern Contacts between Timor-Alor-Pantar and Austronesian Speakers

Marian Klamer

Introduction

The Timor-Alor-Pantar (TAP) family are an outlier “Papuan” group, located some 1,000 kilometers west of the New Guinea mainland, see Figure 3.1 and Figure 3.2.¹ The TAP family constitutes of some 25 languages, and has two subgroups in Timor and one subgroup in Alor and Pantar, as indicated in Figure 3.1 below.

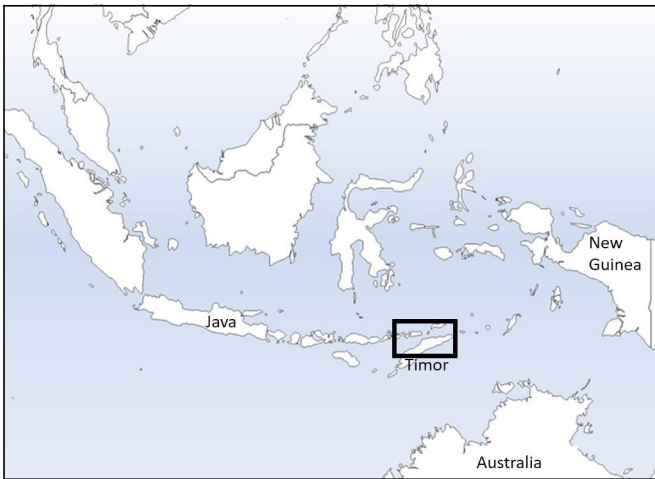


FIGURE 3.1 Location of the Timor-Alor-Pantar languages in Indonesia

1 The term Papuan is used here as a cover term for the hundreds of languages spoken in New Guinea and its vicinity that are not Austronesian (Ross 2005: 15), it says nothing about the genealogical ties between the Papuan families in that area.

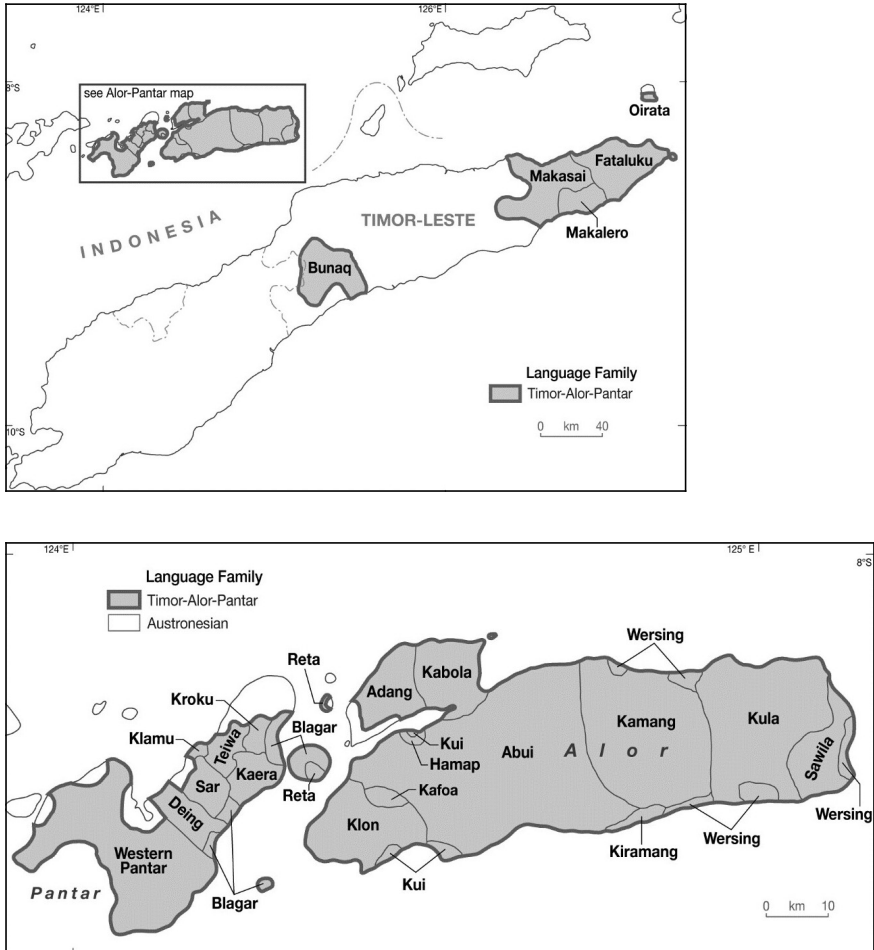


FIGURE 3.2 The Timor-Alor-Pantar languages

The origin and age of the TAP family is unclear. One hypothesis holds that they are descendants of immigrants from New Guinea who arrived in the Lesser Sundas 4,500–4,000 Before Present (BP) and genealogically affiliated with the Trans New Guinea family (cf. Wurm, Voorhoeve, McElhanon 1975, Ross 2005) but the lexical evidence is currently insufficient to support this affiliation (Holton & Robinson 2017b). However, Holton & Robinson (2017b: 183–184) suggest that it is possible that the TAP and the languages on the Bomberai peninsula, West Papua, are related either via a deep genealogical connection or via a more casual contact relationship. If it is a genealogical relationship, it is not yet clear whether they are both part of TNG or whether they share a relationship independent of that family.

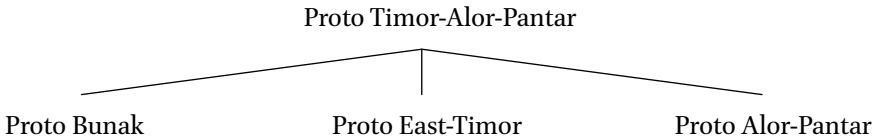


FIGURE 3.3 The three subbranches of the Timor-Alor-Pantar family

HOLTON ET AL. 2012; HOLTON AND ROBINSON 2017B; SCHAPPER ET AL.,
2017

Ancient Malayo-Polynesian (MP)² loans found across the TAP family that show regular sound correspondences suggest that Proto TAP had been in contact with Austronesian languages before splitting up (see section 2 below). As speakers of one or more Austronesian languages are commonly assumed to have arrived in the East Timor area 3,800 BP, that would give the TAP family a maximum age of some 3,800 years. This is relatively young in light of the history of human presence on the islands, which dates back to 42,000 BP in East Timor (O'Connor, Ono & Clarkson 2011), and to 12,000 BP on Alor (O'Connor 2017).

Currently, Alorese is the only one indigenous Austronesian language spoken on the islands of Alor and Pantar. Alorese is closely related to Lamaholot, spoken in the Flores-Lembata region to the west of Pantar (Klamer 2012; Fricke 2019), and speakers of Alorese arrived in the area of Pantar and Alor in the 14th Century (Klamer 2011). On Timor, three TAP languages (Makalero, Makasae and Fataluku) are spoken in contiguous areas in the east of the island, one (Oirata) on Kisar island off the eastern tip of Timor, adjacent to an Austronesian language, and one (Bunak) in the centre of the island, surrounded by Austronesian languages.

The recent publication of the online database *LexiRumah* (Kaiping, Edwards & Klamer 2019) containing lexical data for 357 language varieties spoken in eastern Indonesia and Timor-Leste enables a comparison of lexical data that was previously impossible. In addition, recent years have seen publications of grammar descriptions and historical reconstructions of TAP languages (see the overviews in as well as reconstructions of Austronesian language groups of the Flores-Lembata region (Fricke 2019) and the Timor region (Edwards 2021)). Thus we are now in the position to examine the contact history in the region more closely. Is there lexical evidence that there was contact between speak-

2 In Island SE Asia, languages of the Malayo-Polynesian branch of the Austronesian language family are spoken. This paper refers to these languages interchangeably as 'Austronesian' or 'Malayo-Polynesian (MP)'. However, in the reconstructed forms, a distinction is made between Proto Austronesian (PAN) and Proto Malayo-Polynesian (PMP).

ers of TAP with Austronesian languages? Which languages or regions were the donors, and which were the recipients of lexical and grammatical features? Can we use the evidence to reconstruct stages or regions of contact?

In this chapter, I focus on traces of Austronesian words attested in the lexicon compiled for the TAP languages, as present in the online lexical database *LexiRumah* 3.0.0.; see also below. (TAP borrowings ending up in Austronesian languages of the region are discussed in the chapters by Moro et al. (this volume) and Schapper & Huber (this volume)). In addition, I only focus on ancient and pre-modern borrowings. Ancient loanwords that were inherited throughout the family help us to date the first contact with Austronesian and the age of the TAP family as a whole, as mentioned above. Pre-modern loans are examined because these provide a view on the history of TAP language communities in the period before Indonesian and local Malay became dominant—if we can couple the loans with what little is known about the history of TAP communities in general. For convenience sake, ‘pre-modern’ is defined here as the time between approx. a century ago (100 BP) and the ‘ancient’ period when Proto TAP may have existed, some time around 4000 BP. Over the last hundred years, Malay³ and Indonesian have been increasingly used as languages for interethnic communication in Indonesia; while in Timor Leste, Tetun and Indonesian have (had) that function. This ‘pre-modern’ period is an extremely long time, from which for the Timor-Alor-Pantar region very little is known beyond scattered colonial sources and local oral histories as compiled and analysed in sources such as Hägerdal (2010b; 2010a; 2011; 2012) and Wellfelt (2016). Loans that point to modern contact with Malay, Indonesian or Tetun are outside the scope of the present paper. Such loans, often denoting foreign or non-indigenous objects and concepts, have been adopted across all the TAP languages. Examples include forms similar to Indonesian *dapur* ‘kitchen’, *nangka* ‘jackfruit’, *lampu* ‘lamp’ (< Dutch *lamp*), *lilin* ‘candle’, *tali* ‘rope’, *pasar* ‘market’, *jendela* ‘window’ (< Portuguese *janela*), *gereja* ‘church’ < Portuguese *igreja* ‘church’.⁴

3 Note that on Alor and Pantar, in places like the capital Kalabahi, a local variety of Malay referred to as Alor Malay was already spoken before the advent of Indonesian. Malay has been the lingua franca in eastern Indonesia for centuries. Because of the lexical similarities between Malay and Indonesian, current speakers on Alor and Pantar consider Alor Malay as the colloquial variety of standard Indonesian, even though the two languages have very different histories.

4 Overall, the amount of Indonesian loanwords in word lists of TAP languages is limited. Klammer (2020) found 212 Indonesian loans out a total of 23,247 words listed for the 42 TAP varieties in the LexiRumah database. The average number of words on TAP word list is 553, and the number of loans in each variety range from 1–20 loans, with an average of 3.6% loans.

Why study traces of contact that took place in the pre-modern period? Traditional historical comparison and phylogenetic inference (Kaiping and Klammer 2022) both converge on a pattern where Proto TAP (presumably located in Timor) underwent major splits, separating the AP branch that moved out towards Alor-Pantar, see Figure 3.3 above. The next major split was in the AP branch, with a possible homeland in or around the Straits in the West, separating Pantar from the rest of Alor and the languages of the Alor branch spreading east (Holton et al. 2012). Historical reconstruction thus provides a hypothesis on the homelands and internal dispersal of the TAP family. Studying the traces of pre-modern contact with Austronesian languages can provide a complementary angle on the history of the TAP speakers: with whom did they have contact, and what type of contact was it? The current paper seeks to address these questions.

The paper is structured as follows. Section 1 presents details on the lexical materials and the methodology used in the paper. Section 2 discusses three ancient loans, and section 3 ten pre-modern loans, both organised according to the semantic fields to which the loans belong. In section 4, a summary of the findings is presented, followed by a discussion and conclusions in section 5.

1 Present Study: Methods and Materials

Almost all the lexical data discussed in this paper has been drawn from primary sources compiled and referenced in the online lexical database *LexiRumah* 3.0.0 (Kaiping et al. 2019). Where other sources were used, these are provided in the text. This study investigated the vocabulary of 109 lects (i.e. language varieties or dialects) spoken on the islands of Timor, Alor, Pantar, Flores and Lembata: 54 lects belonging to the Timor-Alor-Pantar family and 55 lects belonging to the Malayo-Polynesian subgroup of Austronesian languages.

To find TAP lexeme sets that contained Austronesian borrowings, I first went on a fishing expedition in LexiRumah, considering lexeme sets for 75 pre-selected concepts in the semantic domains (taken from Haspelmath and Tadmor 2009): *Social and political relations*, *Agriculture and vegetation*, *The house*, *Clothing and grooming*, *Food and drink*, *Warfare and hunting*, *Animals*, *Kinship*, *The physical world*, and *The body*. Crosslinguistically, these concepts cover the spectrum from highly borrowable (*Social and political relations*) to borrowing resistant (*The body*) (Haspelmath & Tadmor 2009).

The results of the expedition were mixed. In many sets that contained loans, the loans were scattered or messy and did not allow interesting generalizations or observations. Sporadically observed loans occurring only in one or two TAP

languages were not considered, as such individual cases may be nonce borrowings and are not good evidence for reconstructing a historical context of contact between communities. Moreover, for most of these sporadic loans, the MP donor language cannot be established. Some sets contained no borrowings. Some sets (e.g. on kinship terminology, or concept for animals such as ‘turtle’) had noisy and unreliable data. Kinship terms are notoriously hard to elicit reliably through a lexical survey, and surveys may render different words for different species of animals, e.g. walking and swimming turtles. Finally, some lexeme sets contained suspected borrowings that were impossible to prove because of lack of reconstructed Austronesian forms to compare them with (more on this below).

For the present paper, I made a selection based on a manual inspection of the results of the initial fishing expedition. I focussed on lexeme sets containing demonstrably Austronesian loans and occurring in a sizeable number of TAP languages, so as to allow some generalizations about the scope, direction or source of the borrowing. I selected 13 concepts from the following semantic fields: *Social and political relations*: ‘king/ruler’, ‘slave’; *Agriculture and vegetation*: ‘maize’, ‘seed’; *Clothing and grooming*: ‘needle’, ‘to weave’, ‘sew’; *Food and drink*: ‘salt’; *Animals*: ‘pig’, ‘deer’; *Kinship* ‘bride price’; and *The body*: ‘navel’, ‘breast’, ‘skin’. The sets discussed in this paper are not an exhaustive listing of the borrowings attested; for reasons of space, some TAP lexeme sets with MP loans are left for future analysis.

To prove that a lexeme set was borrowed into TAP languages, it must be demonstrably Austronesian; that is, there must be a Proto Austronesian (PAN) or Proto Malayo-Polynesian (PMP) reconstructed form that has a similar form and meaning. For this evidence I drew on the etymological database by Blust and Trussel (n.d.), as well as recent historical reconstructions done on daughter stages of PMP that are relevant to the area of Alor Pantar and Timor: Proto Flores-Lembata (PFL) (Fricke 2019), located to the west of Pantar island, and Proto Rote-Meto⁵ (PRM) (Edwards 2021), on Timor.

Rote-Meto is a subgroup within a higher order Timor-Babar (TB) subgroup, see Figure 3.4. The Timor-Babar group comprises all the other languages of Timor and the southern Moluccas, and Proto Timor-Babar is a sister to Proto Central-Timor and Helong (Edwards 2018b; 2019; 2020; 2021). It is yet unknown how Proto Flores-Lembata is related to Proto Timor-Babar and Proto Central Timor, except that all of them are regional, low-level subgroupings within Malayo-Polynesian.

5 Meto = Uab Meto, also known as Dawan, Timorese, or Atoni, see Edwards (this volume).

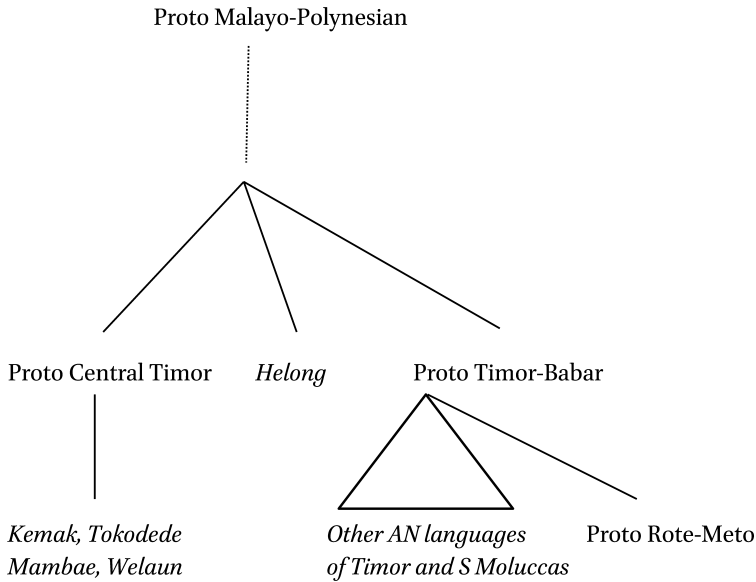


FIGURE 3.4 The MP subbranches of AN languages on Timor

EDWARDS 2018B; 2019; 2020, 2021

In many cases, the sources present reconstructed PMP forms, or they mention sets of related lexemes that cannot (yet) be reconstructed to a common proto form (indicated with a hashtag #).⁶ These two types of Proto Flores-Lembata and Proto Rote-Meto forms were used to compare the TAP data with. In addition, I occasionally considered lexical data from a group of AN languages in central and east Timor that are not grouped under Proto Rote-Meto, but are part of the higher order Timor-Babar subgroup (Edwards 2021), and for which no historical reconstructions are yet available. The diachronic ‘baseline’ form of the TAP languages was determined by considering reconstructed forms for proto TAP (Schapper et al. 2017; Holton et al. 2012; Holton and Robinson 2017a). In sum, I consider both established proto forms and data from low level groups of neighbouring languages to prove that a MP lexeme has entered the TAP languages.

The lexical data is presented below in tables that are organised as follows. The first table presents the available Austronesian data of a particular concept. It contains reconstructed forms from proto MP, Proto Flores-Lembata, and Proto Rote-Meto where available, or it gives representative forms of sets of

⁶ The unreconstructibility of these sets could be due to missing cognates, unexplained irregularities or borrowing.

related lexemes for which no reconstruction was possible (with a hashtag #), and it provides the actual forms of the Austronesian languages of East Timor. The second table contains the TAP data, with reconstructed forms at the top (if any), followed by words attested in the individual languages. In the TAP language table, the languages have been organised by their geographical region, going from west to east: first Pantar-Straits, then West Alor, Central Alor, South Alor, East Alor and ending with East Timor. The organisation of the tables is geographical and does not necessarily reflect genealogical subgroupings.

2 Ancient Loanwords⁷

2.1 *Animals*

2.1.1 'pig'

Pigs appear to have moved through Island SE Asia under human agency as husbanded animals, ultimately from a Southeast Asian source. With the exception of Sulawesi, none of the islands east of the Wallace line possessed endemic populations of pig (*Sus scrofa*, Groves 1981; Glover 1986). In fact, archeological investigations on Flores, Timor, and the northern Mollucas have demonstrated that the first appearance of pigs is associated with the arrival of the 'Neolithic cultural package' during the middle to late Holocene (7000–3500 BP) (Larson et al. 2007).

A form possibly related to PMP *babuy 'pig' is reconstructable as PTAP *baj 'pig' (where /j/ represents a glide), as shown in (1) and (2). The word is inherited across the TAP family with an initial plosive, and follows regular sound changes. This would suggest a very early contact with an Austronesian source at the stage when the TAP family had not yet diversified. If Austronesian groups arrived in

7 Earlier work (Holton et al. 2012:95) has tentatively reconstructed Proto Alor-Pantar (PAP) *bui 'betel nut' as an ancient loan reflecting (< PMP *buaq 'fruit; areca palm and nut', Blust and Trussel n.d.), pointing to the similarity between Alor Pantar lexemes for 'betel nut' and those in nearby Austronesian languages such as Tetun *bua* 'betel', and Tokodede *buo* 'betel'. Here, a discussion of this possible loan has been excluded, because the evidence for it is thin. None of the reflexes in AP languages examined here (except Klamu) has traces of the vowel /a/, instead, virtually all forms reflect /u/ and /i/ or /j/ / (*bui/buj) or reductions thereof (*bu*). In the surrounding Austronesian languages, reflexes include the vowel /a/, so that the formal similarity between AP and AN forms concerns *bu* only. However, Edwards (p.c.) points out that the language of the Babar islands have reflexes of *bui for 'fruit' (< PMP *buaq), and the languages of Aru (e.g. Batuley *bui* 'betel nut', Daigle 2015: 249) do attest an earlier form with a glide, which may constitute support that PAP *bui/buy was indeed an Austronesian borrowing.

the Timor area around 3800 BP (Pawley 2005; Spriggs 2011), that may have been the earliest time the borrowing could have occurred.

The PTAP form *baj reflects the PMP initial plosive /b/ in *babuy, and is a shortened form of the word. In contrast, the reflexes of PMP *babuy attested in the Flores-Lembata and Timor region are all disyllabic, see (1). Also, PFL *vavi or any of its descendants cannot be the donor of PTAP *baj because of the initial fricative. In the Timor region, the AN languages in the east are also unlikely donors because of their initial fricative, as shown in (1). Languages in the west of Timor show reflexes with initial *b, leading to the reconstruction of PRM *bafi. Thus, presently available evidence suggests a loan event involving an ancestor of the Timor languages that is at least as old as PRM, before the other Timor languages underwent lenition of initial *b.⁸ (PTAP did not borrow AN loans for other domestic animals like ‘dog’ and ‘chicken’.)

(1) MP lexeme sets for ‘pig’

PMP *babuy

PFL *vavi	PRM *bafi	AN in East Timor	
		Dadu’a	<i>wawi</i>
		Galolen	<i>hahi</i>
		Waima’a	<i>wau</i>
		Tetun, Suai	<i>fahi</i>
		NW Mambae, Barzatete ⁹	<i>hɛh a</i> ¹⁰
		C Mambae, Hatu-Builico	<i>haih a</i>
		S Mambae, Hatu-Udo	<i>hae</i>
		Naueti	<i>wou</i>

(2) TAP lexeme sets for ‘pig’

PTAP *baj ‘pig’

TAP Pantar-Straits	Deing	<i>bai</i>
	Klamu	<i>bei</i>
	Sar	<i>bai</i>
	Teiwa, Adiabang	<i>baj</i>

8 Lenition was possibly quite late: note that Tetun has initial *b > f and Waima’a has *b > w. These two languages are quite closely related, and Edwards (p.c.) reconstructs **b for their immediate ancestor (Proto Eastern Timor).

9 Mambae, Kemak, Welaun, and Tokodede are placed in a Central Timor subgroup which is (currently) coordinate to Timor-Babar, see figure 3.4.

10 A vertical line ‘|’ separates the non-etymological parts of a word from its etymological part. Accolades ‘{...}’ separate a non-etymological part of a compound from the etymological part.

	Kaera	<i>bej</i>
	Reta, Pura	<i>be:</i>
	Reta, Ternate	<i>bei</i>
	Blagar	<i>be</i>
TAP West Alor	Adang, Otvai	<i>bɔi</i>
	Adang, Lawahing	<i>bi</i>
	Hamap	<i>bi</i>
	Kabola, Monbang	<i>bi?</i>
	Klon, Bring	<i>be:?</i>
TAP Central Alor	Abui, Takalelang	<i>fe</i>
	Papuna	<i>fe</i>
TAP South Alor	Kiraman	<i>bei</i>
	Kui	<i>bei</i>
TAP East Alor	Kamang, Atoitaa	<i>pej</i>
	Suboo	<i>pe</i>
	Tiyei	<i>pe</i>
	Wersing	<i>pei</i>
	Sawila	<i>pi</i>
	Kula	<i>peja</i>
TAP East Timor	Makasae	<i>bai</i>
	Fataluku	<i>pai</i>

2.1.2 'deer'

Deer are ancient animals in eastern Indonesia. They appeared in Timor after 4500 BP (Bellwood 1997: 187), in Sulawesi after 3500 BP (Glover 1986) and in Flores after 2000 BP (Forth 2012: 457). Blust and Trussell (n.d.) give a cognate set of Proto West Malayo-Polynesian *uRsah 'sambhur deer' containing words from Philippine languages, Malay and Toba Batak. Words related to this form are found in Flores-Lembata and Timor, but they may have been borrowed from Malay *rusa* (Edwards 2021). Malay has likely been a regional lingua franca at least since the time of the Sri Wijaya empire (7th–9th Century), and has been used as a trade language in eastern Indonesia since before the colonial times. Antonio Pigafetta's encounter in 1521 with traders from Malacca in Timor (Le Roux 1929: 31) and the Malay word list he collected in Tidore (North Moluccas) (Le Roux 1929: 72–99) is evidence that trade Malay was already used in the region in the early 16th Century.

In east Timor, the word *rusa* sometimes occurs in a compound with *bibi* (Proto Rote-Meto *bibi 'goat', Edwards 2021), or as synonym of *bibi*, see (3). This particular compound is also found in the TAP languages Bunak and Makasae, (4), which suggests that Bunak and Makasae picked it up from one of their neighbours; likely Tetun, the language of interethnic communication.

(3) MP lexeme sets for 'deer'

PWMP *uRsah 'deer'

FL #rusa	PRM #rusa	AN in East Timor	
		Dadu'a	<i>rusa</i>
		Waima'a	<i>ruso, bibi ruso</i>
		Tetun, Suai	<i>bibi rusa</i>
		Naueti	<i>bibi rusa</i>

(4) TAP East Timor lexeme sets for 'deer'

TAP East Timor	Bunak	<i>rusa, bibu</i>
	Makasae	<i>bibi rusa</i>

The Alor Pantar forms suggest a different history. On the basis of the lexeme set given in (5), we can reconstruct proto AP *arusa. Forms reflecting regular correspondences of the consonants are found in all major subgroups of AP: Tubbe r=l, Klamu s=tʃ, Reta r=l, s=h, Adang r=l, Abui r=j, s=t, Kaman r=l, l=zero, s=h.¹¹ On the one hand, this suggests that it is an ancient loan, though it is unclear what the donor language of PAP *arusa may have been. On the other hand, all groups also contain irregular forms, e.g. Teiwa *s>t (no change expected), Klon *s>t (expected *s>h), and Sawila and Wersing retained /s/ (expected *s>t). Irregular *s>t forms may suggest borrowing from a TAP language which underwent that change (e.g. Abui). Further confusing matters, the form could also have been a more recent loan from Malay *rusa*, as has been suggested for the forms attested in Flores-Lembata and Timor. There is no information about when deer appeared in Alor and Pantar.

(5) PAP lexeme set for 'deer'

PAP *arusa 'deer'

TAP Pantar-Straits	Tubbe	<i>lus</i>
	Klamu	<i>raʃfi</i>
	Sar	<i>ru:t</i>
	Teiwa, Lebang	<i>ru:s</i>
	Teiwa, Nule	<i>ru:t</i>

11 Here the symbol '=' is used to denote sound correspondences, not sound changes (which would be represented using '>'). This is done because in some cases it is not sure that the forms in the sets are actually cognates, and so, strictly speaking, we cannot say that a sound 'change' was involved, while a correspondence is obviously there. Some of the correspondences are regular, others are not, and for some correspondences we do not know whether they are regular or not.

	Kaera	<i>rusi</i>
	Reta, Pura	<i>aluha</i>
	Reta, Ternate	<i>aluha?</i>
	Blagar, Bama	<i>rusi</i>
	Blagar, Kulijahi	<i>ruhi, ruhiŋ</i>
	Blagar, Manatang	<i>ʔuruhiŋ</i>
	Blagar, Nule	<i>ruŋ</i>
	Blagar, Pura	<i>haruhiŋ</i>
	Blagar, Tuntuli	<i>rusi</i>
	Blagar, Warsalelang	<i>urusu</i>
TAP West Alor	Adang, Otvai	<i>aru</i>
	Adang, Lawahing	<i>a:lu</i>
	Klon, Hopter	<i>ʔə'rut</i>
TAP Central Alor	Abui, Takalelang	<i>ajut</i>
TAP South Alor	Kiraman	<i>arusu</i>
	Kui	<i>arus</i>
TAP East Alor	Kamang, Atoitaa	<i>au:h</i>
	Suboo	<i>o:h</i>
	Tiyei	<i>a:uh</i>
	Wersing, Maritaing	<i>arus pe</i>
	Sawila	<i>arusu pi</i>
	Kula, Lantoka	<i>aisua pe</i>

2.2 *Subsistence and Trade*

2.2.1 'salt'

Salt is a natural sea product used in barter trade between coastal and inland people in Timor and Alor (Hägerdal 2012: 68, Wellfelt 2016: 145). Across the TAP family, we find reflexes going back to PTAP *asir, a form related to PMP *qasiRa 'salt', compare (6) and (7). The form must be a rather ancient loan. Given the different shape of PFL *hira, this cannot be the donor for PTAP *asir. The languages of west Timor reconstruct to PRM *masi from PMP *ma-qasin 'salty' (Edwards 2019), though Helong in west Timor has *sila* 'salt', a reflex of *qasiRa. The borrowing event of PTAP *asir from an Austronesian source must thus have taken place at a stage preceding PFL *hira or PRM *masi. In east Timor, the Austronesian languages partly reflect *masi (< PMP *(ma-) qasin) (Dadu'a, Galolen, Tetun), and partly *asira (< PMP *qasiRa) with loss of the initial /a/ and the intervocalic /r/ (Tokodede, Kemak, Mambae). Waima'a, Midiki, Naueti either reflect PMP *(ma-)qasin, or they reflect *asira plus loss of the final syllable, as shown in (6). It is thus likely that a form *asiRa was present at the stage of Proto Timor-Babar, a subgroup which includes all AN languages on Timor except those of Central Timor (Welaun, Kemak, Tokodede, Mam-

bae); including west Timor Helong which has *sila*. The Proto Timor-Babar form *asiRa was borrowed as *asir into Proto Timor-Alor-Pantar.

(6) MP lexeme sets for 'salt'

PMP *qasiRa 'salt', PMP *ma-qasin 'salty'

PFL *hira	PRM *masi	AN in East Timor	
		Dadu'a	<i>masi</i>
		Galolen	<i>masin</i>
		Idate	<i>masi</i>
		Tetun Dili	<i>masin</i>
		Tetun, Suai	<i>masin</i>
		Waima'a	<i>asi</i>
		Tokodede	<i>sia</i>
		Kemak	<i>sia</i>
		Mambae	<i>sia</i>
		Midiki	<i>asi</i>
		Naueti	<i>asi</i>

In the Alor-Pantar subgroup, the Pantar-Straits languages show reflexes with metathesized vowels (*asir>isar), see (7). The cognates show regular sound correspondences in Adang, Kafoa and Klon (s=h), Abui (s=t), Kui, Kiraman, Kula and Wersing (s=s), Kamang s=s, r=i, Adang r=i.

(7) TAP lexeme sets for 'salt'

PTAP *asir 'salt'

TAP Pantar-Straits	Tubbe	<i>his:i</i>
	Klamu	<i>je:si</i>
	Sar	<i>hisar</i>
	Teiwa, Lebang	<i>hisar</i>
	Teiwa, Nule	<i>jisar</i>
	Kaera	<i>isar</i>
	Reta, Pura	<i>?ihal</i>
	Reta, Ternate	<i>ihal</i>
	Blagar, Bama	<i>isar</i>
	Blagar, Kulijahi	<i>sija</i>
	Blagar, Manatang	<i>sia</i>
	Blagar, Nule	<i>siah</i>
	Blagar, Pura	<i>sia</i>
	Blagar, Tuntuli	<i>isar</i>
	Blagar, Warsalelang	<i>isar</i>

TAP West Alor	Adang, Otvai	<i>ahēi</i>
	Adang, Lawahing	{ <i>taŋ</i> } <i>hiri</i>
	Kafoa	<i>ahel</i>
	Klon, Hopter	? <i>hir</i>
	Klon, Bring	<i>əhir</i>
TAP Central Alor	Abui, Takalelang	<i>ati</i>
	Papuna	<i>asi</i>
TAP South Alor	Kiraman	<i>ser</i>
	Kui	<i>ser</i>
TAP East Alor	Kamang, Atoitaa	<i>asi:</i>
	Suboo	<i>asi:</i>
	Tiyei	<i>asi:</i>
	Wersing, Maritaing	<i>asir</i>
	Sawila	<i>asira</i>
	Kula, Lantoka	<i>asi</i>
TAP East Timor	Fataluku	<i>asir</i>
	Oirata	<i>asir</i>
	Makasae	<i>gasi</i>

3 Pre-modern Loans

3.1 *Textile Technology*

3.1.1 'needle'

PMP **zaRum* 'needle' is reflected in languages of east Timor as given in (8). It has monosyllabic reflexes where the intervocalic /r/ has been lost in Tetun, Kemak, and Nauti. This form was borrowed into Bunak, see (9). The form without the intervocalic /r/ is also the one attested in the east Alor languages Kula, Sawila and Wersing. Besides east Alor, the loan is not attested elsewhere on Alor or Pantar.

In Dadu'a, Galolen and Waima'a we find reflexes of #*ruma*, a form that may be connected irregularly to **zaRum*. This form is also found in the lexemes in the east Timor TAP languages Makasae and Fataluku, which contain etymons related to both **daun* and #*ruma*.

(8) MP lexeme set for 'needle'

PMP **zaRum* 'needle'

PFL— PRM— AN in East Timor

Tetun

daun (Morris 1984:23)

Kemak

daum

Naueti	<i>dau</i>
Dadu'a	<i>la luma</i>
Galolen	<i>ruma</i>
Waima'a	<i>rumo</i>

(9) TAP lexeme set for 'needle'

TAP Timor	Bunak, Bobonaro	<i>daun</i>
	Bunak, Suai	<i>daun</i>
	Makasae	<i>dauruma</i>
	Fataluku	<i>t̄aruma</i>
TAP East Alor	Kula	<i>dam</i>
	Sawila	<i>da:mu</i>
	Wersing	<i>damu, damu?</i>

In the western part of Alor and Pantar, 'needle' is often expressed with a reflex of PMP *batuR 'weave', showing a semantic shift, compare (10)–(11). The form is likely borrowed from Kedang *batur* into Marica Alorese,¹² and from Alorese into the neighbouring Pantar-Straits and West Alor languages. The source language cannot have been Lamaholot or another Flores-Lembata language like Hewa, as these languages use a different form *lusir/luhi(r)* 'needle'.

(10) MP lexeme set for 'weave'

PMP *batuR 'weave'	AN in Flores-Lembata	
	Kedang	<i>batur</i>
	Alorese, Marica	<i>batur</i>
	Alorese, various dialects	<i>batul</i>
	Alorese, Alor Besar	<i>batu</i>

(11) TAP lexeme set for 'needle' reflecting MP 'weave'

TAP Pantar-Straits	Teiwa, Adiabang	<i>bital</i>
	Teiwa, Lebang	<i>bati</i>
	Teiwa, Nule	<i>bitaj</i>
	Sar	<i>bitai</i>
	Klamu	<i>batu</i>
	Kaera	<i>ba:ti</i>
	Blagar	<i>batul</i>

¹² The final /r/ in Marica *batur* is irregular (inherited words would have lost the final /r/). Marica island is also located closest to the Kedang speaking area of northeast Lembata. Other Alorese dialect change final r>l and some lose it altogether (Fricke, p.c. 2020).

TAP West Alor	Adang, Lawahing	<i>batuŋ</i>
	Adang, Otvai	<i>batij</i>
	Kabola, Monbang	<i>batarŋ</i>
	Klon, Hopter	<i>bah</i>

3.1.2 ‘sew’

In the languages of the Flores-Lembata region, various etymons are used to express ‘sew’, leading to two PFL reconstructions **daru* (< PMP **zaRum* ‘needle’) and **daʔit* (< PMP **zaqit* ‘sew’). In addition, we find reflexes of the form ***sauR* ‘sew’ (Edwards 2021) in Lamaholot and Kedang, see (12). Reflexes of the regional form ***sauR* ‘sew’ are also found in the AN languages of Timor, as shown in (13). The TAP languages of east Alor are likely to have borrowed from (a) language(s) of the Central Timor subgroup, Kemak, Tokodede, or Mambae.

(12) Etymons to express ‘sew’

PMP * <i>zaRum</i> ‘needle’	PMP * <i>zaqit</i> ‘sew’	Regional form (pre-Rote Meto, Edwards 2021) * <i>sauR</i> ‘sew’ ¹³
PFL * <i>daru</i> ‘sew’	PFL * <i>daʔit</i> ‘sew’	Lamaholot, Kedang # <i>saur</i> ‘sew’

(13) Reflexes of regional **sauR* ‘sew’ in AN languages of east Timor

AN Timor	Galolen	<i>sor</i>
	Kemak	<i>sɔra</i>
	Tokodede	<i>sɔr</i>
	Waima’a	<i>sau</i>
	Southern Mambae, Ainaro	<i>sɔr</i>

(14) Reflexes of regional **sauR* ‘sew’ in TAP languages

TAP East Alor	Kula	<i>sua</i>
	Sawila	<i>surə</i>
	Wersing, Maritaing	<i>sɔr</i> { <i>burkinj</i> }
	Wersing, Taramana	<i>sor</i> ‘to sew’, <i>suai</i> ‘to stick’

The words for ‘sew’ in the other TAP languages listed in (15) seem to be related to the regional form ***sədu*(t) ‘weave’, reflected in Tetun *sɔru* ‘weave’ (see (16) below) and Central Lembata *surit* ‘weaving sword’ (Fricke 2017: 88); as well as

13 Edwards (2021:244): “Blust and Trussel (n.d.) reconstruct PCMP **sora*, including Meto as one of their attestations. The cognates in Timor and Flores appear to be better explained by **sauR*, with no final vowel and *R [r] instead of *r [r].”

in PRM *seru ‘weaving sword’ (Edwards 2021). The form that was borrowed into the TAP languages had medial *d changed to /r/.¹⁴

(15) TAP lexeme set meaning ‘sew’, reflecting regional **sədu(t) ‘weave’

TAP Pantar-Straits	Teiwa, Adiabang	<i>rot</i>
	Kaera	<i>səroto</i>
	Reta, Pura	<i>haruata</i>
	Reta, Ternate	<i>arwat:a</i>
	Blagar, Bama	<i>torosi</i>
	Blagar, Kulijahi	<i>rota</i>
	Blagar, Manatang	<i>harota</i>
	Blagar, Nule	<i>rota?</i>
	Blagar, Pura	<i>harota</i>
	Blagar, Tuntuli	<i>torosi</i>
	Blagar, Warsalelang	<i>sorota</i>
TAP West Alor	Adang, Lawahing	<i>naroto?</i>
	Adang, Otvai	<i>harət</i>
	Hamap, Moru	<i>na harot</i>
	Kabola, Monbang	<i>na saroto</i>
	Kafoa	<i>hiota</i>
TAP South Alor	Klon, Bring	<i>{il} hərət</i>
	Klon, Hopter	<i>{il} hərət</i>
	Kui, Labaing	<i>serot</i>
TAP Central Alor	Kiraman	<i>surot</i>
	Papuna	<i>sərowat r</i>
	Abui, Ulaga	<i>tiro:t</i>
	Suboo	<i>suri</i>
	Tiyei	<i>səot</i>

3.1.3 ‘weave’

The forms for ‘weave’ in the Timor AN languages Tetun and Waima’a are reflexes of the regional protoform **sədu(t) ‘weave’ (Edwards 2021, see ‘sew’ above). Similar forms are attested in the Timor TAP languages Bunak and Makasae, see (16).

(16) Forms for ‘weave’ reflecting **sədu(t) ‘weave’ in AN and TAP languages of east Timor

¹⁴ Hawu *pehədu* points to earlier medial *d, not *r (Edwards p.c.).

pre-RM **sədu(t) ‘weave’		
AN East Timor	Tetun, Suai	sɔru
	Waima’a	seru
TAP Timor	Bunak	selu
	Makasae	seru

The regional protoform **sədu(t) ‘weave’ also has reflexes in the TAP languages, but there the forms mean ‘sew’, see (15). In the TAP languages of Central and East Alor, the concept ‘weave’ is expressed by reflexes of borrowed PMP *tənun, *tinun ‘weave’, compare (17)–(18). (The t>s change is unexplained.) The forms could have originated from one or more AN language of Timor, compare Proto Rote-Meto *tenu. However, a direct source in east Timor cannot be established because, as mentioned, the modern AN languages of east Timor do not use reflexes of *tənun/*tinun ‘weave’, but forms of *sauR ‘sew’ instead to denote ‘weave’. It is also possible that the forms of the Central and East Alor languages are (adapted) loans from Malay or Indonesian *tenun*.

(17) MP reconstructions for ‘weave’

PMP *tənun, *tinun ‘weave’		
PFL PFL *tani	PRM *tenu	AN in East Timor

–

(18) TAP lexeme set for ‘weave’

TAP Central Alor	Abui, Takalelang	<i>tinei</i> ¹⁵
	Suboo	<i>sine</i> :
	Tiyei	<i>sine</i> :
TAP East Alor	Kamang, Atoitaa	<i>sine</i>
	Kula, Lantoka	<i>sinə</i> { <i>na</i> }

In Alor there are a few weaving communities along the coasts, but it is not known when the weaving technology was introduced. Oral traditions in Alor mention migrating groups who settled on the south coast as people bringing pottery (Wellfelt 2016, 63), and the same groups tend to be associated with weaving. Pottery and textiles were bartered with people in the interior, where there is a taboo on weaving.¹⁶

15 Abui *tinei* ‘weave (cloth)’ was likely the source for the internal derivation Abui *tij* ‘needle’.

16 A similar taboo on weaving is found in some inland areas of Lembata island (Fricke 2019).

Weaving cloth has been considered as a typical Austronesian cultural feature (Blust 2013:24), but there is some evidence that the weaving tradition in Timor was introduced or disseminated only several hundred years ago (Hägerdal 2012). Pigafetta (1522) reported about a visit to Timor: “The chief with whom I went to speak only had women to serve him. [The women] all go naked, just like the other [women on the other islands]. In their ears they wear small golden earrings with hanging brushes at the side. On their arms they wear many bangles of gold and yellow copper until the elbow. The men go about like the women, apart from that they hang certain golden objects, round like a plate, around their necks, and that they wear bamboo combs in their hair, adorned with golden rings. Some of them wear dried pumpkin stems in their ears instead of golden rings.” (Le Roux 1929). Hägerdal (2012, 18) comments: “The alleged nudity of the women (and, apparently, the men) is more puzzling when regarding the long sarongs worn more recently, but it corroborates a Franciscan travel account from 1670. It is therefore possible that the well-known weaving traditions of Timor were introduced or disseminated at a fairly late stage.” Dutch illustrations of the seventeenth century show Timorese men wearing a kind of loincloth made of straps (Hägerdal 2012, 18). In southeast Alor and other places in Alor bark cloth was widely used for garments until the mid-20th century (Wellfelt 2016, 97).

Today, the (few) weaving centres in Alor produce textiles decorated with techniques that in Indonesian are summarised as *songket*. The textile traditions from the south and east coast of Alor show affinities with Timor, which is congruent with other historical sources, both oral and written, and with the borrowing of *t/sine* ‘weave’.

In West Alor, coastal groups produce textiles with clear affinities to the Solor islands, and with inspiration from Indian textiles called *patola*, produced in Gujarat in North West India from the 11th century onwards (Wellfelt 2016, 63). In the TAP languages of West Alor, Straits, and Pantar, no forms related to PMP **tenun* are attested for ‘weave’; they use lexemes that are reconstructable to **degi* ‘weave’, the source of which (MP or not) is yet unclear.

3.2 Societal Structures

3.2.1 ‘slave’

In the AN languages of Timor, reflexes of PMP **qaRta* ‘outsider(s), alien person(s)’ are found to mean ‘slave’, see (19). Edwards (2021) referring to Mahdi (1994:464 ff.) suggests as the meaning of **qaRta* ‘negrito, black person’. This is based on the semantics across a wide range of MP languages which points to the original meaning being ‘black/Negrito person’ which, depending on the race of the speakers, was applied either to themselves or a subjugated population. In

many languages of Sulawesi and Maluku reflexes of this etymon have the meaning ‘slave’. Of the TAP family, Bunak and Makalero borrowed a reflex of *qaRta ‘slave’, see (20), and both would be unproblematical borrowings from Tetun. In the Flores-Lembata subgroup, PMP *qaRta is reflected as PFL *ata ‘person’ (not ‘slave’). In the TAP languages not spoken on Timor, different etymons are used, see the forms in (21) and (22), further discussed below.

(19) MP lexeme sets for ‘slave’

PMP *qaRta ‘outsider(s), alien person(s)’

PFL *ata ‘person’ PRM *ata ‘slave’ AN in East Timor

Dadu’a	<i>ata</i>
Galolen	<i>ata</i>
Tokodede	<i>a:t</i>
Tetun Dili	<i>ata n</i>
Waima’a	<i>ata</i>
Kemak	<i>ata r</i>
Kemak, Lemia	<i>ata</i>
Idate	<i>w ato</i>
S Mambae, Ainaro	<i>ata</i>
NW Mambae	<i>ata n</i>
C Mambae	<i>ata n</i>
Naueti	<i>ata</i>

(20) TAP lexeme set for ‘slave’

TAP East Timor	Bunak, Bobonaro	<i>ata n</i>
	Bunak, Suai	<i>ata n</i>
	Makalero	<i>ata n</i> ‘herder’ (Huber 2011: 542)

The pre-colonial political economy of Southeast Asia already included slave-raiding. Much of Southeast Asia was underpopulated until the 18th and 19th Centuries, and the key to political control was the control of labour power (Hoskins 1996, 3–4). The Makassarese from South Sulawesi played an important role in the pre-colonial and colonial slave trade, obtaining slaves from Alor, Manggarai and Ende in Flores, Timor, Tanimbar, Buton (Sulawesi), Mindanao (Philippines) and Brunei (Borneo) (Raben 2008, 132; Wellfelt 2016, 45). Most forms of slavery in Southeast Asia seem to have originated in debt bondage, but gradually diversified into complex “closed” systems of enduring social stratification and “open” ones of slaves captured primarily for external trade. As Hoskins (1996:4) writes: “Slaves were one of the most important “local products” exchanged from the hinterland for sale in *entrepôts* along the coasts,

and they were usually obtained by raiding inland communities.” In Timor as well as elsewhere in eastern Indonesia, slaves were an important trade commodity for the colonial Portuguese and Dutch VOC, alongside sandalwood and beeswax (Hägerdal 2012). Slave-raiders not only came from Sulawesi but also from the east. An unpublished grammar sketch of Iha, a Papuan language spoken on the Bomberai peninsula in Southwest Papua (Coenen 1953), mentions that in pre-contact times the Iha speakers went on slave expeditions all the way to the Kei and Tanimbar islands. In turn, there is a tradition in Fataluku (East Timor) that they came from the Kei islands (Voorhoeve 1989). This suggests that maritime contacts existed between the two ends of the chain Papua-East Timor, and a point in between, Kei; and that people movements took place along that chain.

On Alor, oral histories report about inland people such as the Abui being abducted and traded as slaves by coastal populations (Wellfelt 2016, 298, 300). An example are the Kolana (Sawila speakers) on the east coast of Alor. Kolana was allied with Liquiçá on the north coast of East Timor, with whom they traded wax, honey, cattle, and slaves, the latter acquired in wars or by kidnapping (Wellfelt 2016, 100). In 1851 van Lynden mentions Alor and Pantar as a former source of slaves to foreign traders, and the Oecusse enclave in north Timor is mentioned as a recipient of slaves from the Alor and Pantar: ‘In former days, Alor and Pantar provided many slaves and even now there are sometimes slaves being supplied to foreign traders, and to the Timorese (Oekoessie [Oecusse]) who are subject to Portugal [...]’ (Van Lynden 1851:332). According to a Dutch report from 1879, slaves from Alor were sold in Liquiçá via the regent in Lamahala on Adonara island—and the Portuguese commander received a head tax for each imported slave. The year after, in 1880, another report was highly critical of the rulers in Kui on the south coast of Alor and Kolana on the east coast. Both were accused of having brought mountain people from Alor to be sold as slaves in Liquiçá (Wellfelt 2016, 103).

Tetun *malae* refers to foreigners or traders who came from overseas. Reflexes of this word denote ‘slave’ in TAP languages of the Pantar Straits and West Alor, as well as in South and East Alor, and Bunak Maliana, see (21). The use a word similar to *Malay* to refer to a slave would suggest that slaves were associated with people who do not (originally) belong to one’s group.¹⁷ The centuries of slave trade from Alor Pantar to Timor, also involving the Solor islands, may have

17 In (Austronesian) Kemak Kutubaba the Indonesian/Malay word *matroos* ‘sailor’ (originally from Dutch *matroos* ‘sailor’) is used to denote ‘slave’. Just like the case of *malai*, the same word is used here to refer to both a non-indigenous person and a slave.

caused the borrowing of a form similar to the Tetun word *malae* ‘slave’ into languages across Alor and Pantar.

- (21) TAP lexeme sets for ‘slave’ reflecting Tetun ‘foreigner(s), trader(s) from overseas’

Tetun *malae* ‘foreigner(s) or trader(s) from overseas’

TAP Pantar-Straits	Reta, Pura	<i>mala:l</i>
TAP West Alor	Kafoa	<i>madal</i>
	Klon, Bring	<i>məlei</i>
TAP Central Alor	Papuna	<i>maja:</i>
TAP South Alor	Kui, Labaing	<i>marā</i>
TAP East Alor	Kamang	<i>ma:i</i>
	Sawila	<i>malē</i>
	Wersing	<i>məlai</i>
TAP East Timor	Bunak, Maliana	<i>milah</i>

In the Pantar-Straits area, Kaera and Blagar-Tuntuli borrowed the Indonesian/Malay form *jongos* [dʒoŋos] ‘houseboy’ for the notion ‘slave’, as shown in (22). Originally, the word is from Dutch *jongen(s)* [joŋən(s)] ‘boy(s), houseboy(s)’. In Kaera, either the original Dutch form with initial [j] was borrowed (which seems unlikely, because there was no Dutch-speaking population on Pantar), or the initial affricate of the Malay form was simplified to [j] in Kaera because Kaera lacks a phonemic affricate /dʒ/ (Klamer 2014).

- (22) TAP lexeme set for ‘slave’ reflecting Malay/Indonesian [dʒ] *ongos* ‘houseboy’

Malay/Indonesian *jongos* ‘houseboy’

TAP Pantar-Straits	Kaera	<i>joŋos</i>
	Blagar Tuntuli	<i>dʒoŋos</i>

The question may arise why at least three different etymons were borrowed for the same notion. Obviously, part of the answer lies in the different contact histories of the various regions, as the regional differences discussed above indicate. An additional explanation might be that, for many of the word lists used in this paper, the word for ‘slave’ was elicited using the Indonesian prompt *budak*. In Indonesian, this word has various meanings including ‘lad, boy’, ‘servant, underling’, and ‘serf, slave’, and thus it appears to have elicited words of a similar semantic range in the target languages.

In western interpretations, the notion of ‘slave’ means a person who is the servant-property of another person, and who can be bought and sold as such.

In the regions where we did our surveys in Flores, Pantar and Alor, the translations of Indonesian *budak* include this meaning but may also refer to people who are temporary servants ('debt slaves'), or to people who are not, or no longer, part of a particular clan lineage; for instance, orphans, newcomers or strangers. An example of this latter type is reported in Wellfelt (2016, 46). In an Adang village (West Alor) a story tells of a young man from Welai (Abui territory in Central Alor) who was taken from his parents and sold by relatives to traders from Binongko, Sulawesi. The traders ran into a storm and were forced to seek shelter in West Alor. The boy was set free and ended up with an Adang-speaking community in the mountains where he became founder of a new lineage. The abduction and sale of the boy is said to have happened 13 generations (i.e., 300–400 years?, MK) ago. Orphans and newcomers can start their own lineage in a clan, but unless they are adopted into an existing lineage, their lineage will retain a different (often lower) status. For example, they will not be allowed to take part in the ritual negotiations relating to marriage exchanges, but will have practical duties in support of these negotiations, such as organising the food. People in such non-autochthonous lineages may in some ways be considered as servants to the community, but they are not 'owned' by an individual or by a particular autochthonous lineage.

Budak can also be used to refer to war prisoners that are incorporated into the group who captured them, e.g. to become their wives; or prisoners who are given away to another group as part of a peace treaty. In their new environment, such 'slaves' do not necessarily get a lower societal position, nor are they necessarily seen as servants. In fact, they can become normal members of their new group. For example, a captured woman can be treated like all the other women who marry into the clan, and captured children may be adopted by childless couples who bring them up as their own children.

3.2.2 'king, ruler'

The Tetun compound *liu rai* 'king, executive ruler' (lit. 'surpassing (the) earth/estate', cf. Hägerdal 2009, 49), commonly written as *liurai*, has been borrowed into a number of TAP languages on Timor, as well as in languages in South and East Alor that were in contact with Timor (cf. Wellfelt 2016), see (23).

(23) TAP lexeme set for 'king, ruler' reflecting Tetun *liurai*

Tetun *liurai* 'king, executive ruler'

TAP South Alor Klon, Hopter *le:r*

TAP East Alor Kamang *le:i*

 Kula, Lantoka *le:r*

	Sawila	<i>liri</i>
	Wersing	<i>lɛri</i>
TAP East Timor	Bunak	<i>liurai</i>
	Makasae	<i>dai</i>

In TAP languages of Pantar, the concept ‘king’ is expressed with forms related to Malay/Indonesian *rad̥ʒa* ‘king’, possibly through the form that was borrowed into Adonara Lamaholot, as shown in (24), or else directly borrowed from Malay/Indonesian. In both Adonara and the TAP languages of Pantar, the affricate in *rad̥ʒa* has been simplified to [j], because none of these languages have a phonemic affricate [d̥ʒ]. The borrowing may be pre-modern or modern, but cannot be very recent, as currently, the *d̥ʒ* in Indonesian/Malay loans occurring in any of the these languages is not simplified to [j].

(24) MP and TAP lexeme set for ‘king’

Mly/Ind	<i>rad̥ʒa</i>	‘king’
MP Flores-Lembata	Adonara Lamaholot	<i>raja</i> ‘king’
TAP Pantar	Tubbe	<i>raja</i> ‘king’
	Sar	<i>raja</i> ‘king’
	Teiwa, Lebang	<i>raj</i> ‘king’
	Kaera	<i>rai</i> ‘king’

3.3 *Body Parts*

3.3.1 ‘breast’

PMP **susu* is reflected in PFL *(t)usu and PRM **susu*. Reflexes of **susu* are also attested in the AN languages in the north of eastern Timor, see (25). Reflexes of a form with initial /s/ were borrowed into the TAP languages of Timor, see (26) (but Fataluku shows a reflex of PTAP **hami* ‘breast’).

(25) MP lexeme set for ‘breast’

PMP	* <i>susu</i>	
PFL	*(t)usu	PRM * <i>susu</i>
		AN in East Timor
	Dadu’a	<i>susu</i>
	Galolen	<i>susu n</i>
	Tokodede	<i>susu</i>
	Tetun Dili	<i>susu n</i>
	Waima’a	<i>susu {wai}</i>
	Kemak	<i>susu r</i>
	Idate	<i>susu</i>
	S Mambae, Ainaro	<i>susu</i>
	Naueti	<i>susu</i>

(26) Lexeme set for 'breast' in TAP languages of East Timor

TAP East Timor	Bunak, Bobonaro	<i>su:</i>
	Bunak, Maliana	<i>su:</i>
	Bunak, Suai	<i>su:</i>
	Makasae	<i>dudu</i>
	Oirata	<i>susu</i>

The TAP languages Kui and Kiraman on the south coast of Alor borrowed a form with an initial fricative *su*, (27). The source of this borrowing event is also likely to be (the ancestor of) an AN language spoken on the northern Timor coast, as these all have forms starting with /s/. In the TAP languages of Pantar-Straits and West Alor a form with an initial plosive was borrowed, similar to PFL *(t)usu (but dropping the final vowel). The donor is likely to have been Alorese *tuho*. The other languages of Alor Pantar, including those in East Alor and Fataluku on Timor, show reflexes of PTAP *hami 'breast'.

(27) Lexeme set for 'breast' in TAP languages of Alor Pantar

TAP Pantar-Straits	Kaera	<i>tu:</i>
	Blagar, Bama	<i>tu:</i>
	Blagar, Tuntuli	<i>-tu</i> ¹⁸
TAP West Alor	Adang, Lawahing	<i>to?</i>
	Adang, Otvai	<i>to</i>
	Kabola, Monbang	<i>oto?</i>
	Kafoa	<i>tot</i>
	Klon	<i>do:t</i>
TAP Central Alor	Abui, Ulaga	<i>-tuti</i>
TAP South Alor	Kiraman	<i>-su</i>
	Kui, Labaing	<i>-su</i>

3.3.2 'navel'

PMP *pusej is reflected in PFL *pusər (with an irregular final /r/). There are no reflexes of forms with a final glide attested in any of the TAP languages. Forms with an initial plosive /p/ and (reflexes of) a final liquid are found in Blagar and Reta in the Straits, and in Adang, and Kafoa in West Alor, compare (28)–(30). Forms without a medial /s/ in Blagar, Reta and Adang could be loans from an Alorese variety spoken on neighbouring Pantar island, as these varieties have *puhər* 'navel' (while most other Alorese varieties have forms with

18 The bound forms take obligatory inalienable possessor prefixes.

a prefix, as for example Alorese Marica), see (29). The forms with a medial /s/ point to a different source; they may have been borrowed directly from Malay.

The languages of Alor that are spoken further east have a different source. The form *kubu* in Wersing has an unexplained initial syllable that reflects the initial syllable of Tokodede *kupusa*, Alorese Marica *kəpuhɔr*, and is also found in other Western and Central Lamaholot languages (Fricke 2019). This suggests that there once was an older regional form with a prefix, which has modern reflexes in Timor as well as Flores-Lembata. The form in Wersing in particular probably originates from Tokodede on the north coast of Timor, given the geographical proximity and the contacts we know existed between groups in East Alor and North Timor (Schapper & Klamer 2017; Schapper & Wellfelt 2018). A shortened reflex *-bu(:)* is found in the sister languages of Wersing, Kamang and Tiyei.

(28) MP lexeme set for ‘navel’

PMP *pusej (Blust and Trussell n.d.); Malay *pusar* ‘navel’

PFL *pusəɾ PRM *husə AN in East Timor

Tetun Dili	<i>husar</i>
Tetun Suai	<i>husar</i>
Tokodede	<i>ku pusa</i>
Kemak	<i>pusrar</i>
Waimaha	<i>huso</i>
Idate	<i>usar</i>

(29) Lexeme set for ‘navel’ in Alorese varieties

PFL *pusəɾ PRM *husə	Alorese, Helandohi	<i>puhɔr</i>
	Alorese, Wailawar	<i>puhɔr</i>
	Alorese, Munaseli	<i>puhɔr</i>
	Alorese, Pandai	<i>puhɔr</i>
	Alorese, Marica	<i>kə puhɔr</i>

(30) TAP lexeme set for ‘navel’

TAP Pantar-Straits	Blagar, Warsalelang	<i>-pusal</i>
	Blagar, Bama	<i>-pusal</i>
	Blagar, Tuntuli	<i>-pusal</i>
	Blagar, Kulijahi	<i>puar</i>
	Blagar, Nule	<i>puar</i>
	Reta, Pura	<i>puhal</i>
	Reta, Ternate	<i>-pual</i>

TAP West Alor	Adang, Lawahing	<i>ʔa pojʔeŋ</i>
	Adang, Otvai	<i>ʔa puhei</i>
	Hamap, Moru	<i>-puhe</i>
	Kabola, Monbang	<i>-pusu</i>
	Kafoa	<i>-puhai</i>
	Klon, Bring	<i>-pohə gen</i>
	Klon, Hopter	<i>-puhi gen</i>
TAP East Alor	Kamang, Atoitaa	<i>-bu</i>
	Tiyei	<i>-bu:</i>
	Wersing	<i>ku bu</i>

The wide-spread borrowing of an MP word for ‘navel’ across the Alor languages is probably due to its socio-political connotation of ‘centre, head quarters’. The variable patterns of the loans indicate at least three different donor languages: Alorese, Tokodede, and Malay, where borrowing from Malay is likely to have involved separate borrowing events across the island.

3.4 *Subsistence and Trade*

To this semantic domain belong ‘seed’, and ‘maize’, discussed below, but also ‘salt’ (section 2.2.1) and ‘slave’ (section 3.2.1). Including the concept ‘skin’ in this domain is motivated in the relevant section below.

3.4.1 ‘seed’

PMP *binəhiq ‘seed’ is reflected with forms like *fini*, *hini* or *wine* in the AN languages of Timor, see (31). In the TAP languages of Timor, only Bunak-Maliana has a form reflecting the original initial /b/, (33), so Bunak must have borrowed the word either before the sound change *b > w, f, h took place in Timor, or it borrowed the word from an unknown AN source that retained the original /b/. The original /b/ is also found in the loans of Abui, Kamang, Suboo and Tiyee, spoken in Central and East Alor, (33). This might suggest that the forms were borrowed from Bunak, but contact between Central and East Alor and the innerland Bunak seems unlikely. In the west Timor languages, the bilabial stop is also retained, cf. PRM *bini (Edwards 2021) so a predecessor of one of these west Timor languages could also have been the donor of the loans into the Alor languages. Alternatively, the borrowing into TAP may have occurred from an east Timor language before the initial /b/ of PMP *binəhiq started to vary in Timor. In the Flores-Lembata languages no reflexes of *binəhiq are found except in Sika (spoken in the Central Flores region), shown in (32). (Most of the Flores-Lembata languages use a form #kuluk (Fricke 2019), a form that has been borrowed as *kulu* (probably through Alorese) into the TAP languages Blagar Kulijahi and Blagar Nule. This form is not further discussed here.)

(31) MP lexeme set for ‘seed (rice)’

PMP *binəhiq ‘seed rice, rice set aside for the next planting’

LH-KD #kuluk	PRM *bini	AN in East Timor	
		Galolen	<i>hini</i>
		Tetun Dili	<i>fini</i>
		Tokodede	<i>hi:ni</i>
		Kemak	<i>hini</i>
		Waimaha	<i>wine</i>
		Idate	<i>hini</i>
		W Mambae, Barzatete	<i>hina</i>
		NW Mambae, Hatulia	<i>fini</i>
		S Mambae, Hatu-Udo	<i>hiin</i>
		S Mambae, Ainaro	{ <i>na:m</i> } <i>hiin</i>

(32) Reflexes of MP *binəhiq ‘seed’ in Flores-Lembata

AN Flores-Lembata	Sika-Hewa	<i>ihin</i>
	Sika Tana Ai	<i>βini</i>

(33) TAP lexeme sets for ‘seed’

TAP Central Alor	Abui	<i>bi: ka</i>
TAP East Alor	Kamang, Atoitaa	<i>bile:, bini</i>
	Suboo	<i>bile</i>
	Tiyei	<i>bili:</i>
TAP East Timor	Bunak Maliana	<i>bin</i>

3.4.2 ‘maize’

Maize originates from South America and was taken to eastern Indonesia through the Iberian colonial trade network. Maize was first introduced in the Timor region in the period 1540–1650 (Hägerdal 2012:16). In the region under study, lexemes similar to PMP *batad (but with a final /r/) generally mean ‘maize’, as in the AN languages of Flores-Lembata (PFL *vatar ‘maize’) and the AN East Timor forms listed in (34).

PMP *batad ‘millet or sorghum sp. (unident.)’ is listed in Blust and Trussel (n.d.) on the limited evidence of three related forms from the Philippines, to which we can add Bugis *bata?* ‘sorghum’. It is unclear how old sorghum is in Southeast Asia. Lexical and ritual evidence presented in Fox (1991) indicates that it preceded maize as subsistence crop in eastern Indonesia. Makassar has *batara?* ‘millet’ (Cense 1979), and since the Makassarese were involved in inter-regional trade including eastern Indonesia since before the colonial times (see the discussion of ‘slave’ in section 3.2.1), Makassar *batara?* could be the source

of a regional form *batar*, which assimilated millet and/or sorghum and maize (Fox 1991).

A Dominican source mentions maize on Lembata and Pantar shortly after 1641 (Hägerdal 2010, 224). Maize was grown in westernmost Timor by 1658, but must have been known some time before, since by then it was already the main crop (Hägerdal 2012, 50). In contrast, in some parts of Alor, maize was only introduced in the 20th C (Wellfelt 2016, 101). While the food was introduced relatively recently in certain parts of the region, the word seems to have a long history in the AP subfamily, and it may originally have referred to an earlier crop like sorghum, as it did in Kambera on Sumba, where *wataru* means both ‘maize’ and ‘sorghum’ (Forth 1983: 62).

Reflexes of PMP *batad or regional #batar ‘maize’ are not found in the TAP languages of Timor. However, in Alor Pantar, the form is attested everywhere, see (35). This form is strikingly similar to the forms attested in the AN languages of east Timor, see (34), and it is likely to have been borrowed from there. The Flores-Lembata region is an unlikely region of origin, because of the initial fricative in PFL *vatar.

(34) MP lexeme sets for ‘sorghum species’, ‘millet species’, ‘maize’

PMP *batad ‘sorghum sp., *Andropogon sorghum*’

PMP *bətəŋ ‘millet species, probably foxtail millet, *Setaria italica*’

PFL *vatar ‘maize’ PRM *betə ‘millet’ AN in East Timor

Idate	<i>pata:r</i> ‘maize’
Tetun Dili	<i>batar</i> ‘maize’
Tetun, Suai	<i>batar</i> ‘maize’
Mambae	<i>batar</i> ‘maize’

The initial /b/ of *batar shows regular sound correspondences across the AP languages, e.g.: b>f in Abui, b>p in Kula, Sawila and Wersing, see (35). The final /r/ regularly got lost in Abui. If the word was introduced into the AP languages together with the introduction of the new staple food maize since the 17th C, this means that these sound changes must have occurred later than 400 years ago. Alternatively, the word may be an older loan that originally referred to ‘sorghum’ which assimilated the meaning of ‘maize’ after that crop was introduced, as it did in Timor and Sumba (Fox 1991).

(35) TAP lexeme set for ‘maize’

PAP *batar

TAP Pantar-Straits	Tubbe	<i>bat:e</i>
	Klamu	<i>bata</i>

	Sar	<i>batar</i>
	Teiwa, Lebang	<i>batar</i>
	Teiwa, Nule	<i>batar</i>
	Kaera	<i>batar</i>
	Reta, Pura	<i>batal</i>
	Blagar, Bama	<i>batar</i>
	Blagar, Kulijahi	<i>batar</i>
	Blagar, Manatang	<i>batar</i>
	Blagar, Nule	<i>batar</i>
	Blagar, Pura	<i>batar</i>
	Blagar, Tuntuli	<i>batar</i>
	Blagar, Warsalelang	<i>batar</i>
TAP West Alor	Adang, Otvai	<i>bate</i>
	Adang, Lawahing	<i>bati?</i>
	Klon, Hopter	<i>bat</i>
TAP Central Alor	Abui, Takalelang	<i>fat</i>
	Abui, Fuimelang	<i>fa:ti</i>
	Papuna	<i>ba:ti</i>
TAP South Alor	Kiraman	<i>bati</i>
	Kui	<i>batar</i>
TAP East Alor	Kamang, Atoitaa	<i>patei</i>
	Suboo	<i>pati:</i>
	Tiyei	<i>pati</i>
	Wersing, Maritaing	<i>peter</i>
	Sawila	<i>pata</i>
	Kula, Lantoka	<i>pte, p̄te</i>
TAP East Timor	–	

3.4.3 'skin'

PMP **kulit* 'skin; bark' is reflected in the AN languages of Flores-Lembata and Timor, see (36), as well as in modern Malay/Indonesian *kulit*. A reflex of this form is found throughout the TAP languages, where almost all lexemes reflect #*kuli*, with the final /t/ consonant lost, as shown in (37). Some but not all of the TAP lexemes show regular sound changes: PAP *l>i in Kaera *koi*, PAP *l>i and *k>ʔ in Adang *ʔui*. None of the TAP loans have more than two consonants, except Blagar Kulijahi *-ʔulit*, which could be a modern loan from Indonesian *kulit*, and Wersing *klut*, which is similar to Tokodede *kulut*.

- (36) MP lexeme set for ‘skin, bark’
 PMP **kulit* ‘skin; bark’
 PFL **kulit* PRM— AN in East Timor
- | | | |
|--|------------|----------------|
| | Dadu’a | <i>uli k</i> |
| | Midiki | <i>kuli ŋ</i> |
| | Kemak | <i>ulit ir</i> |
| | Tetun Dili | <i>kulit</i> |
| | Tokodede | <i>kulut a</i> |
| | Naueti | <i>kuli</i> |
- (37) TAP lexeme set for ‘skin, bark’
- | | | |
|--------------------|---------------------|-----------------|
| TAP Pantar-Straits | Tubbe | <i>kili</i> |
| | Kaera | <i>koi</i> |
| | Blagar, Warsalelang | <i>pi kol</i> |
| | Blagar, Bama | <i>pi kol</i> |
| | Blagar, Tuntuli | <i>qol</i> |
| | Blagar, Kulijahi | <i>pi ʔulit</i> |
| TAP West Alor | Adang, Lawahing | <i>ʔui</i> |
| | Adang, Otvai | <i>ʔuil</i> |
| | Hamap, Moru | <i>oil</i> |
| | Kabola, Monbang | <i>pi kul</i> |
| | Kafoa | <i>kɔ:l</i> |
| | Klon | <i>kvi</i> |
| TAP Central Alor | Abui | <i>te kul</i> |
| TAP South Alor | Kui, Labaing | <i>ta kuil</i> |
| | Kiraman | <i>kuli</i> |
| TAP East Alor | Kamang, Atoitaa | <i>na kul</i> |
| | Suboo | <i>ne kul</i> |
| | Tiyei | <i>kul</i> |
| | Wersing | <i>klut</i> |
| TAP East Timor | Makasae | <i>uli</i> |

In Alor, bark cloth was widely used for garments until the mid-20th century (Wellfelt 2016:63, 97), and the widely spread borrowing of the concept ‘skin’ could be related to this, because the skin of certain tree were stripped to make bark cloth. There is archaeological evidence that the introduction of bark cloth technology followed the spread of Neolithic culture from southern China into Island Southeast Asia where bark cloth was substituted for other kinds of fibre materials (cf. Wellfelt 2016, 97). This may suggest that the bark clothing was introduced with the Austronesian word for it.

3.5 *Marriage*

3.5.1 'bride price'

Loan forms denoting 'bride price' that are similar to PMP *bəli 'value, price, marriage prestations, brideprice, purchase' are found all over TAP, see (38) and (39). However, they do not show regular sound correspondences. For instance, we do not witness the expected regular correspondence between initial/medial PAP *b > f in Abui, and initial PAP *b > p in Kamang, see (39). Most TAP forms have lost the second syllable of PMP *bəli, with the exception of the disyllabic loans attested in the Pantar-Straits, and in Fataluku.

In the TAP languages, forms with initial *b are attested across the region. It is unclear where the loans originated from. If they came from Timor, the donor form must have had an initial stop. None of the modern AN languages of east Timor retained the initial stop, but PRM did have it, so borrowing could have happened at an earlier stage, before the initial consonant of PMP *bəli started to vary in the Timor region. In Flores-Lembata the initial stop of PMP *bəli was already changed into a fricative at the stage of PFL *veli, so if the donor was a language from the Flores-Lembata region, the borrowing occurred already before the stage of PFL.

Interestingly, in the region that is geographically closest to Flores-Lembata, the Pantar-Straits, no reflexes of PFL *veli are attested, but rather of *beli, see (39). The vowels and the syllable structure of these Pantar-Straits forms are different from the forms attested on Alor, and more similar to modern Malay/Indonesian *beli* 'buy' or *belis* 'bride price'. The word *belis* 'bride price' is generally used in the Malay/Indonesian variety spoken in the eastern province (NTT) of Indonesia (Jones, Hull & Mohamad 2011). It is quite common to hear speakers of local languages use the loanword *belis*, likely because marriages are also frequently arranged between communities with different languages. This may suggest that the forms in the Pantar-Straits represent a different (possibly more recent) borrowing event involving *belis*. In general, the irregular forms suggest that the borrowing of reflexes of PMP *bəli occurred multiple times and from different sources.

(38) MP sets for 'bride price'

PMP *bəli 'value, price, marriage prestations, brideprice, purchase'

PFL *veli¹⁹ PRM *beli AN in East Timor

East Tetun

foli|*n*²⁰

19 PFL *veli 'price; bride price; expensive; buy'.

20 East Tetun *folin* 'price, cost, value; objects for barter' (Morris 1984:35).

	Kemak	<i>heli r</i>
	NW Mambae, Hatulia	<i>heli n</i>
	Naueti	<i>weli</i>
(39) TAP lexeme sets for 'bride price'		
TAP Pantar-Straits	Reta, Pura	<i>bili {pala}</i>
	Reta, Ternate	<i>ta beli</i>
	Blagar, Bama	<i>wili {pala}</i>
	Blagar Manatang	<i>?e bili</i>
	Blagar, Nule	<i>e bali</i>
	Blagar, Tuntuli	<i>ge vili</i>
TAP West Alor	Adang, Otvai	<i>fali</i>
	Kabola, Monbang	<i>?o wol</i>
TAP Central Alor	Abui, Takalelang	<i>he bel</i>
TAP East Alor	Kamang, Atoitaa	<i>fa:l</i>
	Suboo	<i>bal</i>
	Tiyei	<i>bal</i>
TAP East Timor	Bunak	<i>bol</i>
	Fataluku	<i>wala {hana}</i>

The pairs in Reta Pura *bili pala* and Blagar Bama *wili pala* are probably borrowed from Alorese, which has the compound *feling palang* 'dowry paid by the groom's family to the bride's family'.²¹

4 Summary of the Findings

The Austronesian lexical influence on the TAP languages as reflected by the loans discussed above can be characterized as involving animals (pig, deer), textile technology (needle, to weave, to sew); societal structures (slave, king/ruler), body parts (breast, navel), subsistence and trade (salt, seed, maize, skin), and marriage (bride price). The widely spread MP word for the body part 'navel' probably relates to its socio-political connotation of 'centre, head quarters', while 'skin, bark' may have been a trade item as clothing in the region was often made from tree bark until the 19th C (Van Lijnden 1851: 332).

21 The second half of the compound *palang* does not appear to have an independent meaning in Alorese. Thanks to Yunus Sulistyono for checking this with native speakers in Alor and Pantar in July 2020.

The MP loans discussed above differ in their donor region; an overview is given in (40). For some loans the donor region cannot be established, (40a), or the loan may have various different regions of origin, (40b). The loan may also be either from Timor or from Flores-Lembata, or from both (40c). Timor is the region where most of the AN loans investigated in this paper come from (40d). Certain loans from Timor have spread over the entire TAP family ('pig', 'salt'), or all over Alor Pantar ('maize'), while others show more regional diffusion patterns, particularly in the languages of South and East Alor. Where a loan can be seen to originate in (only) the Flores-Lembata region, it has spread to the languages of Pantar, Straits and West/Central Alor, but not beyond to the languages of South and East Alor, (40e). Where an individual language can be identified as donor, it is often a language from Timor, although both Malay and Alorese in the Pantar-Straits region have also been identified as donors, see (40f).

(40) Overview of donor regions of loans discussed in the paper.²²

Concept	PMP or lower proto forms; sets of related forms	Recipient language(s)
<i>a. Unknown donor region</i>		
'deer'	PWMP *uRsah 'deer' FL #rusa PRM #rusa	PAP *arusa 'deer', across AP
'skin'	PMP *kulit 'skin; bark'	#kuli 'skin', across TAP
<i>b. Various donor regions</i>		
'bride price'	PMP *bəli	Across TAP
<i>c. Donor region in Flores-Lembata and/or Timor</i>		
'sew'	pre-Rote Meto **sauR 'sew' Lamaholot, Kedang #saur 'sew'	AP languages in East Alor
'sew'	pre-Rote Meto **sədu(t) 'weave', PRM *seru 'weaving sword'	All AP languages, except East Alor
'weave'	PMP *tenun, *tinun 'weave' PRM *tenu 'weave'	AP languages in Central and East Alor

²² A form with * represents a reconstructible proto form, a form with # represents sets of similar lexemes for which a proto form has not been reconstructed.

‘navel’ PMP *pusej AP languages in Pantar, Straits, West Alor, East Alor

d. *Timor donor region*

‘pig’ PMP *babuy PTAP *baj, across TAP
 ‘salt’ PMP *qasiRa ‘salt’, PTB *asiRa PTAP *asir, across TAP
 ‘slave’ PMP *qaRta ‘outsider(s), alien person(s)’ Bunak, Makalero
 ‘needle’ PMP *zaRum ‘needle’ TAP in Timor, AP languages in East Alor
 ‘weave’ pre-Rote Meto **sədu(t) ‘weave’ Bunak, Makasae
 ‘breast’ PMP *susu, PRM *susu TAP in Timor, AP in South Alor
 ‘seed’ PMP *binehiq ‘seed rice, rice set aside for the next planting’ Bunak, AP in Central and East Alor
 ‘maize’ ?PMP *batad ‘sorghum sp., *Andropogon sorghum*’ AP languages across all of Alor and Pantar
 Regional #batar (< Makassar *batara*??)

e. *Flores-Lembata donor region*

‘needle’ PMP *batuR ‘weave’ AP languages in Pantar, Straits and West Alor
 ‘breast’ PMP *susu, PFL *(t)usu AP languages in Pantar, Straits, West and Central Alor

f. *Individual donor language*

	Donor language	Recipient language
‘deer’	Tetun <i>bibi rusa</i>	Bunak, Makasae
‘king, ruler’	Tetun <i>liurai</i>	Bunak, Makasae
		AP languages in South and East Alor
‘slave’	Tetun <i>malae</i> ‘foreigner’	Across TAP
	Malay/Ind (< Dutch) <i>jongos</i> ‘houseboy’	Kaera, Blagar-Tuntuli
‘king, ruler’	Adonara Lamaholot <i>raja</i> ‘king’ or Indonesian/Malay <i>raḏza</i> ‘king’	AP languages in Pantar
‘skin’	Tokodede <i>kuluta</i> ‘skin’	Wersing <i>klut</i>
‘navel’	Tokodede <i>kupusa</i> ‘navel’	Wersing <i>kubu</i>
	Alorese <i>puhɔr</i> ‘navel’	Blagar <i>puar</i> , Reta <i>pual</i> , <i>puhal</i> , Adang <i>puhei</i>

A relatively high number of pre-modern MP loans appear in (i) Bunak, (ii) South and East Alor, and (iii) the Pantar-Straits. It is possible to identify a few individual donor languages in these regions: Tetun for Bunak, Tokodede and Tetun for languages in South and East Alor, Alorese and Malay for languages of the Pantar-Straits region, see (40f). However, in most cases, the donor language remains unknown.

The three regions can be considered different zones of contacts between TAP speakers and MP communities for two reasons: first, because different lexemes were borrowed in each of the regions, and second, if the same concept was borrowed, as in 'breast' and 'needle', the borrowing involved different forms. It is expected that the TAP languages were in contact with MP in different locations, because Pantar-Straits and South and East Alor as well as Bunak are geographically remote from each other, and there was likely little or no direct contact between them. At the same time, sea currents and sailing proximity allowed speakers in South and East Alor to have contact with communities on the northern coast of Timor island, while communities in the Pantar-Straits were oriented towards the islands Lembata, and Flores beyond it.²³ And the Bunak as inland people on Timor had yet a different set of MP communities as neighbours in central Timor.

5 Discussion and Conclusions

The social context in which the contact between groups takes place plays an important role in determining how linguistic changes caused by contact are shaped and constrained (Muysken 2010). Further, diagnosing contact-induced change may help to reconstruct the history of small-scale speech communities (Ross 2013). Bilingually-induced change is change which bilingual speakers introduce into one of their languages on the model of their other language (Ross 2013: 6). It typically leads to lexical calques (loan translations), grammatical calquing which copies grammatical forms but not their syntax, or syntactic restructuring, which copies both the grammatical forms and their syntax (Ross 2013: 27). Shift-induced change is change introduced by speakers who abandon

23 Numeral systems also present evidence for these regionally bound contacts between AP languages and MP languages in the west and the south: Kedang (Lembata island) has borrowed a unique quinary numeral from Pantar languages, and the north-central Timor languages Tokodede and Mambae have quinary numerals from 'six' through 'nine', a pattern that stands out against the typically conservative numeral systems of the Austronesian languages elsewhere on Timor (Schapper & Klamer 2017).

the community language in favour of another language in their repertoire, the language to which they are shifting. Shift-induced changes mentioned in the literature include phonological transfer, constructional transfer, and simplified (morpho)syntax (Ross 2013: 30). Limited and scattered lexical borrowing from MP into TAP, as discussed in this paper, points to contacts that neither involved bilingualism nor shift.

Recent studies of language contact in the Lesser Sunda region have shown that contact between MP and non-MP (TAP) languages led to different types of language change, and in what follows the findings of the current paper are placed in the context of the different contact situations attested in the region (see also Klamer, to appear).

The first type of contact situation is when there was a relatively *short* period involving a large group of speakers who were bilingual in an MP and non-MP language, followed by a shift to the AN language that was initially spoken as second language by the speakers. This is likely to have happened in the history of Sika (Elias 2018: 119), and in the history of Proto Central Flores (Fricke 2019). The outcome of this type of language contact has been a simplification of the morphology of the MP language they had shifted to, because the shift involved adults who learned the second language imperfectly. The effect of the non-MP substrate language on the MP language is the addition of some new vocabulary (19% in Sika since Proto Flores-Lembata times; Fricke 2019). No syntactic features without accompanying lexicon of the substrate non-AN language ended up in the MP language.

Second, there are several attested cases where there was a *prolonged* period of intense and intimate language contact in the form of bilingualism in a non-MP language and MP language over several generations, which was then followed by a shift to the MP language. This has happened in the history of Proto Flores-Lembata, and again in its descendants Kedang, and Lamaholot (Fricke 2019: 416–417). The effect of the non-MP substrate language on the shifted MP languages Kedang and Lamaholot was the addition of a significant amount of new vocabulary (34% in Western and Central Lamaholot and 24% in Kedang since the time of proto Flores Lembata). In addition, there was a change in the syntax of the MP languages, and some semantic features were added to it (cf. the overview in Fricke 2019: 411–413).

A similar contact situation happened in Timor in the history of Uab Meto in the Proto Rote Meto group. The effect of that contact has been that Meto now has two parallel lexicons, each with their own set of regular sound correspondences: one containing reflexes of Proto MP lexemes, the other containing lexemes for which no MP origin has been found (Edwards 2016; 2018a). The sheer size of the non-MP vocabulary (including basic vocabulary), and the fact

that it has restructured the phonological system of the language, points to a prolonged period of intimate contact between one or more incoming MP language(s), and one or more non-MP languages that were spoken in the region before their arrival, followed by a shift to the MP language.

On Timor, there are also situations where MP speakers are on their way of shifting to a TAP languages: MP Makuva speakers have almost entirely shifted to TAP Fataluku, and MP Naueti and Waima'a show serious Makasae influence. In the past, shifts must have happened in the history of Bunak. The modern lexicon of Bunak contains 30% of MP vocabulary including many items of core vocabulary (Schapper 2011: 37). Certain syntactic constructions in Bunak show a clearly Austronesian (verb-medial) word order (e.g. in the 'give' construction, Klammer and Schapper 2012: 196–197). In some of the loans from MP Tetun, the original Tetun morphology has been reanalysed to fit the Bunak patterns (e.g. the Tetun causative prefix *ha-* has been reanalysed as part of the Bunaq inflectional paradigm, Schapper 2011: 41–42). Large non-inherited vocabularies coupled with morpho-syntactic changes in the target language typically point to a history involving a prolonged or repeated periods of bilingualism.

The third situation is when the bilingualism is *stable* and can go on for centuries rather than generations, without ending in a shift. An example of this situation is MP Alorese, spoken in communities consisting of bilinguals whose first language is non-MP Adang and second language is Alorese as described by Moro (2021, 2018, 2019). After a short period of complexification which likely involved young speakers (Moro 2018; Moro & Fricke 2020), Alorese underwent severe simplification of morphology (Klammer 2011; 2012; 2020; To appear; Moro 2019), and these simplified patterns remained stable over many generations. This implies that the contact must be long-term, intense, and multi-purpose involving a community of bilinguals with a large number of second language speakers (Kusters 2003; Trudgill 2011; Moro 2018). The simplifying second language may (originally) have been used as a trade language or lingua franca, but for any changes to become entrenched in it, it must have been used as a second language in wider communicative contexts. This second language may be the language of a technologically, politically, or culturally dominant group that the speakers of other languages wish to communicate or associate with, but it may also be the language of a community that is incorporating many foreign adults (such as spouses or slaves) with different linguistic backgrounds. The latter is probably what characterizes the Alorese.

A language spoken as a second language can become a shifted language when the second language speakers are a minority and die out, while their offspring grows up speaking the community language as first language. This

is likely what happened in all the cases discussed above, except Alorese. The Alorese case shows that, if the number of second language speakers in a community is sufficiently large (e.g., constituting half or more of the population, Moro 2019), and if there is a constant influx of new second language speakers during many generations, then stable bi-lingual communities can exist for centuries without shifting to either of the languages spoken in the community.

A fourth type of contact situation is when there is relatively superficial contact in limited socio-cultural domains such as trade or marriage negotiations, which does not require a community to be bilingual. I suggest that the contact of AP communities with MP speakers was of this relatively superficial type. (The TAP languages of Timor (e.g. Bunak, discussed above, (Schapper 2011); and Fataluku (McWilliam 2007) had a different, more intense contact history with MP speaking populations.) The evidence for the superficial contact events in AP languages is that the number of MP loans attested in these languages are overall rather limited: (Robinson 2015) estimates that the percentage of Austronesian loanwords on a 200-word Swadesh list for twelve different Alor-Pantar languages is about 8 percent. Above we have seen that loanwords are scattered over various semantic domains. Further, only a few specific donor languages can be traced, and overall, the donor regions are rather diffuse entities, and in the lexeme sets, various levels of (ir)regularity in sound correspondences apply.

Non-lexical evidence of language contact, such as changes where a syntactic structure was borrowed, could consist evidence for an earlier stage that involved a bilingual community. To date, no evidence of MP grammatical structures having diffused into any of the AP languages has been reported. An illustration of a MP influence the syntactic domain would for instance be the change of word order in AP languages with subsequently different grammaticalizations of serial verb constructions. For example, the typical TAP head-final [Object V_1 V_2] serial verb configuration leads to the V_1 developing into a *postposition* as attested across the TAP family (Klamer 2018). In contrast, the typical MP head-initial [V_1 V_2 Object] configuration leads to V_2 becoming a *preposition*. To become fully schematic and entrenched, a new word order must become the most frequent order in a speech community. This type of change needs intense, continued, and long-term contact, typically involving several centuries of bilingualism (Backus, Seza Doğruöz & Heine 2011). While several proto TAP verbs appear to have grammaticalized from serial verbs into postpositions in a similar way across the TAP family, in none of the languages do we find traces of an alternative MP order in the serial verb domain. (In contrast, the MP language Tetun in Timor does reveal traces of non-MP structures in the serial verb domain, Klamer 2018.)

In sum, the lexical evidence presented in this paper suggest that contact between speakers of TAP languages on Alor and Pantar with speakers of MP languages was relatively superficial and limited, unlike the contact between the TAP languages of Timor and MP speakers there. The overall lack of grammatical structures in the AP languages that reflect MP influence suggest that there is no AP language with a history of prolonged bilingualism with an MP language. Neither is there evidence that there once was an MP speaking population that shifted to an AP language. Again, the situation with the TAP languages in Timor is more complex, and for Bunak in particular it must have involved a long and/or repeated history of bilingualism.

To conclude, with the exception of MP Alorese, which has been present in the Alor Pantar area since the 15th century and remains to be spoken in bilingual MP-AP communities until today, current evidence suggests that none of the modern languages of Alor and Pantar has a history involving bilingualism with, or shift from, an MP language. TAP communities have been in contact with MP speaking groups since the stage of proto TAP, thousands of years ago, but the contacts remained superficial, and limited to circumscribed domains involving the transfer of technology, goods and individual people.

Acknowledgements

I would like to thank Sander Adelaar, Owen Edwards, Hanna Fricke and Tom Hoogervorst for their valuable feedback on previous versions of this chapter.

Bibliography

- Backus, Ad, A. Seza Doğruöz & Bernd Heine. 2011. Salient stages in contact-induced grammatical change: Evidence from synchronic vs. diachronic contact situations. *Language Sciences* 33(5). 738–752. DOI: <https://doi.org/10.1016/j.langsci.2011.02.004>
- Cense, A.A. 1979. *Makassaars-Nederlands Woordenboek*. 's Gravenhage: Martinus Nijhoff.
- Coenen, Jules. 1953. De Ihandin taal, gesproken op het schiereiland Onin, aan de westkust van New Guinea. Manuscript.
- Daigle, Benjamin. 2015. A grammar sketch of Batuley: An Austronesian language of Aru, eastern Indonesia. Master Thesis, Leiden University. <https://studenttheses.universiteitleiden.nl/handle/1887/43444>
- Edwards, Owen. 2016. Parallel sound correspondences in Uab Meto. *Oceanic Linguistics* 55(1). 52–86.

- Edwards, Owen. 2018a. Parallel Histories in Rote-Meto. *Oceanic Linguistics* 57(2). 359–409.
- Edwards, Owen. 2018b. Top-down historical phonology of Rote-Meto. *Journal of the Southeast Asian Linguistics Society (JSEALS)* 11(1). 63–90. DOI: <http://hdl.handle.net/10524/52421>
- Edwards, Owen. 2019. Reintroducing Welaun. *Oceanic Linguistics* 58(1). 31–58.
- Edwards, Owen. 2020. *Metathesis and unmetathesis in Amarasi*. Berlin: Language Science Press. Retrieved from <http://langsci-press.org/catalog/book/228>
- Edwards, Owen. 2021. *Rote-Meto comparative dictionary*. Canberra: ANU Press. DOI: <http://doi.org/10.22459/RMCD.2021>
- Elias, Alexander. 2018. Lio and the Central Flores languages. Research Master thesis Leiden University. <https://hdl.handle.net/1887/69452>
- Forth, Gregory L. 1983. Time and temporal classification in Rindi, Eastern Sumba. *Bijdragen Tot de Taal-, Land- En Volkenkunde* 139(1). 46–80.
- Fox, James J. 1991. The heritage of traditional agriculture in Eastern Indonesia: Lexical evidence and the indications of rituals from the outer arc of the Lesser Sundas. *Indo-Pacific Prehistory Association Bulletin* 10. 248–262.
- Fricke, Hanna. 2017. The rise of clause-final negation in Flores-Lembata, Eastern Indonesia. *Linguistics in the Netherlands* 34. 47–62.
- Fricke, Hanna. 2019. *Traces of language contact: the Flores-Lembata languages in eastern Indonesia*. PhD dissertation Leiden University. Amsterdam: LOT Publications.
- Glover, Ian. 1986. *Archaeology in Eastern Timor, 1966–1967*. Canberra: Australian National University.
- Groves, Colin P. 1981. *Ancestors for the Pigs: Taxonomy and Phylogeny of the Genus Sus*. Canberra: Australian National University.
- Hägerdal, Hans. 2009. The Exile of the Liurai: A Historiographical Case Study from Timor. In Hans Hägerdal (ed.), *Responding to the West: Essays on Colonial Domination and Asian Agency*, 45–66. Amsterdam: Amsterdam University Press.
- Hägerdal, Hans. 2010a. Cannibals and pedlars. *Indonesia and the Malay World* 38(111). 217–246. DOI: <https://doi.org/10.1080/13639811.2010.489356>
- Hägerdal, Hans. 2010b. Van Galens memorandum on the Alor islands in 1946. An annotated translation with an introduction. Part 1. *HumaNetten* 25. 14–44.
- Hägerdal, Hans. 2011. Van Galens memorandum on the Alor islands in 1946. An annotated translation with an introduction. Part 2. *HumaNetten* 27. 53–96.
- Hägerdal, Hans. 2012. *Lords of the land, lords of the sea: conflict and adaptation in early colonial Timor, 1600–1800*. Leiden: KITLV Press.
- Haspelmath, Martin & Uri Tadmor. 2009. The Loanword Typology project and the World Loanword Database. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 1–34. Berlin: Mouton de Gruyter.
- Holton, Gary, Marian Klamer, František Kratochvíl, Laura C. Robinson & Antoinette

- Schapper. 2012. The historical relations of the Papuan languages of Alor and Pantar. *Oceanic Linguistics* 51(1). 86–122.
- Holton, Gary & Laura C. Robinson. 2017a. The internal history of the Alor-Pantar language family. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology* (2nd ed.), 49–90. Berlin: Language Science Press.
- Holton, Gary & Laura C. Robinson. 2017b. The linguistic position of the Timor-Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology* (2nd ed.), 147–190. Berlin: Language Science Press.
- Hoskins, Janet. 1996. Introduction: Headhunting as practice and trope. In *Headhunting and the social imagination in Southeast Asia*, 1–49. Stanford: Stanford University Press.
- Huber, Juliette. 2011. *A grammar of Makalero: A Papuan language of East Timor*. PhD dissertation Leiden University. Utrecht: LOT Publications: Leiden University dissertation.
- Jones, Gavin W., Terence H. Hull & Maxnah Mohamad (eds.). 2011. *Changing marriage patterns in Southeast Asia: Economic and socio-cultural dimensions*. Oxon / New York: Routledge.
- Kaiping, Gereon, Owen Edwards & Marian Klamer. 2019. LexiRumah 3.0.0. Leiden University Centre for Linguistics. Retrieved from <https://lexirumah.model-ling.eu/lexirumah/>
- Kaiping, Gereon and Marian Klamer. 2022. The dialect chain of the Timor-Alor-Pantar language family: A new analysis using systematic Bayesian phylogenetics. *Language Dynamics and Change*, 12(2), 274–326. DOI: <https://doi.org/10.1163/22105832-bja10019>
- Klamer, Marian. 2011. *A short grammar of Alorese (Austronesian)*. Munich: Lincom GmbH.
- Klamer, Marian. 2012. Papuan-Austronesian language contact: Alorese from an areal perspective. In Nicholas Evans & Marian Klamer (eds.), *Melanesian languages on the edge of Asia: Challenges for the 21st century*, 72–108. Honolulu: University of Hawai'i Press. Retrieved from <http://scholarspace.manoa.hawaii.edu/handle/10125/4561>
- Klamer, Marian. 2014. Kaera. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar: Sketch grammars*, Vol. 1, 97–146. Berlin: Mouton de Gruyter.
- Klamer, Marian. 2018. Typology and grammaticalization in the Papuan languages of Timor, Alor, and Pantar. In Heiko Narrog & Bernd Heine (eds.), *Grammaticalization from a Typological Perspective*, 235–262. Oxford: Oxford University Press.
- Klamer, Marian. 2020. From Lamaholot to Alorese: Morphological loss in adult language contact. In David Gil and Antoinette Schapper (eds), Amsterdam: Benjamins. In David Gil & Antoinette Schapper (eds.), *Austronesian undressed: How and why languages become isolating*, 339–367. Amsterdam: Benjamins.
- Klamer, Marian. 2023. Word classes in Timor-Alor-Pantar and the Papuan region. In Eva van Lier (ed.), *Oxford Handbook of Word Classes*. Oxford: Oxford University Press.

- Klamer, Marian. To appear. Austronesian-Papuan contact in pre-modern eastern Indonesia. In Nicholas Evans & Sebastian Fedden (eds.), *The Oxford Guide to the Papuan languages*. Oxford: Oxford University Press.
- Klamer, Marian & Antoinette Schapper. 2012. 'Give' constructions in the Papuan languages of Timor-Alor-Pantar. *Linguistic Discovery* 10(3). 174–207.
- Kusters, Christiaan Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. PhD dissertation Leiden University. Utrecht: LOT Publications.
- Larson, Greger, Thomas Cucchic, Masakatsu Fujita, Elizabeth Matisoo-Smith, Judith Robins, Atholl Anderson, ... etc. 2007. Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. *PNAS* 104(12). 4834–4839. DOI: <https://doi.org/10.1073/pnas.0607753104>
- Le Roux, C.C.F.M. 1929. De Elcanos tocht door den Timorarchipel met Magalhães' schip 'Victoria'. In *Feestbundel, uitgegeven door het koninklijk bataviaasch genootschap van kunsten en wetenschappen bij gelegenheid van zijn 150 jarig bestaan: 1778–1928*, 1–99. Kolff: Weltevreden.
- McWilliam, Andrew. 2007. Austronesians in linguistic disguise: Fataluku cultural fusion in East Timor. *Journal of Southeast Asian Studies* 38(2). 355–375.
- Moro, Francesca. 2018. The Plural Word hire in Alorese: Contact-induced change from neighboring Alor-Pantar languages. *Oceanic Linguistics* 57(1). 177–198.
- Moro, Francesca. 2019. Loss of morphology in Alorese (Austronesian): Simplification in adult language contact. *Journal of Language Contact* 12(2). 378–403.
- Moro, Francesca. 2021. Multilingualism in eastern Indonesia: linguistic evidence of a shift from reciprocal to asymmetrical multilingualism. *International Journal of Bilingualism* 25(4). 1102–1119. DOI: <https://doi.org/10.1177/13670069211023134>
- Moro, Francesca & Hanna Fricke. 2020. Contact-induced change in Alorese *give*-constructions. *Oceanic Linguistics* 59(1). 116–147. DOI: [https://doi.org/Oceanic Linguistics 59\(1\), 116–147. doi:10.1353/ol.2020.0008](https://doi.org/Oceanic Linguistics 59(1), 116–147. doi:10.1353/ol.2020.0008).
- Muysken, Pieter. 2010. Scenarios for language contact. In Raymond Hickey (ed.), *The handbook of language contact*, 265–281. Chichester, UK; Malden, MA: Wiley-Blackwell.
- O'Connor, Sue. 2017. Rethinking the Neolithic in Island Southeast Asia, with particular reference to the archaeology of Timor-Leste and Sulawesi. *Archipel [Online]* 90. 15–47.
- O'Connor, Sue, Rintaro Ono & Chris Clarkson. 2011. Pelagic Fishing at 42,000 Years Before the Present and the Maritime Skills of Modern Humans. *Science* 25(334). 1117–1121.
- Raben, Remco. 2008. Cities and the slave trade in early-modern Southeast Asia. In Peter Boomgaard, Dick Kooiman, & Henk Schulte Nordholt (eds.), *Linking destinies: Trade, towns and kin in Asian history*, 119–140. Leiden: KITL Press.
- Robinson, Laura C. 2015. The Alor-Pantar (Papuan) languages and Austronesian con-

- tact in East Nusantara. In Malcolm D. Ross (ed.), *Language contact and Austronesian historical linguistics*. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Andrew K. Pawley, Robert Attenborough, Jack Golson, & Robin Hide (eds.), *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–65. Canberra: Research School of Pacific & Asian Studies, Australian National University. Retrieved from http://www.academia.edu/download/31189925/Ross_Papuan_Pasts.pdf
- Ross, Malcolm. 2013. Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact* 6(1). 5–47. DOI: <https://doi.org/10.1163/19552629-006001002>
- Schapper, Antoinette. 2011. Crossing the border: Historical and linguistic divides among the Bunaq in central Timor. *Wacana* 13(1). 29–49.
- Schapper, Antoinette, Juliette Huber & Aone van Engelenhoven. 2017. The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology (2nd ed.)*, 91–146. Berlin: Language Science Press.
- Schapper, Antoinette & Marian Klamer. 2017. Numeral systems in the Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology (2nd ed.)*, Second edition, 277–328. Berlin: Language Science Press.
- Schapper, Antoinette & Emilie Wellfelt. 2018. Reconstructing contact between Alor and Timor: Evidence from language and beyond. In *Language contact and substrate in the languages of Wallacea, Part 2.*, 95–116. Retrieved from <http://hdl.handle.net/10108/92292>
- Trudgill, Peter. 2011. *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Van Lynden, D.W.C. Baron van. 1851. Bijdrage tot de kennis van Solor, Allor, Rotti, Savoe en omliggende eilanden, getrokken uit een verslag vna de Residentie Timor. *Natuurkundig Tijdschrift Voor Nederlands-Indië* 2. 317–336, 388–414.
- Voorhoeve, Clemens L. 1989. The Masked Bird: Linguistic Relations in the Bird's Head Area. In Paul Haenen & Jan Pouwer (eds.), *Peoples on the move*, 78–101. Nijmegen: Centre for Australian and Oceanic Studies.
- Wellfelt, Emilie. 2016. *Historyscapes in Alor: Approaching indigenous history in Eastern Indonesia*. Linnaeus University dissertation. Retrieved from <http://www.diva-portal.org/smash/record.jsf?pid=diva2:971262>

Phonological Innovation and Lexical Retention in the History of Rote-Meto

Owen Edwards

1 Introduction

In this paper I undertake a historical investigation of Rote-Meto, one low-level Austronesian subgroup in Wallacea, in order to determine the kind of contact these languages may have undergone. This analysis is based on a database of 1,173 reconstructions to Proto Rote-Meto (PRM) or one of its daughter nodes, published as Edwards (2021).¹ I investigate three areas of PRM: segmental inventory, lexicon, and regularity of sound change.

In § 2 I examine the segmental inventory of PRM. I show that when compared with Proto Malayo-Polynesian (PMP), the segmental inventory of PRM has been transformed according to regional norms. Furthermore, certain PRM segments are disproportionately represented in words not known to be inherited from PMP, and certain other segments are over-represented in PMP inheritances. This indicates that the transformation of the PRM segmental inventory mainly occurred due to acquisition of new words with these segments.

In § 3 I examine the lexicon of PRM. I find that PRM has an entirely expected lexical profile for an Austronesian (AN) language. PMP inheritances occur in domains more resistant to borrowing, while words not known to be inherited from PMP show signs of being borrowed. The “Austronesian” nature of the lexicon thus contrasts with the “non-Austronesian” character of the segmental inventory.

In § 4 I examine the regularity of sound change from PRM to its daughter languages, and from PMP to PRM. Examination of regularity of sound change between PRM and its daughters shows that words confined to west Timor show a greater proportion of irregular sound changes. This indicates a larger portion of words confined to west Timor were acquired after the break-up of the proto language. Between PMP to PRM, a large number of unconditioned splits have

¹ Edwards (2021) is freely downloadable in computer searchable formats from <http://hdl.handle.net/1885/251618>.

occurred with several factors contributing including sound change in progress, as well as contact with both AN and non-AN languages.

I conclude in § 5 with a summary of the findings. The history of PRM presents a complex picture. While contact has clearly played a role in the history of this family, there is evidence of multiple kinds of contact at multiple stages of the history of Rote-Meto. This contact has occurred with both AN and non-AN languages.

Furthermore, different domains of PRM potentially attest different kinds of contact. The lexicon attests large-scale borrowing, which could be a result of superficial contact. On the other hand, examination of the segmental inventory paints a different picture, pointing to much more intense contact with substrate languages. This underscores the importance of multiple perspectives in investigations of language history.

1.1 *Language Background*

Rote-Meto is a low-level AN subgroup composed of the languages of Rote Island immediately to the south-west of the island of Timor and the Meto language/dialect cluster which dominates the western part of Timor. Synchronically, the languages of Rote Island and the Meto language/dialect cluster are each comparable to the Romance or West Germanic continua in Europe, whereby speakers of neighbouring varieties are generally able to understand one another, but with mutual intelligibility reduced or blocked between distant varieties.

The island of Rote is divided into nineteen political units known in the anthropological literature as domains (*nusak* or *nusa?* in the languages of Rote), and many speakers claim that each domain has its own language (Fox 2016:233). See Edwards (2021:30–32) for a summary of the different classifications of the languages of Rote that have been proposed in the literature. A map of the domains of Rote is given in Figure 4.1, in which domains are coloured according to one classification of the “languages” of Rote Island.

Metó (a.k.a. Uab Metó, Dawan[ese], Timorese, or Atoni) is a cluster of speech varieties spoken in the western part of Timor. Metó speakers usually identify their speech as a single language but recognise more than a dozen named varieties. These varieties themselves have named “dialects”, with further differences being found between villages of a single “dialect”. A map of self-identified Metó varieties is given in Figure 4.2.

Within the AN family, Rote-Meto belongs to the Timor-Babar subgroup, which contains most, but not all, other AN languages of Timor;² the languages

² Welaun, Kemak, Tokodede, and Mambae are not part of Timor-Babar. These languages form a Central Timor subgroup (Edwards 2019:42–49).

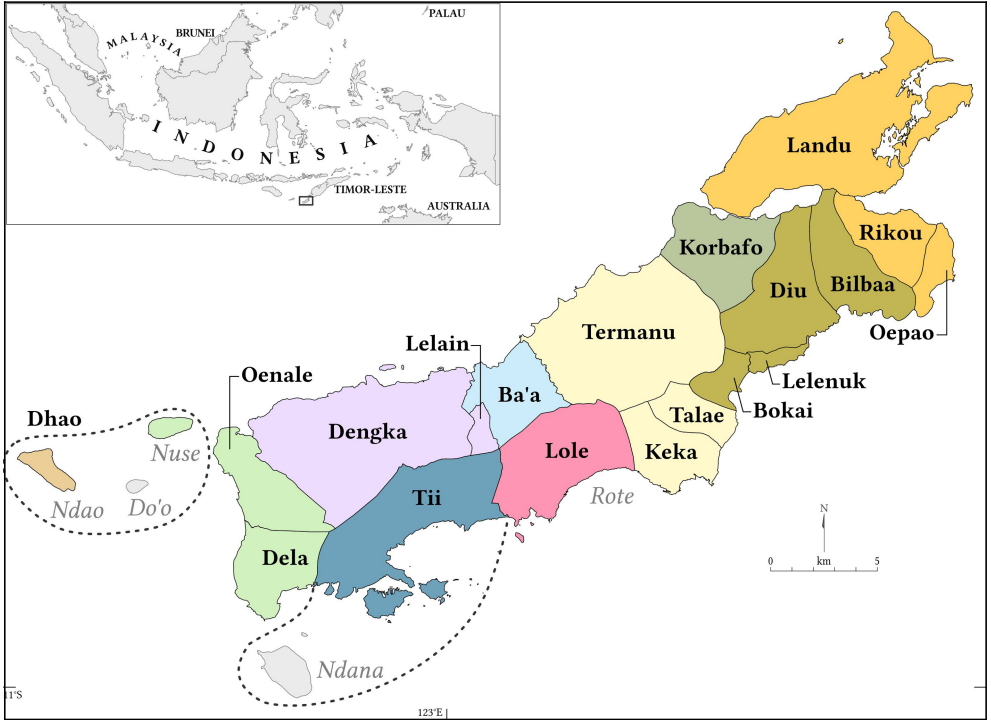
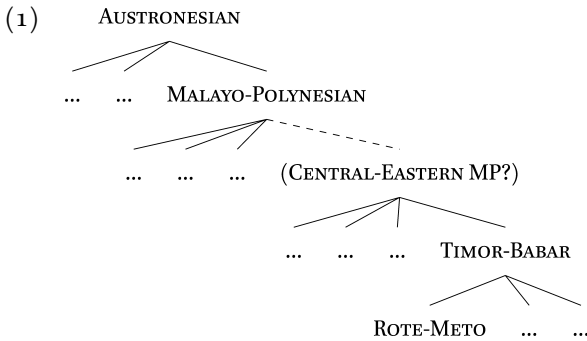


FIGURE 4.1 Domains and Languages of Rote Island

of the islands of the Indonesian regency of Southwest Maluku (*Kabupaten Maluku Barat Daya*), as well as Selaru, Seluwasan, and Makatian of south-west Tanimbar. The position of Rote-Meto within the AN language family is shown in (1).



Phonological evidence for Timor-Babar as a subgroup of MP comes from shared *p > *h, with subsequent *h > ∅ in many cases. Within Timor-Babar, there is

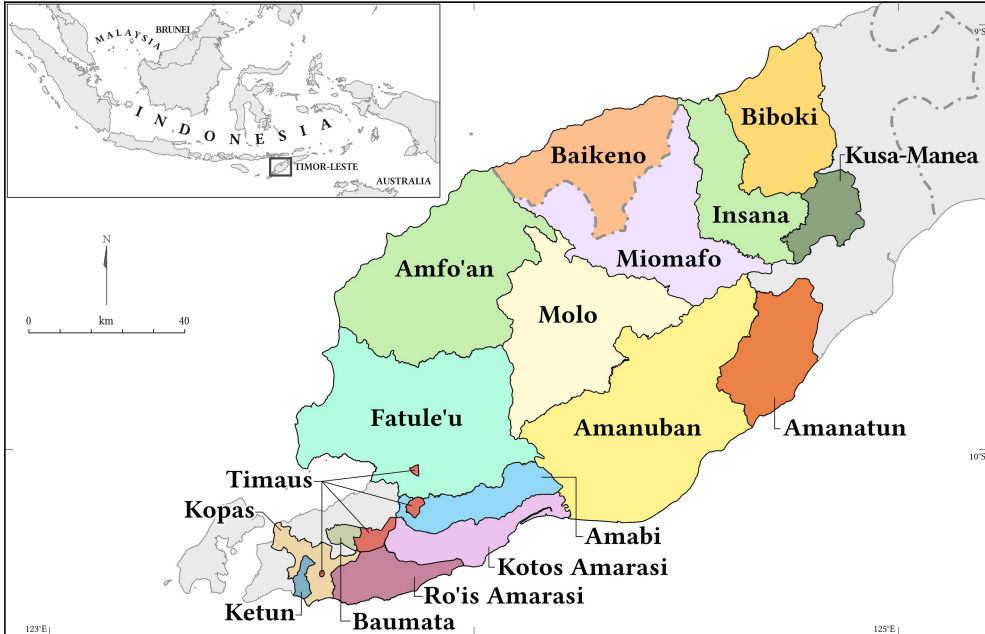


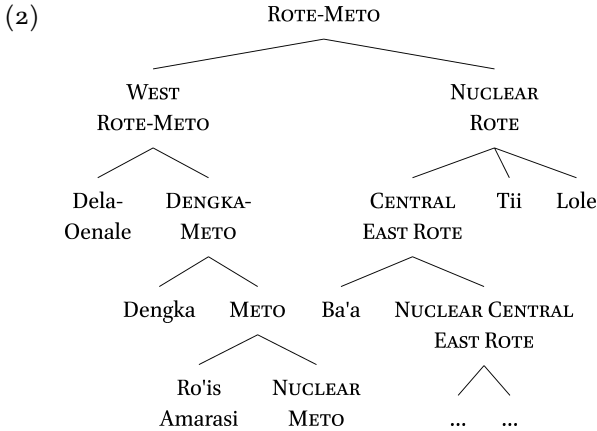
FIGURE 4.2 Self Identified Varieties of Meto

phonological evidence that Rote-Meto forms a distinct subgroup. This evidence comes from shared PMP **wa* > PRM **o* in nine words, lowering of high vowels to mid before word final PMP **R* in eight words, and subsequent loss of **R* in most cases (Edwards 2018b; 2021:76–86).

Within Rote-Meto there are two primary branches. WEST ROTE-METO contains the Meto cluster, along with Dela-Oenale and Dengka of western Rote. NUCLEAR ROTE contains the other languages of Rote Island. NUCLEAR ROTE further contains CENTRAL EAST ROTE, a subgroup which excludes Tii and Lole. A tree diagram showing the structure of the Rote-Meto family is given in (2). Due to space constraints, the internal structure of NUCLEAR CENTRAL EAST ROTE is not shown. See (Edwards 2021:57–66) for details.

The Rote-Meto comparative dictionary on which this paper is based contains 1,173 reconstructions for PRM or one of its lower branches. For the principles on which reconstructions are made, and the levels to which they are assigned, see Edwards (2021:69–71). For the purposes of this paper, it is sufficient to know that reconstructions to branches below PRM are usually only included when possible cognates have been identified in other AN languages.³

3 There are two exceptions to this general rule. Firstly, not all reconstructions to Proto West Rote-Meto have known cognates in other languages; e.g. Proto West Rote-Meto **ka-batus*



Such words were probably present in PRM, but cognates have not (yet) been identified—perhaps due to loss—in other branches.

Thus, for instance, cognates of Proto Meto *metam ‘black’ are not known in the Rote languages. Nonetheless, it is inherited from PMP *ma-qitəm and was almost certainly present in PRM, even though reflexes appear to have been lost in Rote. Similarly, Proto Nuclear Rote *hesu ‘fart’ has no known cognates in West Rote-Meto, but cognates in many regional languages. It was almost certainly present in PRM, with reflexes lost in West Rote-Meto.

1.2 *The Proto Rote-Meto Lexicon by the Numbers*

1,173 reconstructions have been made to PRM or one of its lower branches, and the presence of cognates in certain other languages has been tracked.⁴ The breakdown of these PRM reconstructions according to where cognates are attested is summarised in (3), and mapped in Figure 4.3. (Broken lines in Figure 4.3 serve only to make the distribution of each stratum clearer.)

The set of words in (3a) is referred to as the *Austronesian stratum* (AN), the set of words in (3b) is referred to as the *regional stratum* (regi.), and the set of words in (3c) is referred to as the *west Timor stratum* (wTim.). While the distribution of cognates within the regional stratum has been tracked according to four sub-strata, for the most part no differences were found between these strata.

⁴ ‘sea snail’. Secondly, reconstructions are occasionally made to lower levels when they are semantically and formally similar to a PRM reconstruction; e.g. Proto Nuclear Rote *tengga ‘hand span’ is included as it is similar to Proto West Rote *hanġa ‘hand span’.

4 In addition to the 1,173 reconstructions, Edwards (2021) also contains 84 sets which are a result of borrowing. These sets do not figure in the analyses in this paper.

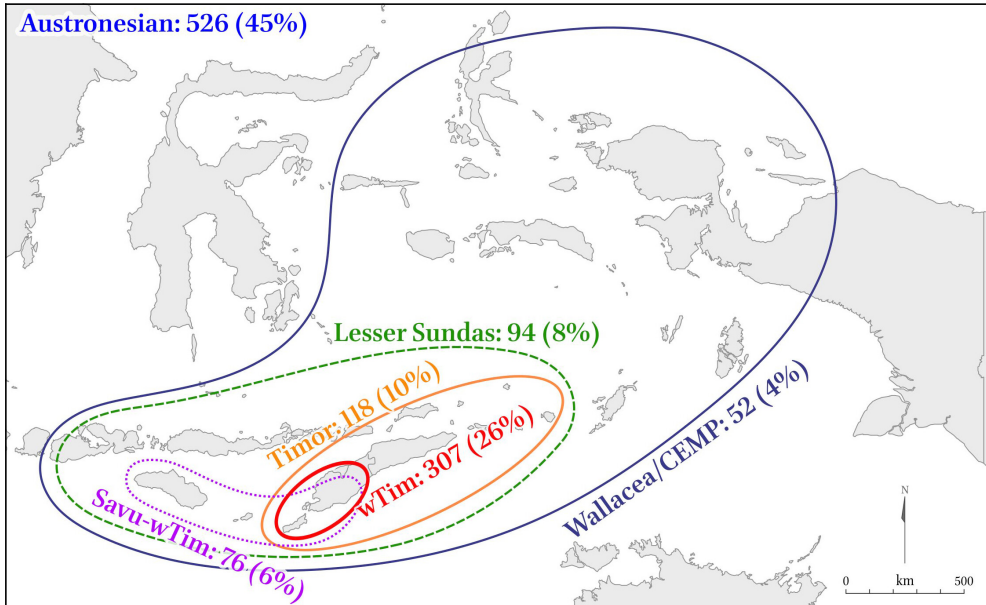


FIGURE 4.3 Proto Rote-Meto Strata

- (3) Proto Rote-Meto strata
- a. Austronesian stratum (AN): 526 (45%) inheritances from PMP
 - b. Regional stratum (regi.) 340 reconstructions have cognates in other regional languages (29%)
 - i. 118 (10%) have cognates in Timor (excluding Helong—see following discussion) and/or south-west Maluku
 - ii. 76 (6%) have cognates in Sumba and/or Hawu
 - iii. 94 (8%) have cognates in the Lesser Sundas
 - iv. 52 (4%) have cognates in other areas of Wallacea, including words reconstructed to putative Proto Central Malayo-Polynesian (PCMP) or Proto Central Eastern Malayo-Polynesian (PCEMP)
 - c. west Timor stratum (wTim): 307 (26%) reconstructions have cognates known only in west Timor (26%)
 - i. 200 (17%) only have known cognates in Rote-Meto languages
 - ii. 107 (9%) also have cognates in Helong (see below)

There are 107 Rote-Meto reconstructions which are currently known to have cognates only in the Rote-Meto languages and Helong. Given that there is no other evidence that Rote-Meto and Helong subgroup together within Timor-Babar, these words probably represent early borrowings between Proto Helong and PRM before the respective sound changes in each language. Such recon-

TABLE 4.1 Cognate sets shared between Rote-Meto and Helong

PRM *gloss PRM	'country' *ijgu	'stocks' *lajge	'snap off' *sejgi	'bright' *maneu-k	'call' *n-oken
Rikou	<i>iku</i>	<i>lake</i>	<i>seki</i>	<i>neu-?</i>	
Termanu	<i>iju</i>	<i>laje</i>	<i>seji</i>	<i>neu-k</i>	<i>n-oke</i>
Oenale	<i>ijgu</i>	<i>lajge</i>	<i>sejgi</i>	<i>meu-?</i>	<i>n-o?e</i>
Kotos Amarasi	<i>iku</i>	<i>nake</i>	<i>n-seki</i>	<i>nmeu</i>	<i>n-o?en</i>
Molo		<i>nake</i>	<i>n-seki</i>	<i>nmeu</i>	<i>n-o?en</i>
Funai Helong	<i>iju</i>			<i>mniu?</i>	
Semau Helong	<i>iju</i>	<i>laje</i>	<i>sijin</i>	<i>niu?</i>	<i>noken</i>

structions usually contain regular sound correspondences, and thus with our current state of knowledge, it is impossible to determine the direction of borrowing. A selection of such cognate sets are given in Table 4.1 to illustrate.

While we could posit that the direction of borrowing was from Proto Helong into PRM this does not provide a solution to the ultimate origin of these reconstructions; it simply shifts the question from PRM to Proto Helong. For this reason, both kinds of reconstructions are treated as a single west Timor stratum.

1.3 *Where Words Come from*

Given that Rote-Meto is an Austronesian subgroup, this raises the question of the origins of the 647 PRM reconstructions which are not known to be inherited from PMP. There are three logical possibilities for new words:

1. derivation
2. language contact
3. coinage / ex-nihilo root-creation

Each of these possibilities, and the extent to which they are known to have contributed to the lexicon of the Rote-Meto languages is discussed below, often with reference to my Amarasi data, which includes a draft dictionary consisting of 2,509 headwords.

One common source of new words is derivation; using morphological or phonological processes of a language to create a new word from a pre-existing word. Derivation includes compounding. My current draft Amarasi dictionary contains 399 derivations, representing 16% of all headwords, with many more derivations undoubtedly existing in the language. In many cases the origin of

a derivation is clear to speakers. Thus, Amarasi *knaa kase* ‘peanut’ is a transparent compound of *knaa?* ‘legume’ and *kase* ‘foreign’ and native speakers are aware of this origin, despite its lexicalised meaning—much like English *blackbird* [ˈblækbɜːd].

However, phonological change (and often semantic shift) can operate to such an extent that a term which was originally polymorphemic is best analysed synchronically as monomorphemic. This process of lexicalisation is known as fusion in the literature (Brinton and Traugott 2005:47–57). For example, Amarasi *atoni?* ‘man, person’ is from PRM *hatahori via regular sound changes.⁵ PRM *hatahori is in turn a compound of PMP *qaRta ‘person of own race’ (semantics from Mahdi 1994:464 ff.) and *qudip ‘alive’. However, this etymology is not known to Amarasi speakers, just like the origin of English *husband* [ˈhʌzbənd] as a compound is unknown to most English speakers.⁶

Another common source of new words is language contact, in particular borrowing. My Amarasi dictionary contains 130 borrowings with an identified source language and a further 329 non-native words occur in my text corpus which may also be borrowings, though some may be insertions and/or code-switching. Borrowing has operated throughout the history of the Rote-Meto languages. Jonker (1908) is a record of the Rote languages as spoken towards the end of the nineteenth century. This work records many borrowings from a variety of sources. Two examples are Termanu *kafa* ‘copper wire’ from Malay *kawat* and *salani* ‘baptise’, ultimately from Arabic *naṣrānī* [nasˤraːniː] ‘Christians’, with initial [na] reanalysed as a third person prefix.

Thirdly, new words can enter the language through coinage, or ex-nihilo root-creation. True coinage—the sheer invention of a word out of nothing—is extremely rare in general usage (McArthur et al. 2018: s.v.) and it is difficult to find examples even in languages whose history is well documented, such as English. The standard English examples are *googol* [ˈgʊːgəl] ‘a very large number, 10¹⁰⁰’ and *Kodak* [ˈkəʊdək].

Words with unknown etymologies probably occur in all languages. One famous English example is *dog* [dɒg] for which none of the proposed etymologies have met with widespread acceptance.⁷ However, the lack of a clear etymology

5 The full pathway was *hatahori > **atahori > **ataholi > **atholi > **atoli > *atoni?*. Intermediate forms are extant in several of the Rote: e.g. Dela *atahori* ‘person’ or Termanu *hataholi* ‘person’.

6 English *husband* [ˈhʌzbənd] is a historic compound of *house* [haʊs] and a now obsolete noun *bond* [bɒnd] ‘householder, master of the house’ (‘husband, n.’ *OED Online*. Oxford University Press, September 2020. Web. 1 December 2020.)

7 The Oxford English Dictionary gives several possible etymologies for *dog* [dɒg] but states “[...]”

does not mean that such words are *ex-nihilo* coinages. Derivation or borrowing are *much* more likely origins. The precise mechanisms of derivation or a donor language may be unrecoverable, but this does not mean we should assume invention. Words are rarely, if ever, made up in a vacuum. Instead, for any novel word to be interpreted and to spread it must build on pre-existing knowledge.

The two kinds of word formation which are closest to coinage are onomatopoeia and conventionalization of baby-talk (nursery language). In western Timor onomatopoeia is mostly limited to names of birds (e.g. Amarasi *koa?* ‘Friarbird, *Philemon* species’) and words describing noises (e.g. Amarasi *tuh-tuu*h ‘uncontrollable sound’). Conventionalized baby-talk is limited to kin terms, such as Amarasi *papa* ‘dad’ and *mama* ‘mum’—both of which may actually be borrowings. Another probable example of conventionalised baby talk in Amarasi is *baba?* ‘parent’s opposite sex sibling’.⁸

Of the three possibilities for new words summarised above, only coinage and language contact are applicable to the lexicon of Rote-Meto as reconstructed in Edwards (2021). Derivation is not included as a possible origin, as derivations with a stem inherited from PMP are placed in the AN stratum,⁹ while derivations involving only elements from unknown sources are included in other strata.¹⁰

As already mentioned, *ex-nihilo* coinage is *extremely* rare. While we can never completely rule out the possibility that some PRM reconstructions were invented in a vacuum after the break-up of PMP, language contact and/or derivation are *much* more likely hypotheses. Only in the case of onomatopoeia and conventionalised baby talk is coinage at all likely. For this reason, the 22 onomatopoeic reconstructions and three likely cases of nursery talk are excluded from the analysis of the lexicon in § 3.

all attempted etymological explanations are extremely speculative. (“dog, n. 1.” *OED Online*. Oxford University Press, September 2020. Web. 3 December 2020.)

8 Amarasi *baba?* may be from PMP *baba, of which Blust and Trussel (2020) state: “As part of universal nursery language this item could have arisen independently in all or many of the languages in which it appears. [...] *baba is just as likely to be an inherited form which resisted regular sound change [...] as a result of the recurrent reinforcement that nursery language provided.” Meto *baba?* has no known cognates in the Rote languages, and is included in the regional stratum in my database.

9 Two examples of reconstructions involving derivation are *[q/k]umaŋ > *saŋuma ‘hermit crab’ and *bañən > *kesufani ‘sneeze’. Both were placed in the AN stratum, even though the origins of initial *saŋj and *kesu are currently unknown.

10 Derivations are not included in the Rote-Meto dictionary when reconstructions of the stem or both members of the compound are already included. Thus, no stratum is over-inflated by the inclusion of many derivations.

It is worth emphasising that the two examples of coinage given above, *googol* [ˈgʊːgɔl] and *Kodak* [ˈkəʊdæk], appear to be the only instances of words created by ex-nihilo coinage in the entire English language which have any degree of currency in the general community. Furthermore, of these, *Kodak* [ˈkəʊdæk] is, at best, a liminal common noun rather than a proper noun. Because ex-nihilo root-creation is so rare, it can be considered a negligible factor in the origin of new words when we consider the lexicon of PRM.¹¹

Finally, when we consider PRM specifically, a particular PRM reconstruction may actually be an inheritance from PMP for which cognates have not yet been identified in other AN languages. While I have consistently searched for etyma of my PRM reconstructions in the online Austronesian Comparative Dictionary (Blust and Trussel 2020), as well as for cognates in certain other regional languages,¹² it is not unlikely that cognates have been missed and/or not yet documented. While future work will almost certainly lead to expansion of the AN stratum, future descriptive work and more data from the Rote-Meto languages will also likely lead to expansion of the regional and west Timor strata. Whatever the final result, the strata not known to be inherited from PMP are likely to remain substantial and their origin(s) will still demand explanation.

To summarise, the extreme rarity of coinage, combined with the fact that derivations are either inherited from MP or of unknown origin, means that the best working hypothesis is that PRM reconstructions not known to be inherited from MP are probably due to contact with non-AN languages.

2 Origins of PRM Segments

The first aspect of PRM which I examine in order to understand its history is the segmental inventory. A number of PRM segments, in particular the series of plosives, are disproportionately represented in different strata. This points to PRM being formed through the meeting of at least two language groups which had distinct segmental inventories; one had an “Austronesian inventory” and one had a “regional inventory”.

11 I know of no traditional cultural practices in western Timor that would encourage coinage in the same way that copyright has in the 20th and 21st centuries.

12 I have consistently checked for cognates in the following languages: Helong (from Balle and Cameron (2014), a draft dictionary with 3,368 headwords), Tetun (from Morris (1984), a dictionary with c. 6,500 headwords), Ili'uun (from de Josselin de Jong (1947), a lexicon of c. 1,500 items), Kisar (from Christensen (in process), a draft dictionary with 2,518 headwords), and Hawu (from Grimes et al. (2008), a dictionary with 1,653 headwords). Jonker's 1908 dictionary also includes copious etymological notes listing putative cognates in other AN languages of his Termanu head-word and these have also been recorded.

TABLE 4.2 Proto Malayo-Polynesian consonants^a

	Labial	Alveolar	Retroflex	Palatal	Velar	Uvular	Glottal
Plos. [-v]	p	t			k	q	
Plos. [+v]	b	d	(d) ^b	g ^j	g		
Affr. [-v.]				(tʃ) ^b			
Affr. [+v.]				dʒ			
Nasal	m	n		ɲ	ŋ		
Fricative		s					h
Lateral		l					
Trill		r~R					
Tap		ɾ					
Glide	w			j			

- a Segments whose traditional transcription differs from IPA are: [d] = *D, [g] = *j, [tʃ] = *c, [dʒ] = *z, [ɲ] = *ñ, [r]~[R] = *R, [ɾ] = *r, and [j] = *y.
- b *D [d] and *c [tʃ] are only distinguished in western Indonesia and are not accepted by all analysts.

The consonant inventory proposed for PMP is given in Table 4.2, transcribed according to their likely phonetic values as assigned by Blust (2013:554–593). While not all analysts agree with this interpretation—see, in particular, Wolff (2010:31–47) for a different proposal—there is broad agreement on the basic system of PMP, with two series of oral stops (voiceless and plain voiced), a set of nasals, fricatives, liquids, and glides.

The PRM system is given in Table 4.3. Note particularly the four series of plosives: voiceless, plain voiced, imploded, and prenasalised. A concise presentation of the evidence for this system is Edwards (2018a: 369–395).

The PRM system, with four series of plosives, presents quite a different typological profile to the PMP system. Furthermore, it is an extremely rare system cross-linguistically. A similar system occurs in only 3% (218/7,302) of lects in Donohue et al. (2013). Thus, it is not the kind of system expected to arise through processes of “normal” language change. Nonetheless, from a more local perspective, the PRM system fits the regional typology well. A similar system occurs in 41% (31/75) of lects spoken in the area between Southeast Sulawesi, Sumbawa, and western Timor (Donohue et al. 2013).¹³

13 That is 31/75 lects in this area have at least one segment belonging to each of these categories: plain-voiceless plosive, plain-voiced plosive, prenasalised plosive, and implosive.

TABLE 4.3 Proto Rote-Meto consonants

	Labial	Alveolar	Palatal	Velar	Glottal
Plosive [-VOICE]	p	t		k	ʔ
Plosive [+VOICE]	b	d	(dʒ) ^a		
Plosive [+GLOTTAL]	ɸ	ɖ			
Plosive [+PRENAS.]	mb	nd		ŋg	
Nasal	m	n		ŋ	
Fricative	f	s			h
Lateral		l			
Trill		r			
Glide	(w) ^a				

a _{PRM} *dʒ and *w currently only have two attestations each.

Many languages of central Flores have exactly the same four series of plosives (e.g. Ende [McDonnell 2009:198]; Keo [Baird 2002:34]; Rongga [Arka et al. 2007:13]), and this system is reconstructed to Proto Central Flores (Elias 2018:101f.).¹⁴ Similarly, the languages of Sumba typically have three series of plosives with two voiced series; either prenasalised and imploded, as in Kambara (Klamer 1998:10), or plain voiced and imploded, as in Laboya (Verdizade 2019:18). Closer to Timor, Dhao has three kinds of voiced stops: plain voiced /b d dʒ g/, implosives /ɸ ɖ ɖʒ g/, and affricates /bʃ dʒ/ (Grimes 2010:256, Balukh 2020:28).

These examples show that while the segmental inventory of _{PRM} differs from _{PMP}, it conforms well to the typology of the region. It is thus highly likely that the _{PRM} system arose through contact with typologically similar languages already present at the time _{PRM} arrived/arose in the Timor region. An investigation of the frequencies of different segments in different strata thus has the potential to shed light on the nature of this contact.

2.1 *PRM Segments by the Numbers*

All _{PRM} proto phonemes have at least some attestation in _{PMP} etyma (see § 4.2 for more details). However, some are disproportionately represented in words not known to be inherited from _{PMP}. This skewing indicates that the

14 One slight difference in the systems of Central Flores is that the implosive series is often optionally glottalised or pre-glottalised.

transformation of the PRM system to a regionally common system occurred partly through the adoption of words with these segments. On the other hand, some segments are disproportionately attested in PMP inheritances. This indicates that while words with new segments were acquired/borrowed by PRM, the “Austronesian inventory” was also retained.

The frequencies of PRM consonants according to whether they occur in known MP inheritances or not are summarised in Table 4.4. Consonants which have a higher representation in either category with statistical significance ($p < 0.003$) compared with the representation of that proto phoneme in the entire data set are shaded red, while figures which are suggestive of significance ($p < 0.01$) are shaded light blue. Consonants which are over-represented by at least 5% but without statistical significance are shaded grey. Statistical significance was calculated with a binomial distribution.¹⁵

In Table 4.4 we see that the segments *d, *ŋg, and *s are over-represented in words not known to be inherited from PMP, with statistical significance. Additionally, *p, *ʔ, *mb, *nd, and *r are over-represented in words not known to be inherited from PMP. These results suggest that these consonants are present in PRM partly through the introduction of new words from non-AN languages which had these consonants. This is explored in more detail in § 2.2 below.

As I discuss further in § 4.2, another source for PRM *d, *mb, *nd, and *ŋg is irregular/minority sound change. However, this is not the case for PRM *s which is a regular reflex of PMP *s. Thus, it may be surprising that PRM *s is not as well represented in MP inheritances. The reason for this is typological. PMP *s is not from alveolar [s], but instead is a reflex of Proto Austronesian *[ʃ] (Blust 2013:585f.) or *[tʰ] ~ *[θ] (Wolff 2010:32).¹⁶ That is, instances of PRM *s in MP inheritances are ultimately from a typologically rare segment. Thus, if instances of PRM *s in words not known to be inherited from MP descend from typologically more common alveolar [s], it would be expected to be more frequent in this stratum.

15 The binomial distribution describes how surprising the observed strata (\pm MP) for each PRM proto phoneme are assuming that all PRM proto phonemes should behave similarly in each stratum. A lower score means that it is very unlikely to see that combination of numbers arising randomly, while a higher score means that the strata for that PRM proto phoneme behave according to the general distribution of proto phonemes. Thus, 32/37 (= 86%) of PRM *b in PMP inheritances, where PMP inheritances contain 43% of all proto phonemes yields a binomial distribution of 0.0000006 while 98/228 (= 43%) of PRM *l in PMP inheritances yields a binomial distribution of 0.05.

16 The segment Blust takes to be Proto Austronesian *[ʃ] is traditionally transcribed <*s>. Proto Austronesian *[s] is transcribed <*S> which then became *h in PMP.

TABLE 4.4 PRM consonants by strata

*C	Total	+MP		-MP	
*p	18	6	33%	12	67%
*t	269	127	47%	142	53%
*k	200	78	39%	122	61%
*ʔ	28	8	29%	20	71%
*b	37	32	86%	5	14%
*d	39	23	59%	16	41%
*dʒ	2	2	100%	—	—
*ʙ	86	34	40%	52	60%
*d̪	102	27	26%	75	74%
*mb	80	27	34%	53	66%
*nd	46	13	28%	33	72%
*ŋg	69	14	20%	55	80%
*f	118	64	54%	54	46%
*s	240	80	33%	160	67%
*h	101	56	55%	45	45%
*m	139	72	52%	67	48%
*n	189	103	54%	86	46%
*ŋ	16	9	56%	7	44%
*l	228	98	43%	130	57%
*r	118	45	38%	73	62%
*w	2	2	100%	—	—
overall	2,127	920	43%	1,207	57%

We also find that certain proto phonemes are over-represented in MP inheritances; in particular PRM *b and *n are over-represented with statistical significance in MP inheritances. Additionally, PRM *f, *h, and *m are over-represented in MP inheritances to an extent which is suggestive of statistical significance. Plain-voiced *d and the velar nasal *ŋ are also over-represented, but without statistical significance. In §2.3 below I take a more detailed look at PRM *d and show that it may pattern with *b in being over-represented in MP inheritances.

The over-representation of PRM *n in MP inheritance is due to the usual merger of PMP *ŋ/*ñ > PRM *n. This merger occurs in 29 cases. If the number of instances of PRM *n in the MP stratum were reduced by 29 (from 103 to 74)

TABLE 4.5 PRM vowels by strata

*V	Total	+MP	-MP	+MP	-MP
*i	393	204	189	52%	48%
*e	393	147	246	37%	63%
*ə	53	26	27	49%	51%
*a	726	349	377	48%	52%
*o	295	63	232	21%	79%
*u	501	264	237	53%	47%
overall	2,361	1,053	1,308	45%	55%

48% of *n would occur in MP inheritances and it would no longer be as over-represented in MP inheritances.

Currently, there does not seem to be a clear explanation for the over-representation of PRM *f, *h, and *m in MP inheritances. It is worth noting that there are usual—though not completely regular (see §4.2)—sound changes from PMP which yield PRM *f and *h. These are PMP *b > PRM *f (60 instances) and PMP *p > PRM *h (39 instances). If these sound changes had not occurred, PRM *f and *h would be massively over-represented in words not known to be inherited from MP.

Apart from the consonants, PMP and PRM are also reconstructed with different vowel systems. PMP is reconstructed with four vowels: *i *ə *a *u, while six are required for PRM: *i *e *ə *a *o *u.

The representation of PRM vowels according to different strata is shown in Table 4.5. This table shows that PRM *e and *o are disproportionately represented in words not known to be inherited from PMP, while all other vowels, except *ə, are over-represented in words known to be inherited from PMP.

The over-representation of mid-vowels *e and *o in words not known to be inherited from PMP may be surprising given that there are (mostly) regular sound changes from PMP which produce each of these vowels in PRM: *au/*wa/*aw > *o, *ai/*ay/*ya > *e, and penultimate *ə > *e. Nonetheless, the data show that the presence of mid-vowels in PRM is mainly attributed to the non-AN strata. That is, the introduction of words from substrate languages with these vowels.¹⁷

17 PCEMP has been reconstructed with the mid-vowels *e and *o. However, of the 478 mid

TABLE 4.6 Implosives and prenasalised plosives #_

Env.	PRM	Total	+MP		-MP	
#_	*b	67	28	42%	39	58%
#_	*d	55	11	20%	44	80%
#_	*mb	49	17	35%	32	65%
#_	*nd	19	3	16%	16	84%
#_	*ŋg	33	6	18%	27	82%
#_	all C ^a	1,045	460	44%	585	56%

a These figures are for the total number of consonants in word initial positions, not just the total number of implosives and prenasalised plosives. This figure is needed to calculate the binomial distribution.

The reason that *i, *u, and to a lesser extent *a, are over-represented in MP inheritances is probably because PMP only had four vowels *i, *ə, *a, and *u, while the substrate language(s) with which PRM has had contact must also have had *e and *o. Thus, the vowels *i, *u, and *a are simply more frequent as a proportion of all vowels in PMP inheritances than they are in other words. MP inheritances are thus more likely to have these vowels than other words.

2.2 Implosion and Pre Nasalisation

In this section I take a more detailed look at the distribution of implosives and pre nasalised plosives. These segments give the PRM consonant system a distinctly different typological profile compared with PMP.

As discussed above, *d, *mb, *nd, and *ŋg are over-represented in words not known to be inherited from PMP, with this skewing being statistically significant for *d and *ŋg. This skewing increases further for *d, *nd, and *ŋg when we examine them according to word position. The frequency of implosives and prenasalised plosives in word initial position is summarised in Table 4.6.

The biggest difference in word initial position is the skewing of *nd towards words not known to be inherited from PMP, with an increase of 12%. In initial position the skewing approaches statistical significance ($p = 0.008$). There are

vowels which occur in words not known to be inherited from PMP, only 6% (31/478) belong to the Wallacean/PCEMP stratum. This indicates that most mid-vowels entered PRM through words that belonged to other, lower level, regional strata.

TABLE 4.7 Implosives and prenasalisation according to strata

	Total	MP		regi.		wTim	
*b	86	34	40%	23	27%	29	34%
*d	102	27	26%	51	50%	24	24%
*mb	80	27	34%	28	35%	25	31%
*nd	46	13	28%	13	28%	20	43%
*ŋg	69	14	20%	25	36%	30	43%
all *C	2,127	920	43%	636	30%	571	27%

only three words known to be inherited from PMP with initial *nd; *zaRum > *ndau ‘needle’, *dakah > *ndake ‘climb, ascend’, and *si-ia > *ndia ‘3SG’.

Further patterns for these segments are revealed when we take a more detailed look at the particular strata in which they occur. Table 4.7 shows the distribution of the implosives and prenasalised plosives in the regional and west Timor strata.

The implosive *d has twice the number of attestations in the regional stratum than it does in either the AN or west Timor stratum. Furthermore, 90% (46/51) of instances of *d in the regional stratum are found in words shared between PRM and other languages of the Lesser Sunda Islands, but currently not known to be more widely distributed—that is, they do *not* occur in words that could be inherited from PCEMP.

The prenasalised plosives, on the other hand, are over-represented in the west Timor stratum. Furthermore, of the prenasalised plosives in the west Timor stratum, 84% (63/75) occur in words which are restricted to Rote-Meto and not shared with Helong.

The more frequent occurrence of *d in the regional stratum and that of the prenasalised plosives in the west Timor stratum may indicate that implosion and prenasalisation developed at different points in the history of Rote-Meto. It may be that implosion developed at a higher node, such as Proto Timor-Babar, while prenasalisation developed at a lower node. However, the implosive *b is most well represented in the west Timor stratum, which may suggest later development of implosion in Rote-Meto. More bottom-up comparative work on other languages of the region may help us better understand the origins of these segments.

What we *can* say with confidence is that while some MP inheritances have developed implosives or prenasalised plosives at a lower level, the presence of

PRM *d, *nd, and *ŋg, in particular, is mainly due to the presence of the non-AN strata. That is, the introduction of words with these segments (or their precursors) into PRM, most likely from substrate languages.

The labials *b and *mb do not show the same skewing towards words not known to be inherited from MP compared with the other implosives and prenasalised plosives. This is partly because these segments are fairly well represented in MP inheritances due to an unconditioned split which affected PMP *b. This is discussed in more detail in § 4.2.

2.3 Plain Voiced Plosives

Table 4.4 shows that there is a difference in distribution between the PRM plain voiced plosives *b and *d. While PRM *b is poorly represented in words not known to be inherited from PMP, *d is fairly well-represented in such words.

However, under my current analysis, two separate correspondence sets attest PRM *d, with these correspondences resulting from an unconditioned split. The correspondence sets for PRM *d are shown in Table 4.8 (shaded grey), along with the reflexes of other PRM voiced coronals to show that neither set attesting *d can be conflated with another correspondence set.

Although I currently propose *d for this correspondence set, this remains a working hypothesis which I acknowledge may need to be revised as more data comes to light. As discussed in Edwards (2021:52–53), an alternate solution would be to posit a segment other than *d to account for the second set of reflexes, marked with a question mark in Table 4.8.¹⁸ We could propose that these reflexes are actually from one of the other PRM proto phonemes—thus shifting the unconditioned split from *d—or we could propose that these words attest another value, such as *d̥ or *nr.

If we separate the two correspondence sets attesting *d from each other, we find that the first pattern for *d is over-represented in MP inheritances with 19/23 (83%) examples, while the second pattern for *d is over-represented in words not known to be inherited from PMP with 12/16 (75%) examples. Furthermore, a binomial distribution shows that both these skewings are statistically significant at $p = 0.0001$ and $p = 0.007$ respectively.

If we exclude the second correspondence set which I have assigned to *d from our analysis, the behaviour of PRM *d is very similar to *b. Thus, it is fair

18 The second set for PRM *d has four examples from PMP: *daŋdaŋ > *dada ‘warm near a fire’, *duyuy > *dui ‘dugong’, *pandak > *mbada-k ‘short in height’, and *ŋadas > *ŋgadas ‘palate, gills’. Part of the evidence for reconstructing *d for this set comes from the fact it reflects PMP *d in these words.

TABLE 4.8 Reflexes of voiced coronals in PRM

PRM	#_		V_V					
	*l	*nd-	*-nd-	*dʒ	*d'	*r	*d	*d(?)
Proto Nuclear Rote	*l	*nd	*nd	*d'	*d'	*r	*d	*r
Tii	<i>l</i>	<i>nd</i>	<i>nd</i>	<i>d'</i>	<i>d'</i>	<i>r</i>	<i>d</i>	<i>r</i>
Lole	<i>l</i>	<i>nd</i>	<i>nd</i>	<i>d'</i>	<i>d'</i>	<i>l</i>	<i>d</i>	<i>l</i>
Ba'a, Termanu	<i>l</i>	<i>nd</i>	<i>n</i>	<i>d</i>	<i>d</i>	<i>l</i>	<i>d</i>	<i>l</i>
Bilbaa, Bokai	<i>l</i>	<i>l</i>	<i>n</i>	<i>d</i>	<i>d</i>	<i>l</i>	<i>d</i>	<i>l</i>
Landu	<i>l</i>	<i>nd</i>	<i>nd</i>	<i>d</i>	<i>d</i>	<i>r</i>	<i>d</i>	<i>r</i>
Rikou	<i>l</i>	<i>r</i>	<i>nd</i>	<i>d</i>	<i>d</i>	<i>r</i>	<i>d</i>	<i>r</i>
Oepao	<i>l</i>	<i>r</i>	<i>r</i>	<i>d</i>	<i>d</i>	<i>r</i>	<i>d</i>	<i>r</i>
Proto West RM	*l	*nd	*nd	*d	*d'	*r	*r	*d
Dela-Oenale	<i>l</i>	<i>nd</i>	<i>nd</i>	<i>r</i>	<i>d'</i>	<i>r</i>	<i>r</i>	<i>r</i>
Dengka	<i>l</i>	<i>nd</i>	<i>nd</i>	<i>l</i>	<i>d'</i>	<i>l</i>	<i>l</i>	<i>l</i>
Proto Meto	*n	*r	*r	*d	*d	*n	*n	*d
Ro'is Amarasi	<i>n</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>n</i>	<i>n</i>	<i>r</i>
Kotos Amarasi	<i>n</i>	<i>k</i>	<i>k</i>	<i>r</i>	<i>r</i>	<i>n</i>	<i>n</i>	<i>r</i>
Amanuban, Molo	<i>n</i>	<i>k</i>	<i>k</i>	<i>l</i>	<i>l</i>	<i>n</i>	<i>n</i>	<i>l</i>
Kusa-Manea	<i>n</i>	<i>k</i>	<i>k</i>	<i>r</i>	<i>r</i>	<i>n</i>	<i>n</i>	<i>r</i>
no.	228	19	27	2	102	118	23	16

to say that the series of plain-voiced plosives as a whole are disproportionately represented in MP inheritances.

Thus, while the PRM implosives and prenasalised plosives are partly due to the introduction of words from substrate languages, these languages appear not to have contributed plain voiced plosives to PRM. By extrapolation, this indicates that these substrate(s) did not have plain voiced plosives. Given the discussion in § 2.2 above which indicates that the pre-RM substrate(s) did have imploded and prenasalised plosives, this leads to the conclusion that these substrates probably had a segmental inventory with three series of stops: voiceless, prenasalised, and imploded.

3 Origins of the Lexicon

In this section I take a detailed look at the lexicon of PRM to determine what it could tell us about the history of Rote-Meto. Unlike the segmental inventory, which has been transformed according to regional norms, the lexicon is expected for an Austronesian language, with non-AN words mainly occurring in more borrowable domains.

As discussed in § 1.2, of the 1,173 PRM reconstructions 647 are not known to be inherited from MP. Of these 340, belong to the regional stratum and 307 belong to the west Timor stratum. Furthermore, as discussed in § 1.3, these reconstructions have one of two likely sources: language contact or coinage. The extreme rarity of coinage means that it can be considered a negligible source, except perhaps in the case of onomatopoeia and nursery language. There are 25 reconstructions in my database which are potentially onomatopoeic or nursery language. Of these, five are inherited from PMP,¹⁹ twelve belong to the regional stratum,²⁰ and eight belong to the west Timor stratum.²¹ These potential cases of coinage are excluded from the analysis throughout the remainder of this section. This leaves 628 reconstructions not known to be inherited from PMP which were probably acquired by language contact. These figures are summarised in Table 4.9.

In this section I investigate the likely borrowings from two perspectives: representation in basic vocabulary (§ 3.1), and distribution in different semantic spheres (§ 3.2). Unlike the segmental inventory, PRM has an expected lexical profile for an AN language. MP inheritances are better represented in basic vocabulary and semantic spheres more resistant to borrowing, while the other strata are more poorly represented in such areas.

Before proceeding with the discussion, it is worth stating that only three of the words not known to be inherited from PMP have known cognates in the non-AN Timor-Alor-Pantar family. This indicates that the non-AN strata were acquired from one or more extinct genealogical lineages.

19 PMP *guru(q) > PRM *ŋguru 'drone, growl', *kaka = *kaka 'older sibling', *kur(u) > *kuru₁ 'call chickens', *təktək > *teke 'gecko', and *uu = *uu 'oink'.

20 Proto Meto *baba-ʔ 'maternal uncle, paternal aunt', *buu 'blow, blowpipe', *fibi 'goat, sheep', *kaa₂ 'crow', *kae 'cockatoo', *koa₁ 'Friarbird', Proto Meto *kumu₁ 'wild dove', *meo₁ 'cat', *ŋgia 'parakeet', *poko 'plop', *roko 'rattle', and *tatə 'boy, older sibling'.

21 *boo₁ 'herd', *di₂ 'whinny', *koa₂ 'cry out', *kuu 'blow', *mee 'bleat', *mbuu 'sound, noise', *ŋguu 'howl, drone', and *tudui 'owl'.

TABLE 4.9 Potential coinages and borrowings in strata

Stratum	Total	Potential coinages	Likely borrowings
PMP	526	5	521
regional	340	12	328
west Timor	307	8	299

3.1 *Basic Vocabulary*

My examination of basic vocabulary used a 254-item wordlist. This wordlist was based on the 226-item Sulawesi Survey Word list used by Mead (1999), with the addition of 28 items not on that list from the 100 item Leipzig-Jakarta word-list given by Tadmor et al. (2012).

A PRM version of this list was compiled and the stratum to which each word belonged was recorded. When more than one reconstruction matched a concept on the wordlist, both were included.²² When no reconstruction matched a concept, that concept was excluded. Similarly, when a reconstruction occurred multiple times (e.g. due to polysemy) it was only counted once. As a result, 32 concepts matched two reconstructions and 44 concepts were not counted. This meant that the final list contained 242 reconstructions. Table 4.10 summarises the composition of basic vocabulary compared to the entire lexicon with regard to the three strata.

We see in Table 4.10 that over two thirds (69%) of basic vocabulary is inherited from PMP. This is much higher than the entire database where inheritances from PMP comprise a little less than half (45%) of PRM reconstructions. The increased ratio of PMP inheritances in basic vocabulary causes a drop among the words not known to be inherited from PMP. Both the regional and west Timor strata drop by 12% compared with their overall representation.

In sum, examination of basic vocabulary does not give us any new information about the origins of the strata. The strong representation of MP inheritances in basic vocabulary is unsurprising given that Rote-Meto is an MP subgroup. While the representation of the regional and west Timor strata in basic vocabulary is lower compared with their overall representation, they both drop by the same amount with this drop accounted for by the increased proportion of MP inheritances in basic vocabulary.

22 Thus, for instance, both PRM *mbana-k and *idu were included for 'nose'.

TABLE 4.10 Comparison of basic vocabulary and entire lexicon

	Basic		Entire lexicon	
AN	166	69%	521	45%
Regional	41	17%	328	29%
west Timor	35	14%	299	26%
	242		1,148	

3.2 *Semantic Fields*

An additional perspective on the composition of the PRM lexicon could be provided by semantic fields. I assigned each reconstruction in my database to one of eighteen semantic fields. These semantic fields and the number of reconstructions within each are summarised in Table 4.11. Fields are arranged from most borrowing-resistant to least borrowing-resistant, following the analysis of Tadmor et al. (2012:41f.) who determined borrowability on the basis of 41 languages from around the world.

The semantic fields in Tadmor et al. (2012) follow those used by Buck (1949) and Key and Comrie (2015), which I modified in some minor ways.²³ Although these semantic fields are problematic in several respects, they have two advantages. Firstly, they allowed assignment of words to semantic fields in a neutral way to avoid confirmation bias, and secondly the resistance to borrowing of each field has been tested by Tadmor et al. (2012:41f.).

Figures which are higher than expected in Table 4.11 with statistical significance compared with the entire lexicon are shaded red, while light blue shading is for figures that are suggestive. This statistical significance was calculated with a binomial distribution.²⁴ Based on the results for all semantic fields and strata it was decided that a score of less than 0.003 was significant, while a score of less than 0.01 was suggestive. The overall figures for several fields are too low to determine statistical significance.

23 The semantic fields in Buck (1949) and Key and Comrie (2015) were modified in the following ways: (1) *Basic actions and technology* was split, with *Technology* then combined with *Tools*. (2) *Warfare and hunting* was split, with *Hunting* combined with *Food and drink* while *Warfare* was combined with *Tools* (nearly all PRM reconstructions relating to warfare are weapons). (3) The semantic fields of *The house*, as well as *Clothing and grooming* (along with *Technology*, and *Weapons*) were combined with *Tools*. (4) The *Law* and *Religion and belief* semantic fields were combined with the *Social and political relations* field.

24 See footnote 15 for information on how the binomial distribution was calculated with respect to proto phonemes.

TABLE 4.11 PRM reconstructions by semantic fields

Semantic sphere		Total	AN	Regi.	wTim.	AN	Regi.	wTim.
1	Sense perception	61	34	11	16	56%	18%	26%
2	Spatial relations	89	40	26	23	45%	29%	26%
3	The body	153	90	35	28	59%	23%	18%
4	People (Kinship)	37	27	4	6	73%	11%	16%
5	Motion	93	30	35	28	32%	38%	30%
6	Physical world	69	36	17	16	52%	25%	23%
7	Emotions & values	22	10	6	6	45%	27%	27%
8	Quantity	25	16	4	5	64%	16%	20%
9	Speech & language	27	6	10	11	22%	37%	41%
10	Time	20	10	6	4	50%	30%	20%
11	Basic actions	86	33	25	28	38%	29%	33%
12	Cognition	17	4	8	5	24%	47%	29%
13	Animals	119	51	37	31	43%	31%	26%
14	Possession	17	4	6	7	24%	35%	41%
15	Food & drink	50	28	14	8	56%	28%	16%
16	Vegetation	167	69	43	55	41%	26%	33%
17	Social & political	29	9	8	12	31%	28%	41%
18	Tools	67	24	33	10	36%	49%	15%
Entire lexicon		1,148	521	328	299	45%	29%	26%

Examination of the lexical strata according to semantic spheres shows that the AN stratum is robustly attested in semantic spheres which are resistant to borrowing. With the exception of *Motion*, the top six most borrowing-resistant spheres are all either better represented than the overall lexicon or equally as well represented. In particular, the AN stratum has a greater proportion of terms in the *Body* and *People* semantic spheres than expected.

While the increased number of *People* terms occurs at the expense of both the regional and west Timor strata, the increase in *Body* part terms is mainly at the expense of the west Timor stratum alone. This might indicate that the west Timor stratum contains more words acquired through contact. Indeed, two of the three most borrowable spheres, *Vegetation* and *Social & political*, have a higher proportion of terms in the West Timor stratum.

The higher proportion of *Vegetation* terms is suggestive of statistical significance ($p = 0.009$). This points to acquisition of new flora terms in Timor. This is

further supported by the fact that of the 43 *Vegetation* terms in the regional stratum nearly half (20/43) occur in the Timor sub-stratum. Given that the homeland of Proto Austronesian is placed in Taiwan and thus in the biogeographical region of Sunda-land, it would be natural for words for the new flora encountered in Sahul-land to be borrowed from substrate languages already present in this area.

Finally, at 49% the proportion of *Tools* is higher than expected in the regional stratum (29% overall). This difference is statistically significant ($p = 0.0002$). Given that the tools sphere is the most borrowable semantic domain, this provides evidence that the regional stratum contains more loans compared to the other strata.

To summarise, examination of the semantic spheres indicates that vocabulary in the regional and west Timor strata have a high proportion of borrowed terms. These results are expected given that PRM is an AN family. Nonetheless, this contrasts with the segmental inventory which has been transformed due to contact with non-AN languages. I return to this apparent contradiction in § 4.2.3 and § 5.

4 Regularity of Sound Change

An examination of the regularity of sound change provides more information about the contact history of PRM. This includes regularity of sound change between PRM and its daughter languages (§ 4.1), as well as between PMP and PRM itself (§ 4.2).

MP inheritances and the regional stratum show about the same degree of irregularity between PRM and its daughters, while the west Timor stratum shows a higher degree of irregularity. This indicates that the regional stratum had mostly been acquired before the formation of PRM proper, while parts of the west Timor-stratum were acquired after the break-up of the proto language.

Examining the regularity of sound change between PMP and PRM reveals a large number of unconditioned splits. Multiple factors have contributed to these splits including sound change in progress, contact with other AN languages, and contact with pre-AN substrate(s).

4.1 *Regularity of Sound Change from PRM*

In this section I examine the regularity of sound change between PRM and its daughter languages in order to gain further insights on the origins of the strata of vocabulary in PRM. The assumption is that a cognate set which spread by bor-

TABLE 4.12 Regular and irregular sound changes

	Reg.	Reg.	Irr.	Reg.	Reg.	Irr.
*gloss	‘moon’	‘fur’	‘ringworm’	‘banana’	‘dry season’	‘cut’
PMP	*bulan	*bulu	*buqəni	*punti		
PRM	*bulan	*bulu-k	*buni	*hundi	*fandu	*fandi ^a
Rikou	<i>bula-ʔ</i>	<i>bulu-ʔ</i>	<i>bu~buni</i>	<i>hundi</i>	<i>fandu-ʔ</i>	
Bilbaa	<i>bula-ʔ</i>	<i>bulu-ʔ</i>	<i>buni</i>	<i>huni</i>	<i>fanu-ʔ</i>	
Korbafo	<i>bula-ʔ</i>	<i>bulu-ʔ</i>	<i>bu~buni</i>	<i>huni</i>	<i>fanu-ʔ</i>	<i>fani</i>
Termanu	<i>bula-k</i>	<i>bulu-k</i>	<i>bu~buni</i>	<i>huni</i>	<i>fanu-k</i>	<i>fani</i>
Tii	<i>bula-k</i>	<i>bulu-k</i>	<i>bu~buni</i>	<i>hundi</i>	<i>fandu-k</i>	
Oenale	<i>fulan</i>	<i>fulu-ʔ</i>	<i>ɸuni</i>	<i>hundi</i>	<i>fandu-ʔ</i>	
Dengka	<i>fula-ʔ</i>	<i>fulu-ʔ</i>	<i>buni</i>	<i>hundi</i>	<i>fandu-ʔ</i>	<i>fandi</i>
Kotos	<i>funan</i>	<i>funu-f</i>	<i>hune</i>	<i>uki</i>	<i>fauknais^b</i>	<i>fani</i>
Molo	<i>funan</i>	<i>funuʔ</i>	<i>hune</i>	<i>uki</i>	<i>fauknais</i>	<i>fani</i>

- a Cognates in languages such as Bima *fati*, *manti* ‘chop’ favour PRM *nd rather than *n.
- b Meto reflexes of *fandu are historic compounds with the first element showing *CV → VC metathesis. Final *nais* is of unknown origin.

rowing and diffusion after the break-up of the proto language is more likely to show irregularities than a cognate set which has been acquired through inheritance from a single etymon.

Table 4.12 shows six cognate sets in the Rote-Meto languages. Three attest PRM *b, and three attest PRM *nd. The first two cognate sets of each show the regular, or most common, correspondences in each daughter language. The third set has irregular correspondences, indicated with grey shading. Reflexes of *buni ‘ringworm’ have irregular *b > ɸ/b in West Rote and irregular *b > h in Meto. In both cases we expect *b > f, as in reflexes of *bulan ‘moon’ and *bulu-k ‘fur’. Reflexes of *fandi ‘cut’ have irregular *nd > n in Meto, where we expect *nd > k, as shown in reflexes of *hundi ‘banana’ and *fandu ‘dry season’.

Such irregular correspondence sets could be because some words are due to diffusion after the break-up of the group, e.g. Meto *hune* ‘ringworm’—while ultimately cognate with the Rote words and from PMP *buqəni—is not a direct inheritance from PRM *buni, but rather a borrowing from another AN language. It could be from a language in which *b > h is regular, or it could be due to interference in the process of borrowing. However, it is also possible that the Meto words have simply undergone irregular *b > h.

In the absence of any identifiable source language, both hypotheses—borrowing or irregular sound change—are equally ad-hoc.²⁵ As a result, I included such words in my database, made a PRM reconstruction and tracked the irregular sound changes required to derive the modern day words from the PRM reconstruction.

When we examine the extent to which PRM reconstructions show irregularities according to strata,²⁶ we find 19% (99/526) of inheritances from PMP have at least one irregularity in the daughter languages, 20% (68/340) of words in the regional stratum have irregularities, and 26% (81/307) of reconstructions in the west Timor stratum have irregularities. If it is correct that a cognate set which spreads after the break-up of the proto language is more likely to display irregular correspondences, then the larger number of irregularities in the west Timor stratum indicates that it contains more such words than the other strata.

The percentage of irregularities in the regional stratum (20%) is almost identical to the AN stratum (19%). This may indicate that most words in this stratum had already been acquired by the time PRM was formed, after which they were inherited regularly into daughter languages. Thus, regularity of sound change from PRM to its daughters probably points to at least two periods of contact in the history of the Rote-Meto languages; one prior to the formation of PRM and one after the break-up of the proto language.

4.2 *Regularity of Sound Change from PMP to PRM*

All PRM segments occur to some extent in MP inheritances. This includes implosives and prenasalised plosives which give the PRM segmental inventory a different typological profile compared with PMP. However, the presence of these segments in MP inheritances is, in general, not due to regular sound changes applying to PMP, but rather is part of a large scale incidence of unconditioned splits which have affected PMP inheritances.²⁷

Seven PMP consonants have (mostly) regular reflexes in PRM: *t = *t, *m = *m, *n = *n, *ñ > *n, *l = *l, *s = *s, and *h > ∅. Of these, the reflexes are completely regular for *m = *m (59 instances), *ñ > *n (8 instances), *n = *n (55 instances), and *h > ∅ (36 instances). Other consonants have occasional irregularities. PMP *t = *t in 117/121 cases (with two cases of *t > *nd and two of *t >

25 Words for which a source language has been identified do not feature in the analysis.

26 Sporadic changes, such as consonant metathesis (e.g. PRM ***man**eu > Proto West Rote-Meto ***nam**eu ‘bright, light’), were not treated as irregular when calculating irregularity.

27 Throughout this section my discussion is limited to consonants in initial and medial position. Word-final consonants in PMP etyma show different sound changes. See Edwards (2018b: 77–80) and Edwards (2021:55) for discussion.

TABLE 4.13 PMP consonants undergoing unconditioned splits^a

Env. ^b	PMP	PRM	No.	%	PMP	PRM	No.	%
	*p	*h	39	74%	*z [dʒ]	*d	7	64%
		*p	6	11%		*d	2	18%
		*b	3	6%		*nd	2	18%
		∅	3	6%	*j [gʲ]	*d	7	44%
#_	*k-	*k-	31	89%	*d	4	23%	
		*h-	2	6%	*r	3	19%	
		∅	2	6%	*dʒ	2	13%	
V_V	*-k-	*k-	37	77%	*g	*k	6	50%
		*-ʔ-	4	8%	*ŋg	6	50%	
		∅	4	8%	*q	∅	45	80%
#_	*b-	*f-	36	34%	*h	8	14%	
		*b-	29	27%	*ŋ	*n	21	66%
		*b-	25	23%	*ŋ	8	25%	
		*mb-	14	13%	*ŋg	4	9%	
V_V	*-b-	*f-	24	71%	*R [r~R]	∅	42	81%
		*-b-	5	15%	*r	9	17%	
		*-mb-	3	9%	*r [r]	*r	9	75%
		*-b-	2	6%	∅	3	25%	
#_	*d-	*d-	16	57%	*wa	*o	9	60%
		*d-	7	25%	*fa	3	20%	
		*r-	4	14%	*wa	2	13%	
V_V	*-d-	*-r-	16	76%				
		*-d-	3	14%				

a Reflexes with a single example and/or which represent less than 5% of all instances are excluded from this table. Thus, not all percentages add up to 100%.

b #_ is word-initial and foot-initial position. Where no environment is given, the split occurs in all word positions.

*d), PMP *l = *l in 97/102 instances (with three cases of *l > *r and two cases of *l > *n), and PMP *s = *s in 60/62 instances (with one case of *s > *nd and one of *s > *d).

All other consonants undergo a split in PRM. The reflexes and their frequencies are summarised in Table 4.13. For some segments conditioning environments play a role in determining the frequency with which certain reflexes occur. For other segments, these splits are completely unconditioned. The only firmly identified conditioning environment is *w > ∅ before or after *i.

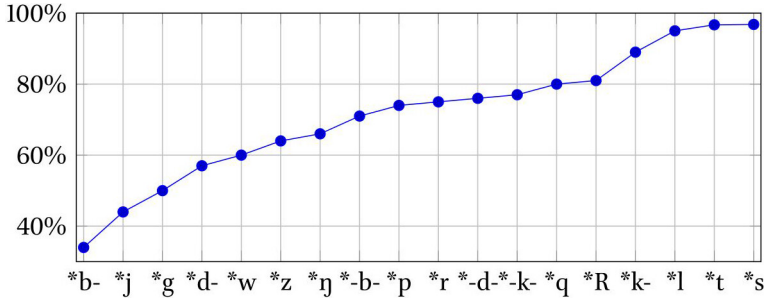


FIGURE 4.4 Frequency of majority reflexes

If we take the most frequent reflex of a PMP segment to be the “regular” reflex, we find that 64% (335/526) of PMP inheritances in PRM are regular, meaning that over a third (191/526 = 36%) of inheritances have at least one irregular sound change.

However, the binary distinction between “regular” and “irregular” is not an adequate description of the PRM data. Instead, there are degrees of regularity. There is a fairly steady cline between the most regular consonants, such as *t or *l, where one reflex predominates, to consonants such as initial *b-, where the most common reflex only occurs in about a third of cases. This cline is shown in Figure 4.4, which graphs the frequency of the most common reflexes among PMP consonants which are not 100% regular. For this reason, I refer to the most common reflex as the *majority* reflex.

None of the implosives and prenasalised stops are the majority reflex of any PMP consonant, with the exception of *z/*j > *d and, arguably, *g > *ŋg. Nonetheless, even though *d is the majority reflex of *z [dʒ] and *j [gʲ], instances of *z/*j > *d only account for 14 cases of *d, which, in turn, represent only 14% (14/102) of all cases of *d.²⁸ As discussed in § 2 (particularly § 2.2), most cases of *d occur in words not known to be inherited from PMP.²⁹

Another source of the PRM prenasalised stops are PMP nasal-stop clusters. However, even the nasal-stop clusters show unconditioned splits. Their reflexes are shown in Table 4.14. A prenasalised plosive is the majority (or only) reflex of PMP *mb, *mp, *ŋk, and *ŋd—though, with the exception of *mb, the absolute

28 There are 27 instances of PRM *d inherited from PMP (Table 4.4, page 114). Apart from the 14 instances of PMP *z/*j > PRM *d, there are 10 instances of PMP *d > PRM *d (Table 4.13), two of PMP *nt/*nd > PRM *d (Table 4.14), and one of PMP *s > PRM *d.

29 It is, of course, also possible that instances of *d in words not known to be inherited from PMP are from earlier **z [dʒ] or **j [gʲ] acquired before the formation of PRM.

TABLE 4.14 PMP nasal-stop clusters

PMP	PRM	No.	PMP	PRM	No.
*mb	*mb	7	*nt	*t	4
	*f	1		*nd	3
*mp	*mb	1		*ɖ	1
*mg	*m	4	*nd	*d	1
	*ʔ	1		*ɖ	1
*ŋk	*ŋg	3		*nd	1
	*k	2		*n	1
*ŋd	*nd	2		*r	1
	*d	1			

numbers are so low that it is hard to be confident that this would be maintained if more reflexes of these clusters were identified in the Rote-Meto languages.

The large number of unconditioned splits between PMP and PRM presents a methodological challenge to the application of the comparative method and naturally raises the question of how these splits are explained. In the following sections I explore three scenarios which may help explain these unconditioned splits:

1. sound change in progress
2. multiple Austronesian strata
3. contact with substrate languages

Each of these scenarios has probably contributed to the unconditioned splits. It must be emphasised that there is not a single unitary solution which can account for all the data. Not only have different scenarios played different roles, they have probably operated to different extents at different points in the history of Rote-Meto. Similarly, different splits may have different origins, and a single split may have multiple causes.

4.2.1 Sound Change in Progress

Part of the likely explanation for the unconditioned splits in PRM is that some splits probably represent incomplete sound changes which had not fully diffused through the lexicon by the time of the break-up of PRM. In this section I discuss two unconditioned splits which appear to be, partly, the result of incomplete sound changes. These are the splits affecting PMP *b and PMP *wa.

The split of PMP *b > PRM *f, *b is one likely example of an incomplete sound change. As summarised in Table 4.13, PMP *b undergoes a four-way uncon-

TABLE 4.15 Reflexes of PMP *b in PRM

Initial *b-			Intervocalic *-b-		
*f-	36	34 %	*-f-	24	71 %
*b-	29	27 %	*-b-	5	15 %
*ḃ-	25	23 %	*-mb-	3	9 %
*mb-	14	13 %	*-b-	2	6 %
total	104			34	

TABLE 4.16 Examples of PMP *b > PRM *f, *b, *ḃ, *mb^a

PMP	PRM	PRM gloss	PMP	PRM	PRM gloss
*bukij	*fui	'wild'	*təbuh	*tefu	'sugarcane'
*bujəq	*fudzə	'foam'	*tuba	*tufa	<i>'Derris elliptica'</i>
*bətaw	*feto-k	'man's sister'	*qubi	*ufi	'wild tuber'
*biRaq	*fia	'wild tuber'	*babuy	*bafi	'pig'
*bulan	*bulan	'node, joint'	*qabu	*afu	'ash, dust'
*bubu	*bufu	'rear'	*balabaw	*ka-lafo	'mouse, rat'
*batu	*batu	'ulcer'	*bukbuk	*ka-fufu-k	'weevil'
*buku	*ḃuku-k	'moon'	*babaw	*bafu	'above'
*buRit	*ḃuit	'fish trap'	*təbiq	*teḃi	'crumble'
*bisul	*ḃisu	'stone'	*libut	*liḃu	'swarm'
*buliR	*mbule-k	'grain head'	*bubuḃ	*ka-fumbu-k	'crown of head'
*buRuk	*mburuk	'rotten'	*qibaw	*kibo	'edible shellfish'

a The number of examples given here corresponds roughly to the overall frequency of each reflex, as summarised in Table 4.15.

ditioned split between PRM *f, *b, *ḃ, and *mb. The reflexes are repeated in Table 4.15. and examples of each change are given in Table 4.16.

It is worth re-emphasising this split is unconditioned. The only role conditioning environments play is in the frequency of reflexes, with PMP *b > PRM *f more common in intervocalic position.³⁰ Indeed, the first piece of evidence

³⁰ The presence or absence of prefixes also does not affect this split. In a small number of

that PMP *b > PRM *f was a change in progress which had not fully diffused through the lexicon by the break-up of PRM comes from these different frequencies.

Intervocalic *-b- > *-f- occurs in 71% (24/34) of cases, while word-initial *b- > *f- occurs in 34% of cases (36/107). These different frequencies can be explained if PMP *b > PRM *f was a change in progress. Intervocalic position is more susceptible to lenition, and *b > *f probably began in this position earlier and had more time to diffuse through in the lexicon, thus resulting in more cases of *b > *f /V_V.

Secondly, Proto West Rote-Meto underwent a “second round” of *b > *f, which affected most instances of *b which had not yet undergone change. There are at least a dozen examples in my database (see Table 4.12 for two examples). This can also be explained by proposing that *b > *f was a change in progress in PRM which continued in West Rote-Meto but was halted in Nuclear Rote after the split between these two branches.

A similar scenario may help explain the split affecting PMP *wa. PMP *wa shows a three way unconditioned split. Again, like the split affecting PMP *b, the split affecting PMP *wa is, essentially, unconditioned³¹ with the outcomes being either PRM *o (nine examples), *wa (two examples), or *fa (three examples). Two examples each of these reflexes in word initial position are given in Table 4.17.

Note that PRM *wa then undergoes a similar split in daughter languages. PRM *wa > *fa in Nuclear Rote but PRM *wa > *o in West Rote-Meto. An additional factor regarding PMP *wa is that Helong—the nearest neighbour of Rote-Meto—has *w > *f (with subsequent *f > p in Semau Helong). Examples include PMP *huaji > **waji > Funai Helong *falin*, PMP *walu > Funai Helong *falu*, and PMP *hawak ‘waist’ > Funai Helong *afa* ‘body, self’ (compare PRM *ao-k ‘body’). The change of PMP *w > *f is much more regular in Helong than in Rote-Meto, and it seems that the change of *wa > *fa in Rote-Meto is a contact-induced change which happened when Rote-Meto came into contact with Helong.

The changes affecting PMP *wa appear to have multiple sources. The evidence indicates that PMP *wa > PRM *o began first, but was not complete when

cases a PMP prefix has become part of the root in PRM. In such cases the historically stem initial consonant is counted as word medial. Thus, for instance, PMP *ma-buhək > PRM *mafu ‘drunk’ and PMP *bañən > PRM *kesufani ‘sneeze’ were both counted as instances of medial PMP *b > PRM *f.

31 The only conditioning affecting PMP *w is that it is lost before *i.

TABLE 4.17 PMP *wa > PRM *o, *wa, fa

*gloss	‘water’	‘root’	‘bee’	‘ySi’	‘eight’	‘day’
PMP	*wahiR	*wakaR	*wani	**waji ^a	*walu	*waRi ^b
PRM	*oe	*oka-k	*wani	*wadi-k	*falu	*fai
PnRote	*oe	*oka-k	*fani	*fadi-k	*falu	*fai
Tii	<i>oe</i>	<i>oka-k</i>	<i>fani</i>	<i>fadi-k</i>	<i>falu</i>	<i>fai</i>
Termanu	<i>oe</i>	<i>oka-k</i>	<i>fani</i>	<i>fadi-k</i>	<i>falu</i>	<i>fai</i>
Rikou	<i>oe</i>	<i>oka-?</i>	<i>fani</i>	<i>fadi-?</i>	<i>falu</i>	<i>fai</i>
PwRM	*oe	*oka-?	*oni	*odi-?	*falu	*fai
Oenale	<i>oe</i>	<i>?oka-?</i>	<i>oni</i>	<i>?odi-?</i>	<i>falu</i>	<i>fai</i>
Dengka	<i>oe</i>	<i>?oka-?</i>	<i>oni</i>	<i>?odi-?</i>	<i>falu</i>	<i>fai</i>
Kotos	<i>oe</i>		<i>oni</i>	<i>ori-f</i>	<i>fanu</i>	<i>fai</i>
Molo	<i>oe</i>		<i>oni</i>	<i>oli-f</i>	<i>fanu</i>	<i>fai</i>

- a Pre-RM **waji is from PMP *huaji with *hua > **wa (widely attested; e.g. Welaun *wali-n*, Buru *wai*). It is glossed by Blust and Trussel (2020) as ‘same sex younger sibling’. and by Wolff (2010:738) as ‘younger sibling’.
- b In addition to PMP *waRi > PRM *fai ‘day, time’, with *wa > *fa, PRM also has *hoi ‘dry in sun’ from PMP *pa-waRi with *wa > *o.

PRM came into contact with pre-Helong. The change *w > *f then started to spread from Helong into Rote-Meto, but affected the two Rote-Meto subgroups to different degrees. Before this change had fully spread throughout Rote-Meto, Proto West Rote-Meto continued earlier *wa > *o. The change of *w > *f then continued to spread and affected all remaining instances of *w.

The changes affecting PMP *wa illustrate well that no single scenario necessarily accounts for all the unconditioned splits we see between PMP and PRM; a change in progress accounts for *wa > *o, while language contact probably explains *wa > *fa.

4.2.2 Contact with Other Austronesian Languages

Another possible explanation for some of the unconditioned splits in Rote-Meto is contact with other AN languages. In particular, it might be proposed that Rote-Meto contains multiple AN strata; an inherited stratum, with one set of correspondences, and a borrowed stratum with another set of correspondences. This scenario has been proposed to account for apparent unconditioned splits in several other AN languages including: Ngaju Dayak (Dempwolff 1922, Dyen 1956), Rotuman (Biggs 1965), and Tiruray (Blust 1992).

Unlike these languages, I have been unable to identify any AN languages which could be a systematic source of one of the sound correspondences seen in Rote-Meto. While contact with other AN languages has undoubtedly played a role in the history of Rote-Meto (as proposed for *wa > *fa above), Rote-Meto does not appear to have multiple AN strata in the same way as has been proposed for Ngaju Dayak, Rotuman, or Tiruray.

Nonetheless, some words with irregular sound correspondences are suspected loans from intermediate AN languages. Two examples are: PMP *zaRum > PRM *ndau ‘needle’ (minority *z > *nd) and PMP *lɔpaw > *lopo ‘shelter, hut’, (minority *p = *p and irregular *ə > *o).³² When we examine MP inheritances according to semantic fields (see § 3.2), we find that words in the *Tools* field—the most borrowable field—have more minority sound changes than other fields. While 39% of all PMP inheritances have a minority sound change, 63% (15/24) of all *Tools* do. A binomial distribution shows that this is suggestive of statistical significance ($p = 0.006$).

Thus, while we cannot currently identify a systematic stratum of borrowed AN vocabulary in Rote-Meto, it would appear that borrowing and contact with intermediate AN languages has played some role in the presence of unconditioned splits. Such contact may have occurred at multiple times in the history of Rote-Meto.

4.2.3 Contact with Substrate(s)

In § 4.2 I showed that certain PRM segments are disproportionately represented in words not known to be inherited from PMP. I further proposed that this is because PRM has acquired many words containing these segments from non-AN substrate(s). In this section, I investigate the role language contact with substrate(s) may have played in producing the unconditioned splits seen between PMP and PRM.

In § 2.2 we saw that the implosive *ɗ and the prenasalised plosives (in particular *nd and *ŋg) are more frequent in words not known to be inherited from PMP. On the other hand, in § 2.3 we saw that the PRM plain-voiced stops *b and *d were more frequent in inheritances from PMP, but mostly lacking in words not known to be inherited from PMP. This indicates that the PRM segmental inventory is a combination of two historically separate systems: an “Austronesian system” (4), which contrasted voiceless and plain-voiced plosives, and a

32 PRM *lopo ‘shelter, hut’ (or some of its reflexes) may ultimately be from Sanskrit *maṇḍapa* [maṇḍapa] ‘temporary shed, pavilion’. Malay has *pendapa* ~ *pendopo* ‘large open pavilion-like veranda’.

locally present “regional system” (5), with three series of plosives: voiceless, implosive and prenasalised.

(4)	Austronesian system	(5)	Regional system
	p t k		p t k
	b d g		ɓ ɗ
			mb nd ŋg

As discussed in § 2, the regional system given in (5) is still present in several of the languages of western Rote, including Dela-Oenale, Tii, and Lole. It is also very similar to the system seen in several languages of Sumba, such as Kampera (Klamer 1998:10).

Some of the unconditioned splits that have occurred between PMP and PRM are probably due to adaptation of the incoming AN system to the regional system. While the exact mechanisms of how this occurred are difficult to discern, I discuss below two possibilities.

Firstly, the unconditioned splits may be due to words from pre-Rote-Meto being borrowed into substrate languages, along with assimilation of plain-voiced plosives to their nearest phonological targets. This process still occurs in the languages of western Rote, whereby non-native plain-voiced /b/ and /d/ are assimilated as implosives and /g/ is assimilated as a prenasalised plosive. Examples of loanwords that have entered Lole via Malay are given in Table 4.18 to illustrate.

A similar process may have occurred at the initial stage of contact between pre-Rote-Meto and substrate languages in western Timor. Thus, for instance, AN /b/ may have been assimilated in loan-words in substrate language(s) as /ɓ/ and/or /mb/, with these words then transferred back into PRM, perhaps by re-borrowing before the extinction of these substrate language(s).

The second way that contact with substrate(s) may have contributed to the unconditioned splits seen in PRM is as the result of language shift from substrate languages with the regional system. Under this scenario, speakers of the substrate language learnt pre-Rote-Meto, but did so with a “regional accent”. This would be akin to some varieties of Indian English in which the “native” dental fricatives /θ/ and /ð/ are dental plosives /t̪h/ and /d̪h/, and the alveolar plosives /t/ and /d/ tend to be retroflex /t̪/ and /d̪/ (Gargesh 2008:237–238).

This scenario may help explain the difference between the largely “Austronesian lexicon” (§ 3) and “non-Austronesian segmental inventory” (§ 2). If PRM is partly a result of language shift we may expect this result. This indeed is the case for some varieties of Indian English which have a largely Germanic lexicon, but south Asian segmental inventory. The difference between the two

TABLE 4.18 Lole assimilation of voiced plosives

Malay			Lole	Gloss
<i>berias</i>	/b̄arias/	→	<i>ɓalias</i>	'make-up'
<i>bangku</i>	/baŋku/	→	<i>ɓaygu</i>	'bench'
<i>bemo</i>	/bemo/	→	<i>ɓemo</i>	'mini-van'
<i>damai</i>	/dame/	→	<i>dame</i>	'peace'
<i>dapur</i>	/dapur/	→	<i>dapu</i>	'kitchen'
<i>dokter</i>	/dokter/	→	<i>ɗotel</i>	'doctor'
<i>gaji</i>	/gadʒi/	→	<i>ŋgadi</i>	'wage'
<i>gelas</i>	/gəlas/	→	<i>ŋgalaas</i>	'glass'
<i>ganti</i>	/ganti/	→	<i>ŋgati</i>	'exchange'

situations is that the non-Germanic “substrate languages” which have provided the segmental inventory of Indian English are still alive, while the non-AN substrate languages which may have provided the segmental inventory of PRM have become extinct.

However, if one event of language shift from a single substrate language were the only factor at play, we would probably expect more regular sound changes between PMP and PRM, such as *b > *ɓ, or *g > *ŋg. The fact that some PMP phonemes undergo multiple splits indicates that no single unitary contact scenario is probably sufficient. There may have been different kinds of contact with different substrate languages at different points in time, and this contact would have combined with sound changes in progress and contact with other AN languages to produce the complex picture we see in the history of PRM.

5 Conclusions

Detailed examination of the Rote-Meto language family produces a complex picture pointing to several kinds of contact. In the current state of our knowledge, nearly half of the reconstructed PRM lexicon is of unknown origin. Some of this lexicon has cognates in other regional languages, while some is limited to west Timor. Almost none of this vocabulary can yet be linked to non-AN languages present in the region, and thus it probably came from extinct pre-AN languages.

Examination of the segmental inventory shows that the inherited AN system with two series of plosives has been adapted to a regionally common system

with four series of plosives; voiceless, plain voiced, implosive, and prenasalised. The representation of proto phonemes in different strata indicates that the implosive *ɖ and the prenasalised plosives mainly entered PRM through the adoption of substrate vocabulary.

The lexicon presents a different picture. While a little over half of the lexicon cannot be traced back to PMP, the MP inheritances are robustly attested in basic vocabulary and semantic spheres which are most resistant to borrowing. They thus are the result of “normal” inheritance.

Regularity of sound change between PRM and its daughters reveals a slight difference between the non-AN strata. The west Timor stratum shows more irregular sound changes than other strata, indicating that this stratum contains more words distributed by borrowing after the break-up of PRM. The regional stratum, on the contrary, shows almost the same amount of irregularity as the inherited MP stratum. This probably points to at least two periods of contact: one before the formation of PRM, and one during and after the PRM period.

There are also a large number of unconditioned splits which have occurred between PMP and PRM. No single explanation accounts for all of these splits; some are cases of sound changes which were in progress but incomplete during the break-up of PRM, some are the result of contact with other AN languages, and some are probably a result of contact with substrate languages. Indeed, in the case of *w there is evidence that the split is a result of both sound change in progress *and* contact.

What does all this mean for the investigation of language history in the wider Wallacea region? Firstly, we must be wary of explanations which try to explain the data in a single way, or as a result of a single contact event. The Rote-Meto data points to multiple different kinds of contact, at different historical points. Even with more than 1,000 PRM reconstructions it is difficult to discern exactly what kinds of contact this language family has undergone. How much less can we say for languages and families for which much more limited data is available?

Secondly, the PRM data shows the importance of multiple lines of investigation. Examination of the segmental inventory paints quite a different picture to that of the lexicon, with regularity of sound change further refining the results we get from these two domains. It would be a natural next step to investigate the morphology and syntax of Rote-Meto to see in what further ways we can deepen our understanding of the history of this family.

Finally, we must not underestimate the role of, now extinct, substrate languages in this region. Rote-Meto has undergone significant contact effects with substrate languages which have transformed the segmental inventory and introduced a large amount of vocabulary. Other analysts have similarly pro-

posed contact to account for grammatical properties of languages in this region (Schapper and Hammarström 2013, Moro 2018, Fricke 2019, Moro and Fricke 2020). The investigation of Rote-Meto builds upon this work and shows that it is not only morpho-syntactic properties, but also phonology and lexicon which have been affected by contact.

Our understanding of language history in the Wallacea region is still in its infancy. This paper is one small contribution towards understanding the history of this region, but many questions remain to be answered, even regarding the Rote-Meto languages. As discussed in §1.1, Rote-Meto is part of a larger Timor-Babar language family, and a detailed bottom-up reconstruction of other branches of this family promises to answer some of the unresolved questions on the history of Rote-Meto, as well as providing more insights in the language history of the greater Timor region.

Acknowledgements

I would like to thank Charles Grimes, Kirsten Culhane, and two anonymous reviewers for comments on this paper which have greatly improved the structure, clarity and analysis. I would also like to thank Gereon Kaiping for help with the statistics. The interpretation of the data and all errors or deficiencies remain my own.

References

- Arka, I Wayan, Jeladu Kosmas, and I Nyoman Suparsa. 2007. *Bahasa Rongga: Tata-bahasa acuan ringkas*. Jakarta: Penerbit Universitas Atma Jaya (PUAJ).
- Baird, Louise. 2002. A grammar of Kéo: An Austronesian language of East Nusantara. Doctoral Dissertation, Australian National University.
- Balle, Misriani, and Stuart Cameron. 2014. Helong dictionary. Unpublished Toolbox file.
- Balukh, Jermy. 2020. A grammar of Dhao: An endangered Austronesian language in eastern Indonesia. Doctoral Dissertation, Leiden University.
- Biggs, Bruce. 1965. Direct and indirect inheritance in Rotuman. *Lingua* 14:383–445.
- Blust, Robert. 1992. On speech strata in Tiruray. In *Papers in Austronesian linguistics*, no. 2, ed. Malcolm Ross, 1–52. Pacific Linguistics: Canberra.
- Blust, Robert. 2013. *The Austronesian languages*. Canberra: Asia-Pacific Linguistics, revised edition.
- Blust, Robert, and Stephen Trussel. 2020. The Austronesian comparative dictionary. URL <http://www.trussel2.com/ACD/>.

- Brinton, Laurel J., and Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University of Chicago Press.
- Christensen, John. in process. Kisar dictionary. Unpublished Toolbox file.
- Dempwolff, Otto. 1922. Entstehung von Nasalen und Nasalverbindungen im Ngaju (Dajak). *Zeitschrift für Eingeborenen Sprachen* 13:161–205.
- Donohue, Mark, Rebecca Hetherington, James McElvenny, and Virginia Dawson. 2013. World phonotactics database. URL <http://phonotactics.anu.edu.au/index.php>, accessed 22/04/2018.
- Dyen, Isidore. 1956. The Ngaju-Dayak 'old speech stratum'. *Language* 32:83–87.
- Edwards, Owen. 2018a. Parallel histories in Rote-Meto. *Oceanic Linguistics* 57:359–409. URL <http://hdl.handle.net/1887/67592>.
- Edwards, Owen. 2018b. Top-down historical phonology of Rote-Meto. *Journal of the Southeast Asian Linguistics Society* 11:63–90. URL <http://hdl.handle.net/10524/52421>.
- Edwards, Owen. 2019. Reintroducing Welaun. *Oceanic Linguistics* 58:31–58. URL <http://hdl.handle.net/1887/79038>.
- Edwards, Owen. 2021. *Rote-Meto comparative dictionary*. Canberra: ANU press. URL <http://doi.org/10.22459/RMCD.2021>.
- Elias, Alexander. 2018. Lio and the Central Flores languages. Master's thesis, Leiden University.
- Fox, James J. 2016. *Master poets, ritual masters: the art of oral composition among the Rotenese of eastern Indonesia*. Canberra: ANU Press.
- Fricke, Hanna. 2019. Traces of language contact: the Flores-Lembata languages in eastern Indonesia. Doctoral Dissertation, Leiden University. URL <http://hdl.handle.net/1887/80399>.
- Gargesh, Ravinder. 2008. Indian English: phonology. In *Varieties of English: Africa, south and southeast Asia*, ed. Rajend Mesthrie, volume 4, 231–243. Berlin and New York: Mouton de Gruyter.
- Grimes, Charles E. 2010. Hawu and Dhao in eastern indonesia: revisiting their relationship. In *East Nusantara: typological and areal analysis*, ed. Michael Ewing and Marian Klamer, 251–280. Canberra: Pacific Linguistics.
- Grimes, Charles E., Bernadus Lado, Thomas Ly, and Simon Tari. 2008. *Lii Hawu (Sabu) online dictionary*. Kupang: UBB-GMIT. URL <http://ubb.or.id/download/kamus-lii-hawu-lexpro/>.
- de Josselin de Jong, J.P.B. 1947. *Studies in Indonesian culture 11: The community of Erai (Wetar) (texts and notes)*. Amsterdam: Noord-Hollandsche Uitgevers-Maatschappij.
- Jonker, J.C.G. 1908. *Rottineesch–Hollandsch woordenboek*. Leiden: E.J. Brill.
- Key, Mary Ritchie, and Bernard Comrie, ed. 2015. *The intercontinental dictionary series*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

- Klamer, Marian. 1998. *A grammar of Kambara*. Berlin/New York: Mouton de Gruyter.
- Mahdi, Waruno. 1994. Some Austronesian maverick protoforms with culture-historical implications: II. *Oceanic Linguistics* 30:431–490.
- McArthur, Tom, Jacqueline Lam-McArthur, Lise Fontaine, and Roshan McArthur. 2018. *The Oxford companion to the English language*. Oxford: Oxford University Press.
- McDonnell, Bradley. 2009. A preliminary description of Ende phonology. *Journal of the Southeast Asian Linguistics Society* 2:195–226.
- Mead, David. 1999. *The Bungku-Tolaki languages of South-Eastern Sulawesi, Indonesia*. Canberra: Pacific Linguistics. D-91.
- Moro, Francesca, and Hanna Fricke. 2020. Contact-induced change in Alorese Give-constructions. *Oceanic Linguistics* 59:116–147. URL <https://hdl.handle.net/1887/70891>
- Moro, Francesca R. 2018. The plural word *hire* in Alorese: Contact-induced change from neighboring Alor-Pantar languages. *Oceanic Linguistics* 57:177–198. URL <http://hdl.handle.net/1887/70891>.
- Morris, Cliff. 1984. *Tetun—English dictionary*. Canberra: Pacific Linguistics. C-83.
- Schapper, Antoinette, and Harald Hammarström. 2013. Innovative numerals in Malayo-Polynesian languages outside of Oceania. *Oceanic Linguistics* 52:423–456.
- Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. 2012. Borrowability and the notion of basic vocabulary. In *Quantitative approaches to linguistic diversity*, ed. Søren Wichmann and Anthony P. Grant, 35–55. John Benjamins Publishing Company.
- Verdzade, Allahverdi. 2019. Selected topics in the phonology and morphosyntax of Laboya. Master's thesis, Stockholm University.
- Wolff, John U. 2010. *Proto-Austronesian phonology with glossary*. Ithaca, NY: Cornell Southeast Asia Program Publications.

The Mixed Lexicon of Lamaholot (Austronesian): A Language with a Large Lexical Component of Unknown Origin

Hanna Fricke

1 Introduction*

Eastern Indonesia, an area of linguistic diversity and contact, is characterised by the presence of Austronesian languages and languages of non-Austronesian ('Papuan') families which have co-existed and influenced each other since about 3,500 years. This contact has led to linguistic features diffusing between languages regardless of their genealogical affiliation (Klamer, Reesink, and van Staden 2008, 10, 136; Ewing and Klamer 2010).

Genetically, the population of eastern Indonesia is heterogeneous. Archaeology and population genetics reveal two major waves of modern human (*homo sapiens*) migration into island Southeast Asia: an earlier arrival of non-Asian populations starting about 50,000 years ago and a later influx of Asian populations about 4,000–5,000 years ago (Hudjashov et al. 2017, 2439–2440, 2447). In contrast to the western part of the country, in eastern Indonesia a high degree of ancestry from the earlier population is attested (Bellwood 2017:86–87). The second migration wave is associated with the Malayo-Polynesian branch of the Austronesian language family and its speakers moving from Taiwan into island Southeast Asia, including eastern Indonesia (Karafet et al. 2010, 1833; Bellwood 2017, 181).

In this chapter, I examine the lexical side of language contact between Austronesian (AN) and non-Austronesian (non-AN) languages in eastern Indonesia by taking the AN Flores-Lembata languages, and in particular the Lamaholot subgroups, as a sample case. The Flores-Lembata languages are spoken in eastern Flores and the Solor Archipelago in the Indonesian province of Nusa Tenggara Timur (cf. section 2).

* This chapter is based on Part II of the author's dissertation (Fricke 2019a). I would like to thank Marian Klamer, Owen Edwards and Francesca Moro for their comments on earlier versions of this chapter, and Naonori Nagaya and Alex Elias for their critical reviews which raised very

So far there has been little research which systematically examined the (entire) lexicon of an AN languages of eastern Indonesia looking at AN versus non-AN origin. Reid (1994) is a study on possibly non-Austronesian vocabulary in the AN Negrito languages of the Philippines. Blust (2013, 691–694) discusses AN languages in Melanesia with very high lexical replacement rates. However, differently to the Negrito languages of the Philippines, the non-AN component in the lexicon of these Melanesian languages is so high that their classification as Austronesian becomes debatable. Elias (2020, 331) shows AN retention rates of 60–70% for basic vocabulary in the Central Flores languages Lio, Keo and Rongga, but does not go into details about the innovated part of the lexicon. Edwards (2021) is a comparative dictionary of the AN Rote and Meto languages on the eastern Indonesian island of Timor which reveals an equally large amount of non-AN vocabulary as attested in the Flores-Lembata languages. The AN and the non-AN components of the Rote-Meto lexicon each have different sets of regular sound correspondences (Edwards 2016; 2018; this volume). Other studies on AN-non-AN contact have often focussed on the diffusion of grammatical and lexical features over AN and non-AN languages in larger linguistic areas, such as East Nusantara (Klamer, Reesink, and van Staden 2008; Ewing and Klamer 2010; Holton and Klamer 2017) and Wallacea (Schapper 2015). Based on this background, the present study adds to the more thorough investigation of non-AN vocabulary in individual AN languages and low-level families.

The focus of this study is the non-AN vocabulary in the Flores-Lembata languages, in particular that of the Lamaholot subgroups. As non-AN influence has been proposed for several structural features in all Lamaholot subgroups (Fricke 2019a, Part III), non-AN traces are also expected in the lexicon. Based on this hypothesis, this study addresses the following research questions.

- 1) How big are the AN and the non-AN components of the reconstructed Proto Flores-Lembata (PFL) lexicon? (section 5)
- 2) How big are the AN and the non-AN components of the lexicon in individual varieties of each Flores-Lembata subgroup? (section 6.2)

As it turns out that the non-AN component of the individual varieties (Question 2) is much bigger than the non-AN component of the PFL reconstructions

good additional points. Furthermore, this work would not have been possible without the Dutch Research Council (NWO)'s VICI grant for the project *Reconstructing the past through languages of the present: The Lesser Sunda Islands* by Prof. dr. Marian Klamer (project number: 277-70-012).

(Question 1), the non-AN component of the individual Flores-Lembata languages is investigated further through the third research question.

- 3) Which and how many non-AN lexemes in the Flores-Lembata languages cannot be reconstructed to PFL but are attested with regular correspondences in more than one subgroup of the family? (section 6.3)

Based on the lexical findings, I argue that the Lamaholot languages went through a prolonged period of bilingualism where speakers spoke earlier versions of the present-day Lamaholot languages and one or more unknown non-AN languages. Code-switching must have been a very common, or even the main, way of communication which finally led to the distinct traces of contact in the Lamaholot lexicon.

This chapter is structured as follows. Section 2 introduces the Flores-Lembata languages and their linguistic context. Section 3 describes the dataset and methodology used to investigate the Flores-Lembata lexicon. Section 4 summarizes the historical phonology of Flores-Lembata, including reconstructed PFL phonemes and subgroup-defining sound changes. The results of the lexical study are presented in section 5 on the origins of the reconstructed vocabulary of PFL and in section 6 on the AN and non-AN components of the present-day Flores-Lembata languages. Section 7 discusses the implication of the results for the reconstruction of a possible contact scenario which led to the non-AN part of the lexicon. In section 8, I provide a summary of the main conclusions of this study.

2 The Flores-Lembata Languages

The Flores-Lembata languages, displayed on the map in Figure 5.1, are spoken in the eastern part of Flores and in the Solor Archipelago in the Indonesian province Nusa Tenggara Timur. Based on exclusively shared innovations, I distinguish five Flores-Lembata subgroups: Sika, Western Lamaholot, Central Lamaholot, Eastern Lamaholot, and Kedang (cf. section 4 and Fricke (2019a, 222–226) for more details). In total, there are about 500,000 speakers of Flores-Lembata languages, out of which Western Lamaholot with about 300,000 speakers and Sika with around 175,000 speakers form by large the biggest groups (Fricke 2019a, 156–160).

The Flores-Lembata languages are a subgroup within the Malayo-Polynesian branch of Austronesian, see Figure 5.2. The Flores-Lembata family is part of the larger low-level subgroup of Bima-Lembata (Fricke 2019a, 229). As indicated on the map above, all neighbouring languages are also Austronesian, except for the non-Austronesian languages on the islands of Alor and Pantar, and

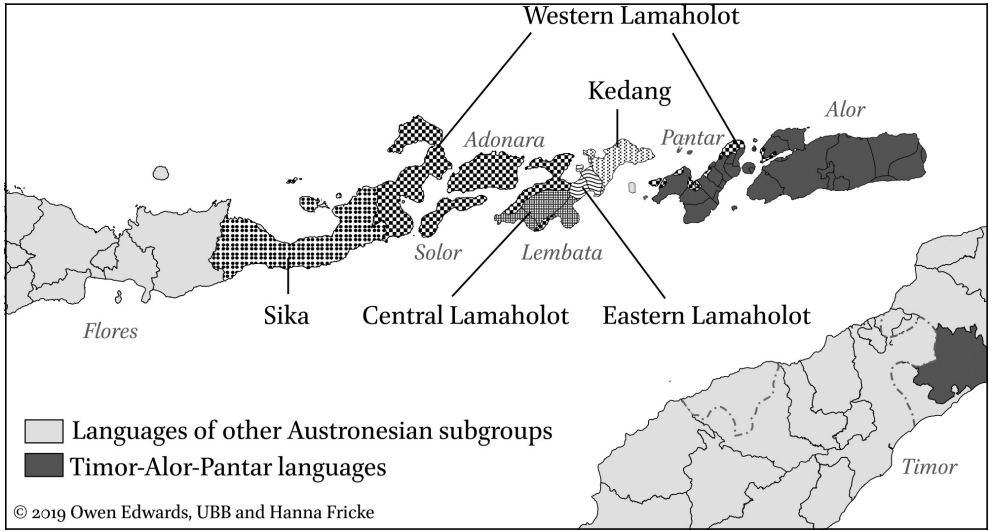


FIGURE 5.1 The Flores-Lembata languages and their linguistic context

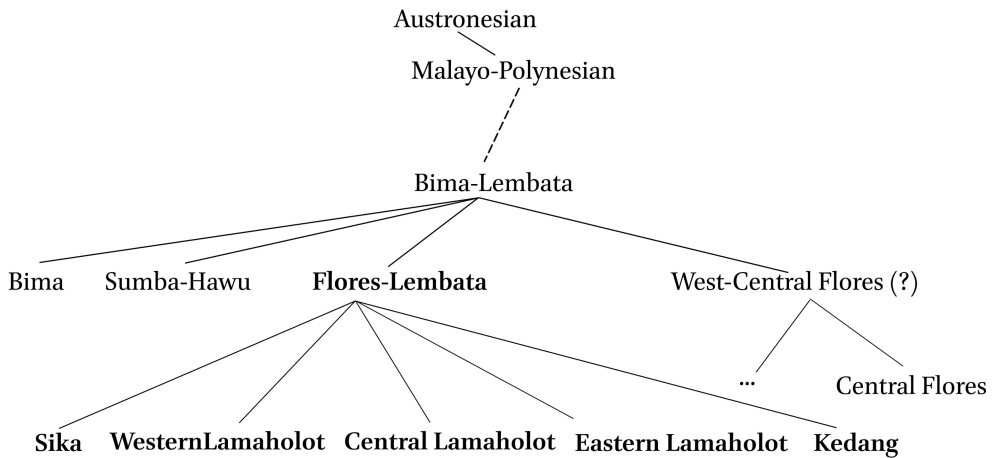


FIGURE 5.2 The Flores-Lembata languages and their genealogical affiliation

parts of Timor, which belong to the Timor-Alor-Pantar family. Towards the west of the Flores-Lembata languages, the Austronesian Central Flores languages are spoken (Elias 2018). On the island of Timor, southeast of the Solor archipelago, the Austronesian Timor-Babar languages and Central Timor languages are found (Edwards 2018; 2019).

Previous comparative studies on the Flores-Lembata languages and Lamaholot have not considered Central Lamaholot and Eastern Lamaholot as independent branches of Flores-Lembata (Keraf 1978; Fernandez 1996; Doyle 2010;

Grangé 2015). Keraf (1978, Appendix), based on lexicostatistics, is the only study which established three Lamaholot subgroups but, different to the proposal here, he connected all three groups to one Lamaholot node, which then connects to Sika and Kedang on a higher level without name. In all other studies, the geographic areas of Central Lamaholot and Eastern Lamaholot were included in the so-called Lamaholot dialect chain or cluster but remained linguistically undescribed.

Most published linguistic research on individual varieties of the Lamaholot dialect chain has been conducted on varieties of the Western Lamaholot group (Arndt 1937; Fernandez 1977; Keraf 1978; Pampus 1999; Nishiyama and Kelen 2007; Nagaya 2011; Klamer 2011; Grangé 2015; Kroon 2016; Akoli 2010; Michels 2017). Only little has been published on Central Lamaholot varieties (Akoli 2010; Krauße 2016; Fricke 2019a; 2019b; 2019c), while no publication is available yet on Eastern Lamaholot varieties. Several descriptive linguistic works are also published on Sika (Arndt 1931; Rosen 1977; 1986; Lewis and Grimes 1995; Bolscher 1982; Pareira and Lewis 1998; Fricke 2014) and Kedang varieties (Samely 1991; Samely and Barnes 2013).

In this chapter, I use the term “Lamaholot” to refer to the three subgroups Western Lamaholot, Eastern Lamaholot and Central Lamaholot as a unit of closely-related subgroups that have been in close contact. However, there is no evidence that Lamaholot, thus the three subgroups, forms an innovation-defined subgroup within Flores-Lembata (Fricke 2019a, 226–228). The reasons for not abandoning the label and concept of Lamaholot encompassing the three subgroups as a whole is (i) the fact that the speakers of the three Lamaholot subgroups see themselves as belonging to one socio-cultural unit opposed to their neighbours Kedang in the east and Sika in the west and (ii) the three subgroups have been in contact until today and share certain structural features that are not attested in Sika and Kedang, such as clause-final negation and an alienability distinction in the possessive construction (Fricke 2019a, Part III). Also lexically, they are more similar to each other than to Sika and Kedang.

Among the three Lamaholot groups, there is little inter-group intelligibility and within each group, various varieties are attested with differences in lexicon, phonology and grammar. For Western Lamaholot, high mutual intelligibility among the group-internal varieties is reported by Michels (2017, 12). Generally, among Western Lamaholot varieties, mutual intelligibility decreases with increasing geographical distance. No empirical data is available on mutual intelligibility within the varieties of Eastern Lamaholot and the varieties of the Central Lamaholot group.

3 Methodology

3.1 *Parallel Vocabularies*

This study approaches the lexicon of a group of languages by investigating the origin of a large set of words in a binary way—Austronesian (AN) versus non-Austronesian (non-AN) origin. Non-Austronesian here means of unknown origin. It is hypothesized that the non-Austronesian component of the lexicon is acquired through language contact, forming a lexical substrate. A similar approach is taken by Edwards (2016; 2018; 2021; this volume) when studying the phonological and lexical features of the Austronesian Rote-Meto languages of Timor.

A potential shortcoming of the method is obviously the fact that more language documentation in (eastern) Indonesia may reveal that some of the non-Austronesian vocabularies are found more wide-spread than previously assumed. However, the number of lexical items of unknown origin will certainly also increase with more documentation, which then consequently will not change much in the distribution of AN versus non-AN components.

3.2 *Data Representation*

Proto Malayo-Polynesian (PMP) phonemes in this chapter are transcribed as in Blust and Trussel (2010). Most of the symbols used by Blust and Trussel (2010) are equivalents to symbols of the International Phonetic Alphabet (IPA). Only the PMP graphemes listed in Table 5.1 do not correspond to the IPA symbols that represent their assumed pronunciation (Ross 1992; Wolff 2010; Blust 2013, 245, 554, 588, 601). To allow a differentiation between schwa [ə] and the unrounded front vowel [e] on lower levels, I re-transcribe all PMP *e with *ə̄. In all other cases, I keep the transcriptions in Blust and Trussel (2010).

The Proto Flores-Lembata (PFL) reconstructions are given using IPA symbols. Only for the palatal approximant [j], I use the symbol *y* instead of its IPA symbol [j] to avoid confusion with a voiced palatal affricate [dʒ] which is often represented with *j* in orthographic transcriptions elsewhere.

Reflexes of entire words or phonemes that are attested in present-day languages are given in *italic* and transcribed in phonemic IPA. Again with the exception of the palatal approximant [j], which I represent as *y* to avoid confusion with a voiced palatal affricate [dʒ]. In data from other sources, the symbol *w* is re-transcribed as *v* for the Flores-Lembata languages, as it is realised as a voiced fricative [v] or approximant [ʋ] in all languages of Flores-Lembata. The vowels *ɛ* and *æ* in some Lamaholot sources are both re-transcribed as *e*, as they are not phonemic. The same is done for *ɔ* and *ɜ* which are re-transcribed as *o* and schwa *ə* respectively for the same reason.

TABLE 5.1 Non-IPA symbols in PMP forms in Blust and Trussel (2010) and in this chapter

Blust and Trussel 2010	This chapter	IPA symbol
⟨j⟩	⟨j⟩	[g] / [ɣ] / [gʲ]
⟨z⟩	⟨z⟩	[dʒ] / [d͡ʒ]
⟨R⟩	⟨R⟩	[r]
⟨e⟩	⟨ə⟩	[ə]
⟨y⟩	⟨y⟩	[j]

3.3 Dataset

The basis for this study are wordlists of 46 Flores-Lembata varieties accessible through the lexical database LexiRumah (Kaiping and Klamer 2018; Kaiping, Edwards, and Klamer 2019), originating from various sources as indicated in the database. Each wordlist contains between 200 and 600 lexical items. In total 607 different concepts of basic, as well as special, vocabulary are covered. For this study, additional information from dictionaries was added for some of the concepts.

From the wordlists, over 400 lexeme sets were extracted using the online tool EDICTOR (etymological dictionary editor) at <https://digling.org/edictor/>. From a lexical database, EDICTOR creates sets of words with similar forms and meanings. The tool also aligns similar sounds within the sets which helps to discover sound correspondences.

I define a lexeme set as a set of formally similar words that appear across languages. There are two types of lexeme sets, cognate sets and similarity sets. Cognate sets trace back to a reconstructible proto form in a proto language, such as Proto Flores-Lembata (PFL), my own reconstructions, or/and Proto Malayo-Polynesian (PMP), as attested in Blust and Trussel (2010). Similarity sets cannot (yet) be reconstructed to a common proto form but they are so similar that they must have some common history. Table 5.2 shows two lexeme sets as examples. The set for the concept ‘seven’ is a cognate set which traces back to a PFL and a PMP reconstruction. For a similarity set, the form given is marked by a hashtag (#), such as #dahe-k ‘near’.

In my dataset of Flores-Lembata vocabulary, I establish a lexeme set if a similar form occurs in at least two of the Flores-Lembata subgroups. Occasionally, my database contains sets based on lexemes that are only found in one Flores-Lembata subgroup but these are only considered if they go back to a PMP form. An example for this is the Sika (Hewa variety) word *roun* ‘leaf’ which traces

TABLE 5.2 Two types of lexeme sets in the Flores-Lembata languages

Lexeme sets		
	Cognate set 'seven'	Similarity set 'near'
	PMP *pitu	-
	PFL *pitu	LH-KD #dahe-k
Western Lamaholot	<i>pito</i>	<i>dahe</i>
Central Lamaholot	<i>pito</i>	<i>dae k</i>
Eastern Lamaholot	[...]	<i>dahe</i>
Kedang	<i>pitu</i>	<i>dehi ?</i>
Sika	<i>pitu</i>	-

PMP = Proto Malayo-Polynesian, PFL = Proto Flores-Lembata, LH = Lamaholot, KD = Kedang, [...] = no data available for this concept, - = no related lexeme, | = historic morpheme boundary

back to PMP *dohun 'leaf'. Among Flores-Lembata languages, only Sika has a reflex of this PMP form, all other languages have replaced this concept with a new lexeme, but as the Sika form clearly goes back to PMP and shows regular sound correspondences, I reconstruct PFL *doun 'leaf' based on the known regular sound changes (see section 4).

The PFL forms in this study are my own reconstructions based on the analysis of the historical phonology of the Flores-Lembata languages (see section 4). I reconstruct a PFL form for a lexeme set if the following criteria are fulfilled. First, the sound correspondences between the reflexes in different subgroups have to be regular. Second, there are two possible conditions that lead to a PFL reconstruction: (i) If the lexeme set can be traced back to a PMP form, then the set is always reconstructed to PFL, or (ii) if no related PMP form is known, the set must be attested in at least Sika and Kedang to be reconstructed to PFL. This means that a form that appears in only one or two Flores-Lembata subgroup and has a PMP form is always reconstructed to PFL. However, if no PMP form exists, only items that are attested in Sika and also in Kedang are reconstructed to PFL. Sika and Kedang are the two Flores-Lembata subgroups that are geographically the furthest apart and therefore, the occurrence of related forms in these two languages points to inheritance from Proto Flores-Lembata (PFL) rather than to diffusion after the split of the family. For most lexeme sets of this study, reconstructions are presented in tables throughout

this chapter. For reflexes of these in the individual languages, please consult the LexiRumah database (Kaiping, Edwards, and Klamer 2019) or the appendix of Fricke (2019a).

3.4 Analysis

The wordlists and established lexeme sets (see section 3.3) were analysed in several ways to answer the research questions listed in section 1. To answer question (1) *How big are the AN and the non-AN components of the Proto Flores-Lembata (PFL) lexicon?*, I compared the number of PFL forms which trace back to a known PMP form, thus being of AN origin, with those that do not, thus being classified as non-AN.¹

To answer question (2) *How big are the Austronesian (AN) and the non-Austronesian (non-AN) components of the lexicon in individual varieties of each Flores-Lembata subgroup?*, I selected, for each subgroup, the variety with most lexical data available. Each item in the wordlist was then classified as AN if it traces back to a PMP form and as non-AN if no known PMP is attested.

To answer question (3), *Which and how many non-AN lexemes in the Flores-Lembata languages cannot be reconstructed to PFL but are attested with regular correspondences in more than one subgroup of the family?*, I counted lexeme sets which cannot be reconstructed to PFL according to the criteria above but, nevertheless, show regular sound correspondences among the subgroups in which the lexemes are attested. An example for such a regular but unreconstructible set is the similarity set in Table 5.2 above. For such a set, a potential reconstruction is established and marked with a hashtag (#) and a subgroup abbreviation, such as LH-KD for Lamaholot-Kedang, to indicate in which subgroups a lexeme of this set is attested.

4 The Historical Phonology of Flores-Lembata

As a background for section 5 and 6, I provide an overview of the reconstructed Proto Flores-Lembata phonology and the exclusively shared sound changes that define the five subgroups of Flores-Lembata, see FIGURE 2 in section 1. No evidence for mid-level subgroups that unite one or more Flores-Lembata subgroups has been found (Fricke 2019a, 226). The reconstruction of the Flores-

1 PFL forms and individual lexemes for which no PMP reconstruction is available could be borrowings from (unknown) non-Austronesian sources but also language-internal innovations, or they could be ultimately of Austronesian origin but due to a lack of data, their PMP origin has not been reconstructed yet. For the purpose of this study, I classify lexemes and proto forms without PMP reconstructions as non-AN.

TABLE 5.3 Proto Flores-Lembata vowel inventory

	Front	Central	Back
High	*i		*u
Mid	*e	*ə (non-final)	*o
Low		*a	

TABLE 5.4 Proto Flores-Lembata consonant inventory

	Labial	Coronal	Dorsal	Glottal
Voiceless stop	*p	*t	*k	*ʔ
Voiced stop	*b (non-final)	*d (non-final)	*g (non-final)	
Affricate		*dʒ (initial)		
Fricative	*v	*s		*h (non-final)
Nasal	*m	*n	*ŋ (non-initial)	
Rhotic		*r		
Lateral		*l		
Approximant			*y [j] (final)	

Lembata phonology and the establishment of the sound changes defining each subgroup are essential to reconstruct the PFL forms presented in section 5 and to prove the regularity of the lexeme sets presented in section 6.3.

Table 5.3 is an overview of the Proto Flores-Lembata vowel inventory and Table 5.4 is a summary of the consonant inventory. The reconstructed sounds are taken from Fricke (2019a, 219–220). If a sound is only attested in certain positions, this is indicated in the tables.

Table 5.5 is an overview of the sound changes attested in each of the five Flores-Lembata subgroups following Fricke (2019a, 224–226). The sound changes are classified as subgroup-defining when they are exclusive to this subgroup. In each subgroup, there are also other sound changes attested, listed in the right column of the table, but these occur in more than one of the subgroups, thus they cannot be regarded as exclusive.

Western Lamaholot is defined by the sound change of PFL *r > ʔ which is regularly attested in intervocalic and final position. In initial position, it only occurs in some lexemes or only in certain varieties but the change is not completed. Central Lamaholot is defined by three exclusively shared sound

TABLE 5.5 Attested sound changes in the Flores-Lembata subgroups

	Subgroup-defining	Other
Western Lamaholot	PFL *ɾ > PWL *ʔ / V_V; #_	PFL *-d- > r PFL *dʒ- > r / #_ PFL *s > h
Central Lamaholot	PFL *-d- > PCL *-dʒ- / V_V PFL *h > PCL ∅ PFL *ʔ > PCL ∅	PCL *s > h (some varieties) PCL *y > dʒ (some varieties) PCL *dʒ > y (some varieties, sporadic) PCL *v > f (some varieties)
Eastern Lamaholot	<i>none</i>	PFL *-d- > r PFL *dʒ- > r / #_ PFL *s > h PFL *k > ?
Kedang	PFL *g > k PFL *-d- > (*dʒ >) y/∅ / V_V	PFL *s > h PFL *k > ?
Sika	PFL *d > r PFL *-ŋ- > n / V_V PFL *mp- > b / #_ PFL *mt- > d / #_	PFL *k > ? PFL *d > r PFL *s > h

PFL = Proto Flores-Lembata, PWL = Proto Western Lamaholot, PCL = Proto Central Lamaholot, V_V = intervocalic position, #_ = final position, #_ = initial position, ∅ = zero (reflex lost)

changes: PFL *-d- > PCL *-dʒ- in intervocalic position and the loss of PFL *h and *ʔ in all positions. In addition to exclusively shared sound changes, the subgroups Western Lamaholot and Central Lamaholot are also defined by further shared lexical and morpho-syntactic innovations, such as the PWL clause-final negator *hala or the PCL plural suffix *-dʒa (Fricke 2019a, 224–226).

Eastern Lamaholot does not undergo any exclusive sound change. All sound changes attested in Eastern Lamaholot are shared with neighbouring languages. PFL *-d- > r is also attested in neighbouring Western Lamaholot varieties, PFL *k > ? is also attested in neighbouring Kedang and PFL *s > h is attested in Western Lamaholot and Kedang. Therefore, these changes are not good evidence for subgrouping. However, Eastern Lamaholot shows some exclusively shared lexical innovations, such as əso ‘tree’ ≠ PFL *kayu ‘tree; wood’ < PMP *kahiwi ‘wood; tree’.² And all other Flores-Lembata languages can be grouped

² The Eastern Lamaholot word əso for ‘tree’ could be related to forms in Alor-Pantar languages, such as Kula *asaka* ‘tree’ or Sawila *asəkə* ‘tree’.

into Western Lamaholot, Central Lamaholot, Sika or Kedang. Therefore, Eastern Lamaholot is nevertheless classified as a subgroup within Flores-Lembata.

Kedang is defined by the exclusively shared sound changes PFL *g > k in all positions and PFL *-d- > (*dʒ >) y / ∅ in intervocalic position. Due to missing historic documentation, there is no direct evidence for the intermediate stage of PFL *-d- > *dʒ in Kedang. However, it is very likely that Kedang went through this stage before > y / ∅. There is evidence from loanwords, such as Kedang *yendela* ‘window’ from Indonesian *dʒendela* and Kedang *yadi* ‘become; happen’ from Indonesian *dʒadi*, that Kedang y in initial position comes from an earlier dʒ (Samely and Barnes 2013, 712; Fricke 2019a, 191).

Sika is defined by four exclusively shared sound changes: PFL *d > r in all positions, PFL *-ŋ- > n in intervocalic position, PFL *mp- > b in initial position and PFL *mt- > d in initial position.

5 Proto Flores-Lembata (PFL) Reconstructions and Their Origin

5.1 Overview

Out of 210 PFL reconstructions in my dataset, about 82% (n=173), listed in section 5.2, are of Austronesian origin, i.e., trace back to a PMP form. Only a small number (n=37) of PFL reconstructions, listed in section 5.3, cannot be connected to any known PMP form. Section 5.4, summarizes and discusses the features of the PFL vocabulary. At the current stage of research, it remains unclear whether the PFL forms without known PMP source can be regarded as a non-Austronesian substrate of PFL. This is because many of these 37 forms are likely to be inherited from an earlier ancestor as similar forms are also found in other Austronesian languages of the region.

5.2 PFL Reconstructions with PMP Sources

Table 5.6 lists 173 PFL reconstructions that have a PMP source and are reflected with largely regular sound correspondences in the Flores-Lembata subgroups. The rightmost column of the table indicates in which subgroups reflexes of the PFL forms are attested. For the purpose of simplicity, the Lamaholot subgroups are grouped together as LH located in the centre of the Flores-Lembata family. LH thus means that a reflex is attested in one or more Lamaholot subgroups. For the last category of PFL reconstructions that only contain reflexes in Lamaholot varieties, LH only in the end of the table, this means that reflexes are attested in at least two Lamaholot subgroups.

TABLE 5.6 PFL reconstructions with PMP source (n=173)³

PFL	PFL meaning	PMP source	Reflex in
*aku	'1SG'	*i aku	SK, LH, KD
*kami	'1PL.EXCL'	*kami	SK, LH, KD
*kita	'1PL.INCL'	*kita	SK, LH, KD
*hida	'3PL'	*si ida	SK, LH, KD
*tudu	'accuse'	*tuzuq	SK, LH, KD
*pənikɪ	'bat'	*paniki	SK, LH, KD
*vani/*blani	'bee'	*wani	SK, LH, KD
*manuk	'bird; chicken'	*manuk	SK, LH, KD
*m-paʔit	'bitter'	*paqit	SK, LH, KD
*mitəm	'black'	*ma-qitəm	SK, LH, KD
*puhun	'blossom; flower'	*pusuŋ 'heart; heart of banana'	SK, LH, KD
*prupi/*plupi	'blow'	*upi	SK, LH, KD
*vulu-k	'body hair'	*bulu	SK, LH, KD
*luri	'bone'	*duRi	SK, LH, KD
*vuhur	'bow'	*busuR	SK, LH, KD
*(t)usu	'breast'	*susu	SK, LH, KD
*mamaʔ	'chew'	*mamaq	SK, LH, KD
*pipi/*klipi	'cheek'	*pipi	SK, LH, KD
*ana(k)	'child; small'	*anak	SK, LH, KD
*piliʔ	'choose'	*piliq	SK, LH, KD
*hakay	'climb'	*sakay	SK, LH, KD
*mai	'come'	PAN *um-aRi	SK, LH, KD
*vatar	'corn; maize'	*batad 'millet; sorghum'	SK, LH, KD
*lədav	'day; sun'	*qaləjaw 'sun'	SK, LH, KD
*matay	'die'	*m-atay	SK, LH, KD
*gali	'dig'	*kali	SK, LH, KD
*bagi	'divide'	*baqagi	SK, LH, KD
*ahu	'dog'	*asu	SK, LH, KD
*-inu	'drink'	*inum	SK, LH, KD
*mada	'dry; thirsty'	*maja	SK, LH, KD
*pa-vari	'dryinsun'	*waRi	SK, LH, KD

3 In all tables, a hyphen (-) indicates a general morpheme boundary, < > indicates an infix. V indicates an unknown vowel, PAN stands for Proto Austronesian

TABLE 5.6 PFL reconstructions with PMP source (n=173) (*cont.*)

PFL	PFL meaning	PMP source	Reflex in
*kVan	'eat'	*kaən	SK, LH, KD
*təlur	'egg'	*qatəluR	SK, LH, KD
*mata	'eye'	*mata	SK, LH, KD
*ama	'father'	*ama	SK, LH, KD
*api	'fire'	*hapuy	SK, LH, KD
*ikan	'fish'	*hikan	SK, LH, KD
*təməla	'flea'	*qatiməla	SK, LH, KD
*vuda	'foam'	*bujəq	SK, LH, KD
*ləpət	'fold'	*lipət	SK, LH, KD
*tuʔan	'forest'	*tuqan	SK, LH, KD
*vua-n	'fruit; betelnut'	*buaq	SK, LH, KD
*m-pənu-k	'full'	*pənuq	SK, LH, KD
*bəli	'give'	*bəRay	SK, LH, KD
*udu	'grass; bush'	*udu	SK, LH, KD
*lima	'hand, arm, five'	*qalima	SK, LH, KD
*kutu	'headlice'	*kutu	SK, LH, KD
*dəŋəR	'hear'	*dəŋəR	SK, LH, KD
*bərat	'heavy'	*(ma)bəRəqat	SK, LH, KD
*pida	'how many'	*pija	SK, LH, KD
*bə-ləma ⁴	'inside; deep'	*daləm	SK, LH, KD
*una	'inside; house'	*qunəj 'pith of plant; core'	SK, LH, KD
*viri	'leftside'	*kawiri	SK, LH, KD
*tave	'laugh'	*tawa	SK, LH, KD
*ʔapur	'lime'	*qapur	SK, LH, KD
*vivir	'lips'	*biRbiR 'lower lip'	SK, LH, KD
*isi-k / *ihi-k	'meat'	*isi	SK, LH, KD
*vulan	'moon'	*bulan	SK, LH, KD
*ina	'mother'	*ina	SK, LH, KD
*ili	'mountain'	*qilih	SK, LH, KD
*vava	'mouth'	*baqbaq	SK, LH, KD
*nadan	'name'	*ŋajan	SK, LH, KD
*pusəR	'navel'	*pusəj	SK, LH, KD
*vəru	'new'	*baqəRu	SK, LH, KD
*niduŋ/*iduŋ	'nose'	*ŋijuŋ/*ijuŋ	SK, LH, KD

4 The prefix b- is a nominaliser in Central Lembata Lamaholot.

TABLE 5.6 PFL reconstructions with PMP source (n=173) (cont.)

PFL	PFL meaning	PMP source	Reflex in
*m-tuʔa	'old (people)'	*ma-tuqah	SK, LH, KD
*əha	'one; alone'	*əsa	SK, LH, KD
*uti	'penis'	*qutin	SK, LH, KD
*ata	'person'	*qaRta 'outsider, alien people'	SK, LH, KD
*vavi	'pig'	*babuy	SK, LH, KD
*bayu	'pound'	*bayu	SK, LH, KD
*veli	'price; brideprice; expensive; buy'	*bəli	SK, LH, KD
*udan	'rain'	*quzan	SK, LH, KD
*uay	'rattan'	*quay	SK, LH, KD
*vanan	'rightside'	*ka-wanan	SK, LH, KD
*m-tasak	'ripe'	*ma-tasak	SK, LH, KD
*lalan	'road'	*zalan	SK, LH, KD
*ramut	'root'	*Ramut	SK, LH, KD
*layar	'sail'	*layaR	SK, LH, KD
*m-pədu	'salty'	*qapəju 'gall' > *ma-pəju	SK, LH, KD
*sama	'same'	*sama	SK, LH, KD
*ənay	'sand'	*qənay	SK, LH, KD
*garu	'scratch'	*garut	SK, LH, KD
*tahik	'sea'	*tasik	SK, LH, KD
*pitu	'seven'	*pitu	SK, LH, KD
*iu	'shark'	*qihu	SK, LH, KD
*m-tidəm	'sharp'	*tazim 'whet'	SK, LH, KD
*meya	'shy; ashamed'	*ma-həyaq	SK, LH, KD
*ənəm	'six'	*ənəm	SK, LH, KD
*ular	'snake'	*hulaR	SK, LH, KD
*mətala	'star'	*mantalaq 'Venus'	SK, LH, KD
*t<m>akav	'steal'	*takaw	SK, LH, KD
*tai	'stomach; belly'	*tian	SK, LH, KD
*vatu	'stone'	*batu	SK, LH, KD
*mulur	'straight'	*lurus	SK, LH, KD
*təvu	'sugarcane'	*təbuh	SK, LH, KD
*naŋi	'swim'	*naŋuy	SK, LH, KD
*luu	'tear'	*luhəq	SK, LH, KD
*pulu	'ten'	*sa-ŋa-puluq	SK, LH, KD
*m-kapal	'thick'	*ma-kapal	SK, LH, KD

TABLE 5.6 PFL reconstructions with PMP source (n=173) (cont.)

PFL	PFL meaning	PMP source	Reflex in
*rivu/*ribu	'thousand'	*Ribu	SK, LH, KD
*təlu	'three'	*təlu	SK, LH, KD
*panav	'walk'	*panaw	SK, LH, KD
*kayu	'tree; wood'	*kahiw	SK, LH, KD
*dʒua ⁵	'two'	*duha	SK, LH, KD
*uta	'vegetable; bean'	*qutan	SK, LH, KD
*vaʔir	'water'	*wahiR	SK, LH, KD
*apa	'what'	*apa	SK, LH, KD
*budaʔ	'white'	*budaq	SK, LH, KD
*aŋin	'wind'	*haŋin	SK, LH, KD
*binay	'woman; sister'	*binay 'woman'	SK, LH, KD
*sala	'wrong'	*salaq	SK, LH, KD
*vadi	'younger sibling'	*huaji	SK, LH, KD
*hakay	'ascend'	*sakay	LH, KD
*raya	'big'	*Raya	LH, KD
*tuno	'burn; grill'	*tunu	LH, KD
*tanem	'bury'	*tanəm	LH, KD
*doa ⁶	'far; long'	*zauq	LH, KD
*pukət	'fishnet, fishtrap'	*pukət	LH, KD
*kaviʔ ⁷	'fishhook'	*kawil	LH, KD
*əpat	'four'	*əpat	LH, KD
*paluk	'hit'	*palu	LH, KD
*k-silap	'lightning'	*silap 'sparkle; drizzle'	LH, KD
*təkek	'lizard'	*təktək	LH, KD
*a(m)pu	'mother's brother'	*əmpu 'grandparent /grand-child'	LH, KD
*nusu	'mouth'	*ŋusu	LH, KD
*kiput	'narrow'	*kiput	LH, KD
*garaŋ	'rough'	*garaŋ	LH, KD
*takut	'scared'	*takut	LH, KD
*kələm	'sky'	*kələm 'dark, overcast, obscure'	LH, KD

5 PFL *dʒ- < PMP *d- is an irregular reflex.

6 PMP *-au- > PFL *-oa- is an irregular change.

7 Sika *kavir* 'fishhook' is related but has irregular initial *k =k rather than expected *k > ?/∅.

TABLE 5.6 PFL reconstructions with PMP source (n=173) (cont.)

PFL	PFL meaning	PMP source	Reflex in
*diri	'stand'	*diRi	LH, KD
*lahe-k	'testicles'	*lasəR	LH, KD
*m-nipih-i	'thin'	*ma-nipis	LH, KD
*basa	'wash'	*basəq	LH, KD
*tani ⁸	'weave'	*tənun	LH, KD
*kapik ⁹	'wing'	*kapak	LH, KD
*tuun ¹⁰	'year'	*taqun	LH, KD
*modip	'alive, live'	*ma-qucip	SK, LH
*ʔavu	'ash, dust'	*qabu	SK, LH
*uma ¹¹	'garden'	*quma	SK, LH
*leba	'burdenstick'	*lemba	SK, LH
*tani ¹²	'cry'	*taŋis	SK, LH
*taʔi	'excrement'	*taqi	SK, LH
*puhun	'heart'	*pusuŋ 'heart; heart of banana'	SK, LH
*laki	'husband; male'	*laki	SK, LH
*gatər	'itchy'	*gatəl	SK, LH
*lotur	'knee'	*qulutuhud	SK, LH
*siva	'nine'	*siwa	SK, LH
*meran	'red'	*ma-iRaq	SK, LH
*gəvalik ¹³	'return'	*balik	SK, LH
*padi	'riceplant'	*pajay	SK, LH
*tali	'rope'	*talih	SK, LH
*plari/*kari	'run'	*lariw	SK, LH
*kulit	'skin'	*kulit	SK, LH
*g-nilu-k ¹⁴	'sour'	*ŋilu	SK, LH
*ikur	'tail'	*ikuR	SK, LH

8 The vowel changes from PMP to PFL are irregular.

9 (i) Sika *kapik* 'wing' is related but has irregular initial *k =k rather than expected *k > ʔ/∅.
(ii) PMP *a > PFL *i is an irregular change.

10 PMP *-aqu- > PFL *-uu- is an irregular change.

11 Kedang *lumar* 'garden' could be related.

12 Intervocalic PFL *-n- < PMP *-ŋ- is irregular.

13 PMP *balik > PFL *gəvalik is most likely PMP *b > *w > *v with the addition of a verbalising prefix g-.

14 Kedang *kiru* 'sour' could be related.

TABLE 5.6 PFL reconstructions with PMP source (n=173) (cont.)

PFL	PFL meaning	PMP source	Reflex in
*m-panau	'tinea'	*panaw	SK, LH
*puki	'vagina'	*puki	SK, LH
*hapu	'wipe'	*sapu	SK, LH
*sika	'chase away'	*sika	LH
*buŋa/*puŋa	'flower'	*buŋa	LH
*(kə)namuk	'fly' (n.)	*ñamuk 'mosquito'	LH
*tuma	'louse on clothing'	*tumah	LH
*ta(ke)	'no; not'	*taq	LH
*bukat	'open'	*bu(ŋ)kas	LH
*mula	'plant'	*mula	LH
*(v)uvuŋ ¹⁵	'ridge'	*bubuŋ	LH
*hira	'salt'	*qasiRa	LH
*tudu	'sleep'	*tuduR	LH
*ipe	'teeth'	*(n)ipən	LH
*baŋun	'wake up'	*baŋun	LH
*an	'what'	*anu	LH
*muav	'yawn'	*ma-huab	LH

5.3 PFL Reconstructions without PMP Sources

Table 5.7 lists 37 regular PFL reconstructions that, based on the current stage of knowledge, do not go back to a PMP form. If a related or resemblant form is known to also occur in other languages of the region outside of the Flores-Lembata family, this is indicated in the last column with "Flores" meaning the Austronesian languages of Flores, "Timor (AN)" meaning the in the Austronesian languages of Timor, "Timor (TAP)" meaning in the Timor-Alor-Pantar languages of Timor, and "Alor-Pantar" meaning in the Alor-Pantar languages on the islands of Alor and Pantar. I do not consider the possible occurrences of the lexemes in languages outside of the East Nusa Tenggara and Timor-Leste region. Further research on the lexicon of the languages in this area and beyond will probably increase the number of these regionally spread items. Currently, 14 out of 37 lexeme sets listed here are also found outside of the Flores-Lembata family. The remaining 23 reconstructions are considered innovations of PFL.

15 Sika *puvun* 'ridge' could be related.

TABLE 5.7 PFL reconstructions without PMP source (n=37)

PFL	PFL meaning	Regional spread
*təmisi	'ant'	
*dasan	'ask; report'	
*muku	'banana'	Flores, Timor (AN), Timor (TAP), Alor-Pantar
*təmayuŋ	'bedbug'	Flores, Timor (AN)
*giki	'bite'	Flores, Timor (AN), Timor (TAP), Alor-Pantar
*vəki	'body'	Flores
*tena	'canoe'	
*laku	'civet cat'	Flores, Timor (AN), Alor-Pantar
*rusu / *ruhu	'coral reef'	
*pati	'cut'	Flores, Timor (AN)
*gurit	'dig'	
*bao	'float'	
*lodonŋ	'fall down; descend'	
*voda-k	'fat'	Flores
*pə-vunu	'fight'	
*napu-k	'flat; stream; river'	
*pau ¹⁶	'mango'	Flores, Timor (AN)
*motonŋ	'moringa'	Alor-Pantar
*osan	'mat'	
*k⟨n⟩əpuŋ/*həpuŋ	'mosquito'	
*kəmeruŋ	'rice ear bug'	Timor (AN)
*(n)ubak	'stream; river'	
*vura	'sand'	
*labur	'shirt'	Flores, Maluku
*kpali-k/*kwali-k	'shoulder'	
*kamak	'skin; bark of tree'	
*kə-melu	'smooth'	
*m-potaŋ	'spit' (v.)	
*(k)rəvun	'sweat'	
*səru-k	'sweet'	
*alis	'tendon'	Flores

16 Could be related to Proto Western Malayo-Polynesian (PWMP) *qambawan 'manggo'.

TABLE 5.7 PFL reconstructions without PMP source (n=37) (*cont.*)

PFL	PFL meaning	Regional spread
*kera	'turtle'	Flores, Timor (AN), Alor-Pantar ¹⁷
*ale	'waist'	
*hogo	'wake up'	
*gəbi/ *gnəbin	'wall'	Flores
*(l)oyor	'wave; sea'	
*nora	'with'	Flores, Timor (AN)

5.4 *Summary and Conclusions*

The 210 PFL reconstructions are to a great extent of Austronesian origin, for 82% of them, there is a known PMP source. About one fifth of the PFL vocabulary remains of unknown origin. PFL, as a descendant of PMP, has thus replaced about 20% of the vocabulary for the concepts in this study since PMP times, i.e. around 4000 years ago (Pawley 2005). When selecting only basic vocabulary forms (see Appendix for a list of basic concepts) from the sample, around 124 PFL forms remain. Out of these basic forms only 13% are not of PMP origin. This lower percentage of non-PMP vocabulary in PFL basic vocabulary compared to the whole database confirms that lexical replacement in basic vocabulary is less likely to occur than in other parts of the vocabulary.

The PFL vocabulary which is not of PMP origin could be regarded as a non-Austronesian lexical substrate in PFL. However, at the current stage of research, it is not entirely clear if the set of lexical items in PFL that do not trace back to PMP can be part of a substrate in PFL because it is unknown how much of this vocabulary traces further back to an earlier ancestor of PFL. In section 5.3, I have shown that about 30% of the non-AN vocabulary in PFL has related forms elsewhere in the region which suggests inheritance from an earlier ancestor. As this number is based on an initial survey, more in-depth systematic investigation into the lexicon of the languages of the region and even beyond may shed light on how far this vocabulary can be traced back. Some of it may even ultimately go back to PMP. It is possible that with further research, the number of PFL reconstructions without PMP source becomes so small that one could account for it by lexical replacement that naturally occurs in any language for different reasons, such as avoidance of homophony, semantic change, derivation, borrowing and invention of new words.

¹⁷ Proto Central Eastern Malayo-Polynesian (PCEMP) *kəRa or *keRa 'turtle'.

6 The Present-Day Lamaholot Lexicon and Its Origins

6.1 *Overview*

In contrast to the previous section, which concerns the reconstructed vocabulary of Proto Flores-Lembata, this section examines the present-day lexicon of the Flores-Lembata languages and its Austronesian, i.e. tracing back to a PMP form, versus non-Austronesian origins, i.e. cannot be related to any known PMP form. The Lamaholot subgroups contain, with about 50% of their lexicon, the greatest amount of non-AN vocabulary among the Flores-Lembata languages. This vocabulary is of interest because only little of it can be traced back to PFL (see section 5). Therefore, it must have entered the languages after PFL split into subgroups. Section 6.2 present the results for individual Flores-Lembata languages, while section 6.3 provides insights into the non-AN vocabulary of the individual languages which is shared and shows regular sound correspondences among at least two subgroups.

6.2 *The Lexicon of Individual Varieties*

In all three Lamaholot groups, only about 50% of the present-day lexicon trace back to an Austronesian source, as shown in Table 5.8. In the sister languages, the AN component is higher in Kedang with 57% AN origin and again higher in Sika with 62% AN origin.¹⁸

The data in Table 5.8 is based on one variety per subgroup, named in brackets in the table. I have not observed significant variation between the varieties of one subgroup regarding the distribution of PMP versus non-PMP vocabulary. Therefore, the varieties with the largest amount of data available were chosen.

The percentage of non-AN vocabulary is stable across the three Lamaholot groups, even though the size of the datasets varies. The Eastern Lamaholot dataset (n=128) is much smaller than the one of the Central group (n=333) and Western group (n=276) and contains proportionally more basic vocabulary than the larger datasets of Central and Western Lamaholot. Therefore, Eastern Lamaholot shows a slightly higher percentage of AN vocabulary compared to Central and Western Lamaholot.

The results in the table lead to two observations. (1) All three Lamaholot varieties have a very similar percentage of non-AN lexical items, and therefore most likely had a similar history of lexical replacement, and (2) the non-AN component in Lamaholot is higher than in their closest relatives Kedang and Sika.

¹⁸ When only examining basic vocabulary (see Appendix), the AN components are about 10% higher for all five varieties examined. This again confirms that basic vocabulary is replaced less frequently (cf. section 5.4).

TABLE 5.8 AN and non-AN lexemes in individual varieties of the Flores-Lembata subgroups

	AN	Non-AN	Total
Western Lamaholot (Lewoingu)	49% 134	51% 142	276
Central Lamaholot (Central Lembata)	47% 158	53% 175	333
Eastern Lamaholot (Lamatuka)	54% 69	46% 59	128
Kedang (Leubatang)	57% 131	43% 97	228
Sika (Hewa)	62% 136	38% 84	220

The AN component of the present-day Lamaholot lexicon traces back to the 173 reconstructed PFL forms with AN origin, listed in section 5.2. The non-AN component of the Lamaholot lexicon is of further interest because it consists of much more vocabulary than the small set of 37 PFL reconstructions without AN origin, listed in section 5.3. In the following section, the non-AN lexicon of the Lamaholot groups is examined in more detail.

6.3 *The Shared Non-AN Vocabulary*

The non-AN component of the Lamaholot lexicon can be divided into four categories (1) non-AN lexical items with attested regular sound correspondences in two or all Lamaholot subgroups (n=71), (2) non-AN lexical items attested with regular sound correspondences in at least one Lamaholot subgroup and in Kedang (n=73), (3) non-AN lexical items attested with regular sound correspondences in at least one Lamaholot subgroup and in Sika (n=41), and (4) non-AN lexical items only attested in one Lamaholot subgroup (not counted). The last category of non-AN lexical items which only occur in one of the Lamaholot subgroups is rather small and was not systematically counted in this study.

The main interest lays on those non-AN lexemes in category (1), (2) and (3), listed in Table 5.9, which are spread over more than one subgroup and show regular sound correspondences between these groups, thus form lexeme sets (cf. section 3). From the numbers of lexeme sets, it becomes clear that Lamaholot shares most non-AN vocabulary among the three subgroups or shares it with Kedang, while considerably less non-AN vocabulary is shared with Sika.

The last column of the table indicates in which other language groups of the region known related forms are attested. The same categories are used as for Table 5.7 above.

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185)

Lexeme set	Meaning	Regional spread
Lexeme sets only attested in Lamaholot subgroups (n=71)		
#əvan	'accuse'	
#tapan	'answer'	Timor (TAP)
#svaol	'all'	
#knaru	'back'	
#navak	'body'	
#ravuk	'body hair'	Timor (AN)
#esari nai	'breathe' (v.)	
#hopi	'buy'	
#kiri	'comb'	Alor-Pantar (PAP *kir (Robinson 2015))
#oli	'come; arrive'	
#suda	'command; order' (v.)	
#bisu	'cook'	
#kluok	'cooked rice; uncooked rice'	
#vekan	'divide'	
#knavi	'door'	Alor-Pantar (?)
#ləŋat	'fall from above'	
#gəni	'fight'	
#vahak	'finished'	
#lerek	'flat; below'	
#kənito	'forehead'	
#alus	'good'	
#pehen	'grasp; hold'	
#madu	'grasshopper'	
#latar	'hair'	
#kote	'head'	Timor (AN)
#soron	'hide'	
#dani	'hit (drum)'	
#uməŋ	'hole'	

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185) (*cont.*)

Lexeme set	Meaning	Regional spread
#plati/kati	'hot'	
#maluv	'hungry'	
#bati	'hunt'	
#gekay	'laugh'	
#səmekiŋ	'left side'	
#loit	'let go'	
#pavaŋ	'lie' (position for things)	
#kleak ¹⁹	'light (weight)'	
#kmoruŋ	'locust'	
#vuda	'lungs'	Alor-Pantar
#elam	'meat; flesh'	
#vətəm ²⁰	'millet'	Flores
#vala	'mud'	Alor-Pantar
#niləŋ	'necklace'	
#magun	'old people'	
#toʔu	'one'	
#gesak	'other'	
#głasa	'play'	
#nakiŋ	'promise'	Alor-Pantar
#vidu	'pull'	Flores
#magar	'rack above hearth'	
#tue	'return'	
#(a)luŋu	'river; stream'	
#bua	'sail' (v.)	
#sodam	'smell'	Timor (AN)
#m⟨an⟩akap	'sorcerer'	
#pəriŋo	'spit'	
#piʔuk	'squeeze'	
#puka	'stem'	Flores
#mopa	'straight'	
#kebol	'sugar palm'	
#luvak	'sun'	Alor-Pantar

19 Sika *heak* 'light (weight)' and Kedang *ʔahaʔ* 'light (weight)' could be related to #kleak.

20 Kedang *vereʔ* 'millet' could be related to #vətəm.

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185) (*cont.*)

Lexeme set	Meaning	Regional spread
#blolo/golo	'tall'	
#luʔo	'thatch for roofing'	
#tnakar	'thatched roof'	
#pənəŋe	thick	
#prəvak	thick	
#petən	'think; miss'	
#məna	'vagina'	Flores
#rio	'wake someone up'	
#ga(ŋe)	'where'	Alor-Pantar, Timor (TAP)
#henaku	'who'	Timor (AN)
#ugadak	'wound'	

Lexeme sets attested in Kedang and Lamaholot (n=73)

#soloi	'answer' (v.)	
#gəter	'ask question'	
#bovoŋ	'bark'	
#həbu ²¹	'bathe'	
#malu	'betel vine'	Timor (AN), Timor (TAP)
#puur	'blow'	Flores, Timor (AN), AP
#papi	'burn; clear land'	
#letuʔ	'close' (v.)	
#kova ²²	'cloud; fog'	
#korok	'chest'	
#tapu	'coconut'	
#hekan	'condition; time; garden'	
#mudəŋ	'correct; the following'	
#bəpap	'crocodile'	Alor-Pantar
#belu	'cut; kill'	Flores
#sedu	'dance'	
#klebit	'deaf'	

21 Central Lamaholot *ləbo* 'bathe' could be related.

22 Sika *kova* 'cloud' could be related but would involve an irregular retention of PMP *k = Sika *k*. This lexeme set might trace back to PMP *awaŋ 'atmosphere, space between earth and sky' with an insertion of initial *k*- and an irregular change of PMP *a > PFL *o.

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185) (*cont.*)

Lexeme set	Meaning	Regional spread
#butu	'eight; bunch; group'	Flores, Timor (AN), AP ²³
#gokal	'fall over'	
#bəka	'fly'	
#lei	'foot, leg'	
#(kəne) breuŋ ²⁴	'friend'	
#neʔi	'give'	Timor (AN)
#gedi	'go up; ascend'	
#dikə-n ²⁵	'good; person'	
#vurek	'gravel'	
#tava ²⁶	'grow; stem'	
#pohiŋ	'help'	
#vuok	'hole'	
#vetak	'house; barn'	
#nara bone gaku	'how'	
#kverak	'jackfruit'	Alor-Pantar
#kudul	'knee'	
#lolo	'leaf'	
#ləpa	'leaf; sheet; lontar leaf'	
#benehik	'light (not dark)'	
#(kutu) kihan	'louse eggs'	
#kabe	'man; husband; person'	

23 PCEMP *butu 'group, crowd, flock, school, bunch, cluster'.

24 Sika *deuy* 'friend' could be related but would involve an irregular correspondence of Lamaholot/Kedang *br-* and Sika *d-*.

25 The set #dikə-n could derive from PMP *diqɑq 'good' with an irregular change of PMP *-q- > PFL *-k- before ə. However, as also the change of PMP *-ɑq > PFL *-ə in this word remains unexplained, PFL *dikə 'good; correct' might also be unrelated to PMP *diqɑq. The original meaning of this set is probably 'good; correct'. The word 'good' is combined with another word for 'person', i.e. PFL *ata, such as still in used for example in Central Lembata *ata dikən* 'person'. This was probably done as an opposition of members of another group that were enemies. Over time, also the second part of the compound acquires the meaning 'person'. However, in some subgroups, such as for example in Kedang and Eastern Lamaholot, both meanings 'good' and 'person' are retained. In Alorese, a reflex of PFL *dikə means 'right side'.

26 Eastern Lamaholot *nava* 'stem' could be related.

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185) (*cont.*)

Lexeme set	Meaning	Regional spread
#rai-k ²⁷	'many'	
#tudak	'narrow'	
#dahe-k	'near'	
#vuli	'neck'	Alor-Pantar
#batul	'needle'	Alor-Pantar
#payam	'papaya'	
#volar	'ridge'	
#vadək	'rope'	
#doru ²⁸	'rub; wipe'	Alor-Pantar
#taʔu	'salt'	
#bota(n)	'sand'	
#kəburak	'scabies'	Flores
#kuluk	'seed'	Alor-Pantar
#durum	'sell'	
#saur	'sew'	Timor (AN), Alor-Pantar
#məkul	'short'	
#tobe	'sit'	
#təguʔ	'skewer'	
#molan	'sorcerer'	
#gala(r)	'spear'	Flores
#təmidu ²⁹	'spit'	Timor (AN)
#bəta	'split'	
#tubak	'stab'	
##(kə)boti	'stomach; belly'	
#kebaŋ	'storage house; barn'	
#pola	'sugar palm'	
#soŋa	'tie'	
#ebel	'tongue'	
##(bela) bayan	'treaty'	Alor-Pantar
#deko	'trousers'	Flores, Timor (AN), Alor-Pantar

27 #rai 'many' could trace back to PMP *Raya 'big'.

28 Western Lamaholot *doruk* 'rub; wipe' could be related but would involve an irregular retention of PFL *r = WL r.

29 This could be related to PWMP *qizuR 'saliva; spittle'.

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups (n=185) (*cont.*)

Lexeme set	Meaning	Regional spread
#ləvu	'village'	
#luanj	'vomit'	
#hamu	'wipe; sweep'	Timor (AN)
#kumas	'yellow'	
#evian	'yesterday'	
Lexeme sets attested in Sika and Lamaholot (n=41)		
#supel	'arrow'	Flores, Alor-Pantar (?)
#baka	'bite'	Flores
#(sə)mei	'blood'	
#nahi	'breath'	Flores
#ihere	'close' (v.)	
#kōbu	'crocodile'	
#gasik	'count'	Timor (AN)
#kəbehar	'cuscus'	
#baŋak	'flow'	Flores
#-ai	'go'	
#volonj	'hill; ridge'	Flores
#tara	'horn'	
#(raʔi) etan	'know'	Timor (AN)
#blavir	'long; far'	
#koli	'lontar palm'	Flores, Alor-Pantar
#(meiŋ) ʔetan	'meat'	
#təker	'narrow'	Flores
#lusir	'needle'	
#guman	'night'	Timor (AN), Alor-Pantar
#dʒəma	'night, time unit'	
#pehan ³⁰	'other'	Flores
#likat	'oven'	Flores
#əpak	'palm of hand; footprint'	
#pahat	'plant yam'	Flores
#tubu	'pull'	
#gide	'pull'	

30 Kedang *palan* 'other' could be related.

TABLE 5.9 Regular but unreconstructible lexeme sets among Flores-Lembata subgroups
(n=185) (*cont.*)

Lexeme set	Meaning	Regional spread
#gualok	'round'	
#madi	'say'	Flores
#kəmekot	'scorpion'	
#buʔu	'short'	Flores
#blara	'sick; painful'	
#tuʔay ³¹	'sleep'	
#nuhi	'smoke'	Flores, Timor (AN)
#pemek	'squeeze'	Alor-Pantar
#robak	'stab'	
#hukut	'think; remember; miss'	
#kleka ³²	'thunder'	
#papa lele	'trade'	
#puʔu	'wash'	Flores
#kəsako	'whisper'	
#ledan	'wide'	

6.4 Summary and Conclusions

It has been shown that about 50% of the present-day lexicon of Lamaholot cannot be traced back to an Austronesian origin. Most of this non-AN vocabulary is shared among all Lamaholot subgroups, and often also shared with Kedang, less frequently with Sika. The shared vocabulary shows regular sound correspondences among the subgroups. However, as none of the 185 lexeme sets in this section is attested in both Sika and Kedang, the western and eastern most languages of the Flores-Lembata family, this vocabulary cannot be reconstructed to Proto Flores-Lembata (cf. section 3). This stands in contrast to the 37 non-AN lexical items which are reconstructible to PFL (cf. section 5.3).

The fact that the non-AN lexical items show regular sound correspondences over the subgroups suggests that these vocabulary additions cannot be very recent. They must have become part of the language before the respective sound changes had occurred or were still ongoing, thus had not ceased to be

31 Kedang *tɛʔel* 'sleep' could be related.

32 CL-Kalikasa *kəlagor* 'thunder' could be related but would require an irregular change of the last syllable #ka to Kalikasa *gor*.

active yet. Therefore, this vocabulary must have been added at some point after the split off PFL into subgroups but before the groups terminated their individual sound changes.

7 Discussion

7.1 *Non-Austronesian Features in Lamaholot*

The results of the lexical study presented in section 5 and 6 have shown that most of the Proto Flores-Lembata (PFL) vocabulary can be attributed to an Austronesian source (82% AN). This means that, lexically, PFL was a largely Austronesian language.³³ However, when examining the present-day vocabulary of the descendants of Proto Flores-Lembata, it becomes clear that the amount of non-AN vocabulary increased after the split of the proto language into subgroups. About half of the lexicon of the present-day Lamaholot varieties does not trace back to an Austronesian source (51% non-AN in Western Lamaholot, 53% non-AN in Central Lamaholot and 46% non-AN in Eastern Lamaholot). To a lesser extent, this is also observed in the sister languages Kedang (43% non-AN) and Sika (38% non-AN).

The non-AN vocabulary covers virtually all semantic domains. There are large amounts of basic vocabulary denoting properties or verbal concepts, such as #plati/kati 'hot' in the Lamaholot subgroups or #tuʔay 'sleep' in Sika and Lamaholot. Also body part nouns are a rather big group with 22 non-AN terms, out of which only 5 can be reconstructed to PFL. In addition to that there is special vocabulary in the domains of flora and fauna, such as #kobu 'crocodile' in Lamaholot and Sika or #tapu 'coconut' in Lamaholot and Kedang. In total, the database contains 19 non-AN animal terms, out of which 6 are reconstructible to PFL, and 17 non-AN terms in the semantic domain of plants, out of which 4 trace back to PFL.

Table 5.10 compares the Flores-Lembata languages with their closest Austronesian neighbours, the Rote-Meto languages on Timor (Amarasi and Termanu in the table), and the Central Flores languages in central Flores (Rongga, Keo, and Lio in the table). The table is sorted by increasing percentage of AN lexical retention in the basic vocabulary.

The Lamaholot subgroups show only slightly higher rates of non-AN basic vocabulary than most other Austronesian languages of the region. Thus, the Lamaholot subgroups fit into the regional pattern when it comes to the com-

33 Structurally, PFL innovated several non-AN features, these include word order patterns in

TABLE 5.10 Regional comparison of AN and non-AN components of the lexicon

	Basic vocabulary		Entire lexicon	
	AN	Non-AN	AN	Non-AN
Central Lamaholot	57%	43%	47%	53%
Amarasi (Edwards pers. com.)	58%	42%	–	–
Western Lamaholot	61%	39%	49%	51%
Eastern Lamaholot	62%	38%	54%	46%
Termanu (Edwards pers. com.)	62%	38%	–	–
Rongga (Elias 2020: 331)	63%	37%	–	–
Kedang	64%	36%	57%	43%
Keo (Elias 2020: 331)	64%	36%	–	–
Lio (Elias 2020: 331)	69%	31%	–	–
PRM (Edwards this volume)	69%	31%	45%	55%
Sika	75%	25%	62%	38%
PFL (Fricke 2019a: 248–249)	87%	13%	82%	18%

position of the lexicon. However, the very high AN retention rate of PFL is striking. The other proto language in the table, Proto Rote-Meto, does also have a slightly higher rate than the present-day Rote-Meto languages but the difference between proto language and present-day languages is not as big as for PFL and its descendants. Two possible reasons can be proposed for this difference, (1) PFL could be older than PRM, as with more time obviously more vocabulary can be replaced, or (2) PFL could have been less influenced by non-AN languages than it was the case for PRM.

A possible shortcoming of the comparison in Table 5.10 is that the percentages come from different studies with somewhat different methodologies and definitions of basic vocabulary. Therefore, it cannot be excluded that some differences are due to methodology.

In the following section, I discuss reasons for the increase in non-AN vocabulary after the split of Proto Flores-Lembata into subgroups. I argue that this added non-AN vocabulary is a lexical substrate and points to a contact scen-

the nouns phrase, property nouns, and the clause-final deictic motion verbs ‘come’ and ‘go’ (Fricke 2019a, Part III).

ario with now extinct non-Austronesian languages in the area. This hypothesis is supported by non-Austronesian structural features which are attested in the Lamaholot subgroups and are not found, or only to a much lesser extent, in Sika and Kedang. All three Lamaholot groups innovated clause-final negation, an alienability distinction in the possessive construction, clause-final deictic motion verbs encoding elevation, and the Central Lamaholot group innovated a general plural suffix for nouns (Klamer 2012; Fricke 2017; 2019a).³⁴ Due to lexical differences in the clause-final negators, the deictic motion verbs and the ways the alienability distinction is realized, the innovations likely developed independently in each of the groups. However, it appears that they are all caused by contact to typologically very similar languages that are now extinct. These contact languages that triggered the innovation of the features just mentioned were probably non-Austronesian with a typological profile similar to the Timor-Alor-Pantar (TAP) languages spoken towards the east of the Lamaholot area (see FIGURE 1 in section 2). This is proposed because the present-day TAP languages have exactly these structural features which are innovations in Lamaholot but retentions in the TAP languages. The lexical substrate, however, does not point to the TAP family, as there are only a few lexical items which have similar forms in the TAP languages (cf. Table 5.9 in section 6.3).

7.2 *Reconstructed Contact Scenarios for Lamaholot*

Depending on the circumstances, contact-induced language change can affect any feature of a language (Thomason and Kaufman 1988, 14). The social scenario in which the contact takes place plays an important role in determining constraints on contact-induced change for a particular contact situation (Muysken 2010). Analysing the outcome of languages contact, such as the innovated vocabularies and grammatical features of Lamaholot, a possible contact scenario can be reconstructed.

As discussed in Fricke (2019a, 415–416), the evidence for non-AN grammatical features in the Flores-Lembata languages suggests that the ancestors of the Flores-Lembata people were bilingual speakers of at least one AN and one non-AN language over several generations. This led to convergence in word order and new morpho-syntactic categories based on semantic distinctions. These kind of changes can be attributed to bilingual copying, a term which Ross (2013, 6, 23) uses for “change which bilingual speakers introduce into one of their lan-

34 As there is almost no data on Eastern Lamaholot, it is only known for sure that Eastern Lamaholot has clause-final negation. The other features remain to be investigated.

guages on the model of their other language.” In PFL only syntactic changes are attested but no additional features. The same holds for Sika. In Kedang and the Lamaholot varieties, features were added and this means an increase in complexity (Ross 2013, 32). This qualitative difference in contact outcomes between PFL and Sika, on the one hand, and Kedang and Lamaholot, on the other hand, is also found in the amounts of new non-AN vocabulary. The increase of new lexical items in PFL and Sika is lower than in Kedang and the Lamaholot varieties.

I argue that for the case of Lamaholot, and possibly also for Kedang, contact-induced introduction of additional vocabulary is the most likely explanation for most of the non-AN vocabulary. It is not realistic to assume that a community invented all this new vocabulary. According to Thomason (2007), deliberate language change is still rare. Nevertheless, I am not excluding that some of the new vocabulary was indeed invented due to reasons, such as taboo or esoterogany, or are more recent borrowings.

The large amount of new vocabulary is more likely to be a remnant of code-switching by highly proficient bilinguals. The new vocabulary is basic as well as special vocabulary (see section 7.1). No specific semantic domain is clearly favoured. A social situation that can lead to such an unsystematic mixing of vocabulary is a community where all speakers are fluent bilinguals and where code-switching is the most common form of communication. The “fossilisation” of such type of code-switching can lead to a so-called bilingual mixed language (Thomason 2001, 198, 215). In the development of Lamaholot such a mixed code could have become the main way of communication in the community. After a few generations, this way of speaking became then more standardized and finally the only language of the community.

The central American language Garifuna is such a language which is the result of language mixing over several generations (Haurholm-Larsen 2016). In the case of Garifuna, more is known about the history of the Garifuna people and the language material clearly shows two source language families, Arawak and Carib, none of them being extinct. The social scenario behind the Garifuna language is the following. All Arawak male speakers were killed by invading Carib male speakers who then lived on with the Arawak women. The Arawak language became their common language, however, was heavily influenced by Carib grammatical structures and lexical items. A clear relict of the dual origin of the lexicon are parallel lexemes, one Arawak term and one Carib term with the same meaning (Haurholm-Larsen 2016, 289–290).

The Lamaholot variety Central Lembata shows a similar phenomenon. Table 5.11 lists 15 lexeme pairs in Central Lembata with the same meaning but two origins. One of each pair is of Austronesian origin, a PMP form is provided

TABLE 5.11 Parallel lexemes in Central Lembata Lamaholot

Meaning	Central Lembata	PMP origin
'old (for people)'	<i>tuan</i> (for women)	*ma-tuqah
	<i>magun</i> (for men)	-
'body'	<i>navak</i>	*hawak 'waist; back of the waist'
	<i>vəki</i>	-
'belly'	<i>tai</i>	*tian
	<i>kboti</i> / <i>kaluŋ</i> 'gut'	-
'corner'	<i>bnelok</i>	*beluk 'bend'
	<i>snikup</i>	-
'to fight'	<i>punu</i>	*bunuq 'kill'
	<i>punu geni</i> (only in combination)	-
'garden'	<i>maan</i>	*quma
	<i>ekan</i>	-
'to give'	<i>bee</i>	*bəRay
	<i>noto</i>	
'male'	<i>lakin</i> (for animals)	*laki
	<i>lamen</i> (for humans)	
'name'	<i>nadzan</i>	*ŋajan
	<i>maken</i>	
'person'	<i>ata</i>	*qaRta 'outsider, alien-people'
	<i>dikən</i>	-
'to speak'	<i>tutu</i>	*tutur
	<i>pnua</i>	-
'to stand up'	<i>baŋu</i>	*baŋun
	<i>boko</i>	-
'to steal'	<i>takav</i>	*takaw
	<i>lavit</i>	-
'stem'	<i>puuk</i>	*puqun
	<i>tava</i>	-
'to swim'	<i>naŋe</i>	*naŋuy
	<i>dulo</i>	-

in the right column, and the other one of each pair is of unknown origin. Most pairs are synonyms, some are near-synonyms. Some of these pairs, but not all, are often used in combination, such as *vəki navak* 'body', *tutu pnua* 'to discuss', and *ata dikən* 'person'. Knowing about the case of Garifuna, it can be proposed that the non-AN parts of the pairs are relicts from the now extinct non-AN contact language(s).

Nowadays no non-Austronesian languages are spoken anymore in the whole area of Flores-Lembata. Therefore, also the contact scenarios of the proto languages of Lamaholot must have reached the stage of language shift towards the Austronesian languages at some point. When finally all speakers shifted, the languages had already been heavily influenced by the non-Austronesian languages due to a long and intensive period of bilingualism. It may even be possible that, as the whole society became bilingual, speakers did not differentiate the languages any more but the mixed code became their new language. Nevertheless, the Lamaholot languages remain overall more Austronesian than non-Austronesian in lexicon and grammar. Therefore, assuming a mixed code does not mean an equal mix that leads to doubts on the genealogical affiliation of these languages. However, the non-Austronesian component in lexicon and grammar is considerably large, going beyond some instances of borrowing. This amount of non-AN features suggest a language mixing based on long-term bilingualism with code-switching practices, at least up to a certain degree.

Additional evidence for the historic presence of speakers of unrelated languages, especially in the Lamaholot and Kedang areas, are irregularities in the lexeme sets. In the cognate sets and similarity sets, listed in section 5.2, 5.3 and 6.3, there are 9 sets with irregular reflexes attested in individual subgroups. These irregularities are: (1) sporadic consonant changes in the first person pronouns '1SG', '1PL.EXCL', '1PL.INCL', (2) sporadic lenition of PFL *b > v in the sets 'thousand', 'woman' and 'tongue', (3) unexpected non-occurrence of the sound change *s > h in the set 'salt' and the sound change *d > dʒ in the set 'how much', and (4) the sporadic change of *t > d in the set 'forest' (for details see Fricke 2019a, 144–148). These irregular reflexes are mainly attested in Kedang or Lamaholot. The Sika reflexes are largely regular.

8 Conclusions

In this chapter, I have shown that the vocabulary of the Austronesian Flores-Lembata languages, as well as of their ancestor Proto Flores-Lembata, is to varying degrees of non-Austronesian origin. While PFL has only little lexical

items which do not have an Austronesian origin, this amount raises to about 50% of the lexicon in the Lamaholot varieties.

I have argued that this mixed lexicon emerged out of bilingual speech communities which were fluent in Lamaholot as well as in at least one unknown non-Austronesian language. These non-AN languages were most likely typologically very similar to the neighbouring Timor-Alor-Pantar languages. Lexically, however, no clear relation to the TAP languages could be established. As there are three Lamaholot subgroups today which share most of the non-AN vocabulary, and show regular sound correspondence in this added vocabulary, the vocabulary was most likely added before the establishment of these three subgroups based on regular sound changes. However, it must have been added after the split up of PFL because only very little of it can be reconstructed to PFL. Also the added non-AN grammatical features in the Lamaholot subgroups support this scenario.

This case study is an example of how a language contact scenario in the past can be reconstructed by analysing non-inherited features in grammar and lexicon. Investigating both, lexicon and grammar, draws a more detailed picture, in the case of Lamaholot, a bilingual community where code-switching was a common, if not the main way of communication.

Acknowledgements

This chapter is based on Part II of the author's dissertation (Fricke 2019a). I would like to thank Marian Klamer, Owen Edwards and Francesca Moro for their comments on earlier versions of this chapter, and Naonori Nagaya and Alex Elias for their critical reviews which raised very good additional points. Furthermore, this work would not have been possible without the Dutch Research Council (NWO)'s VICI grant for the project *Reconstructing the past through languages of the present: The Lesser Sunda Islands* by Prof. dr. Marian Klamer (project number: 277-70-012).

Appendix

List of Basic Concepts

The classification as basic concepts is based on the Leipzig-Jakarta Basic Vocabulary list (Tadmor, Haspelmath, and Taylor 2010, 238–241) with my own extensions, concerning in particular regionally relevant concepts. In total, the following 192 concepts have been classified as basic for the purpose of this study:

1pl exclusive; 1pl inclusive; 1sg; 2sg; 3pl; 3sg; all; ant; ash, dust; back; banana; bathe; betel vine; big; bird, chicken; bite; bitter; black, dirty; blood; blow; body, self; body hair; bone, seed; breast, milk; burn, shine; child, small; cloud, fog; coconut; come; cry; cut, kill; day, sun; deaf; die; dog; dream; drink; drop, fall from above; dry, thirsty; ear; eat; egg; eight; excrements; eye; fall from above, descend; fall over; far, long; fat; fingernail; finished; fire; fish; flat, below, river; flower, blossom; fly; fly (n.); flying fox; foot, leg; forehead; forest; four; fruit, betelnut; full; give; go; good; grass, bush; hair; hand, arm, five; head; headlice; hear; heart; heavy; here; hide; hillwards, above; hit; horn; hot; house; how much, how many; how?; hungry; inside, deep; inside, liver, house; itchy; knee; knife; know; laugh; leaf; lie down (non-human); liver; man; many; meat, flesh; meeting house; moon, market; mosquito; mother; mountain; mouth; name; narrow; navel; near; neck; needle; new; night; nine; no, not; nose; old; one, alone; person; pound; price, bride price, expensive, buy; rain; rat; rattan; red; rice; road; roof rafter; root; rope; round; run; salt; sand, soil; say; say; sea, wave; see; seven; short; sick, painful; sit; six; skin, bark of tree; sky; sleep, lie down; smoke; snake; soil; spit; stand; star; stomach, belly; stone; storage house, barn; suck; sugar palm; sugarcane; sun; sweet; swim; tail; teeth; ten; that; thatch for roofing; thatched roof; thick; this; thousand; three; tie; tongue, say; tree, wood; two; vomit; wake someone up; wake up; walk; wash, bathe; water; what; where; white; who; wide; wife, husband; wind; wing; woman, sister; yellow; yesterday; younger sibling.

References

- Akoli, Marcelinus. 2010. "Lamaholot Dialect Variations." Master thesis, Melbourne: Monash University.
- Arndt, Paul. 1931. *Grammatik der Sika-Sprache*. Ende: Arnoldus.
- Arndt, Paul. 1937. *Grammatik der Solor-Sprache*. Ende: Arnoldus.
- Bellwood, Peter. 2017. *First Islanders: Prehistory and Human Migration in Island Southeast Asia*. Hoboken: John Wiley & Sons.
- Blust, Robert. 2013. *The Austronesian Languages*. Canberra: Pacific Linguistics.
- Blust, Robert, and Stephen Trussel. 2010. "The Austronesian Comparative Dictionary: Web Edition." <http://www.trussel2.com/acd/>.
- Bolscher, H. 1982. *Tati Lalang Sara Sikka*. Maumere: Lela.
- Doyle, Matthew. 2010. "Internal Divisions of the Flores-Lembata Subgroup of Central Malayo-Polynesian." Master thesis, Leiden: Leiden University.
- Edwards, Owen. 2016. "Parallel Sound Correspondences in Uab Meto." *Oceanic Linguistics* 55 (1): 52–86.

- Edwards, Owen. 2018. "Top-Down Historical Phonology of Rote-Meto." *Journal of the Southeast Asian Linguistics Society* 11 (1): 63–90.
- Edwards, Owen. 2019. "Reintroducing Welaun." *Oceanic Linguistics* 58 (1): 31–58.
- Edwards, Owen. 2021. *Rote-Meto Comparative Dictionary*. Canberra: Australian National University Press.
- Edwards, Owen. this volume. "Phonological Innovation and Lexical Retention in the History of Rote-Meto." In *Traces of contact in the lexicon: Austronesian and Papuan studies*, edited by Marian Klamer and Francesca Moro. Leiden: Brill.
- Elias, Alexander. 2018. "Lio and the Central Flores Languages." Master thesis, Leiden: Leiden University.
- Elias, Alexander. 2020. "Are the Central Flores Languages Really Typologically Unusual?" In *Austronesian Undressed: How and Why Languages Become Isolating*, edited by David Gil and Antoinette Schapper. Amsterdam: John Benjamins.
- Ewing, Michael, and Marian Klamer, eds. 2010. *East Nusantara: Typological and Areal Analyses*. Canberra: Pacific Linguistics.
- Fernandez, Inyo. 1977. "Bahasa Lamaholot Ile Mandiri." Bachelor thesis, Yogyakarta: Universitas Gadjah Mada.
- Fernandez, Inyo. 1996. *Relasi historis kekerabatan bahasa Flores: Kajian linguistik historis komparatif terhadap sembilan bahasa di Flores*. Ende: Nusa Indah.
- Fricke, Hanna. 2014. *Topics in the Grammar of Hewa: A Variety of Sika in Eastern Flores, Indonesia*. Munich: Lincom Europe.
- Fricke, Hanna. 2017. "The Rise of Clause-Final Negation in Flores-Lembata, Eastern Indonesia." *Linguistics in the Netherlands* 34 (November): 47–62.
- Fricke, Hanna. 2019a. *Traces of language contact: The Flores-Lembata languages in eastern Indonesia*. Amsterdam: LOT Publications.
- Fricke, Hanna. 2019b. Central Lembata Corpus. Nijmegen: The Language Archive, Max Planck Institute for Psycholinguistics. <https://hdl.handle.net/1839/c12bb123-1c4b-47cf-8d21-f185daf0557e>.
- Fricke, Hanna. 2019c. *Kamus Tiga Bahasa. Atadei Demon—Indonesia—Inggris*. Kupang: Unit Bahasa & Budaya (UBB).
- Grangé, Philippe. 2015. "The Lamaholot Dialect Chain (East Flores, Indonesia)." In *Language Change in Austronesian Languages: Papers from 12-ICAL*, edited by Malcolm Ross and Wayan Arka. Canberra: Asia-Pacific Linguistics.
- Haurholm-Larsen, Steffen. 2016. "A Grammar of Garifuna." Dissertation, Bern: University of Bern.
- Holton, Gary, and Marian Klamer. 2017. "The Papuan Languages of East Nusantara and the Bird's Head." In *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*, edited by Bill Palmer, 569–640. Berlin: Mouton de Gruyter.
- Hudjashov, Georgi, Tatiana Karafet, Daniel Lawson, Sean Downey, Olga Savina, Herawati Sudoyo, Stephen Lansing, Michael Hammer, and Murray Cox. 2017. "Complex

- Patterns of Admixture across the Indonesian Archipelago." *Molecular Biology and Evolution* 34 (10): 2439–2452.
- Kaiping, Gereon, Owen Edwards, and Marian Klamer. 2019. "LexiRumah 3.0.0. Leiden: Leiden University Centre for Linguistics." <https://lexirumah.model-ling.eu/lexirumah/>.
- Kaiping, Gereon, and Marian Klamer. 2018. "LexiRumah: An Online Lexical Database of the Lesser Sunda Islands." *PLoS One* 13 (10): e0205250.
- Karafet, Tatiana, Brian Hallmark, Murray Cox, Herawati Sudoyo, Sean Downey, Stephen Lansing, and Michael Hammer. 2010. "Major East–West Division Underlies Y Chromosome Stratification across Indonesia." *Molecular Biology and Evolution* 27 (8): 1833–1844.
- Keraf, Gregorius. 1978. *Morfologi Dialek Lamalera*. Ende: Arnoldus.
- Klamer, Marian. 2011. *A Short Grammar of Alorese (Austronesian)*. Munich: Lincom Europe.
- Klamer, Marian. 2012. "Papuan–Austronesian Language Contact: Alorese from an Areal Perspective." In *Melanesian Languages on the Edge of Asia: Challenges for the 21st Century*, edited by Marian Klamer and Nicholas Evans, 72–108. Honolulu: University of Hawai'i Press.
- Klamer, Marian, Ger Reesink, and Miriam van Staden. 2008. "East Nusantara as a Linguistic Area." In *From Linguistic Areas to Areal Linguistics*, edited by Pieter Muysken, 95–149. Amsterdam: John Benjamins.
- Krauß, Daniel. 2016. "A Brief Grammar of the Eastern Atadei Language of Lembata, Indonesia." *Linguistik Indonesia*, 34 (2): 113–128.
- Kroon, Yosep. 2016. "A Grammar of Solor Lamaholot: A Language of Flores, Eastern Indonesia." Dissertation, Adelaide: University of Adelaide.
- Lewis, Douglas, and Charles Grimes. 1995. "Sika." In *Comparative Austronesian Dictionary*, edited by Darrell Tryon and Malcolm Ross, 601–609. Berlin: Mouton de Gruyter.
- Michels, Marc. 2017. "Western Lamaholot. A Cross-Dialectal Grammar Sketch." Master thesis, Leiden: Leiden University.
- Muysken, Pieter. 2010. "Scenarios for Language Contact." In *The Handbook of Language Contact*, edited by Raymond Hickey. Malden: Wiley-Blackwell.
- Nagaya, Naonori. 2011. "The Lamaholot Language of Eastern Indonesia." Dissertation, Houston: Rice University.
- Nishiyama, Kunio, and Herman Kelen. 2007. *A Grammar of Lamaholot, Eastern Indonesia: The Morphology and Syntax of the Lewoingu Dialect*. Munich: Lincom Europe.
- Pampus, Karl-Heinz. 1999. *Koda Kiwā: Dreisprachiges Wörterbuch Der Lamaholot-Sprache, Dialekt von Lewolema*. Stuttgart: Deutsche Morgenländische Gesellschaft Steiner.
- Pareira, Mandalangi, and Douglas Lewis. 1998. *Kamus Sara Sikka Bahasa Indonesia*. Ende: Nusa Indah.

- Pawley, Andrew. 2005. *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*. Canberra: Pacific Linguistics.
- Reid, Lawrence A. 1994. "Possible Non-Austronesian Lexical Elements in Philippine Negrito Languages." *Oceanic Linguistics* 33 (1): 37–72.
- Robinson, Laura. 2015. The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara. In Malcolm Ross & Wayan Arka (eds.), *Language change in Austronesian languages: Papers from 12-ICAL*, 19–33. Canberra: Asia-Pacific Linguistics.
- Rosen, Joan. 1977. "Partikel kata kerja dan anapfor kosong dalam bahasa Maumere: Paper pada konferensi bahasa daerah di Jakarta." In Konferensi bahasa daerah (24–29 Januari 1977). Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- Rosen, Joan. 1986. "Phonemes, Verb Classes and Personal Endings in Maumere." *NUSA: Linguistic Studies of Languages in and around Indonesia*, Miscellaneous of Indonesian and other languages in Indonesia Part VIII, 25: 39–69.
- Ross, Malcolm. 1992. "The Sound of Proto-Austronesian: An Outsider's View of the Formosan Evidence." *Oceanic Linguistics* 31 (1): 23–64.
- Ross, Malcolm. 2013. "Diagnosing Contact Processes from Their Outcomes: The Importance of Life Stages." *Journal of Language Contact* 6 (1): 5–47.
- Samely, Ursula. 1991. *Kedang (Eastern Indonesia): Some Aspects of Its Grammar*. Hamburg: H. Buske.
- Samely, Ursula, and Robert Barnes. 2013. *A dictionary of the Kedang language: Kedang-Indonesian-English*. Leiden: Brill.
- Schapper, Antoinette. 2015. "Wallacea, a Linguistic Area." *Archipel* 90: 99–151.
- Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. 2010. "Borrowability and the Notion of Basic Vocabulary." *Diachronica* 27 (2): 226–246.
- Thomason, Sarah. 2001. *Language Contact*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah. 2007. "Language Contact and Deliberate Change." *Journal of Language Contact* 1 (1): 41–62.
- Thomason, Sarah, and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Wolff, John. 2010. *Proto-Austronesian Phonology with Glossary: Volume 1*. Ithaca: Cornell Southeast Asia Program Publications.

Entwined Histories: The Lexicons of Kawaimina and Maka Languages

Antoinette Schapper and Juliette Huber

1 Introduction

Research into Austronesian-Papuan language contact in eastern Indonesia has to date mainly centred on identifying Austronesian lexical influences on Papuan languages and Papuan morphosyntactic influences on Austronesian languages (see Schapper forthcoming for a recent overview). In Timor also, linguists have commented on the large number of Austronesian etyma found in the Papuan languages of the region ever since they were identified as being non-Austronesian in the first half of the 20th century (see Schapper 2020a for a history and references). A similar picture of profound Austronesian influence on Papuan-speaking populations emerges from anthropological research: according to McWilliam (2007), for instance, the Papuan-speaking Fataluku of Timor-Leste are culturally so thoroughly Austronesian that he characterizes them as “Austronesians in linguistic disguise”. He adduces numerous culturally significant lexemes borrowed from Austronesian languages to support his characterisation. In this paper, we draw attention to a different scenario and examine how lexical transfer has potentially occurred from Papuan languages into Austronesian languages. We seek to highlight the need to go beyond the Austronesian-Papuan dichotomy in characterising the lexical histories of languages in Timor, showing that many lexemes that are shared between neighbouring Austronesian and Papuan languages resist classification as belonging to one or the other.

At the far eastern end of Timor, an expansive and influential Papuan-speaking community lives alongside smaller Austronesian-speaking groups. The Papuan language in question is Makasae. Together with its close linguistic relative Makalero, Makasae belongs to the Eastern Timor subgroup of the Timor-Alor-Pantar (TAP) family. Their Austronesian neighbours are a small group of four closely related languages: Waima’a, Naueti, Kairui and Midiki, known collectively since Hull (1998) as “Kawaimina” languages.¹ The existing literat-

1 Two reviewers flagged problems with this name. In the absence of any alternatives in the published literature, we maintain its use here.

ure suggests that individual Kawaimina languages have been impacted by the neighbouring Papuan languages to different degrees. According to Hajek and Himmelmann (2006: 10), hardly any Makasae loans are found in Waima'a. Closely related Naueti, on the other hand, is suggested by Hull (2004: 34) to have "a strong presence of Papuan lexical elements". In the absence of comparative lexical studies of the languages in question, however, these claims are impossible to verify.

In this paper, we assess the evidence for lexical borrowing from the Papuan languages of Eastern Timor, in particular the Maka languages (as we shall collectively refer to Makasae and Makalero), into their Kawaimina neighbours. We highlight the existence of multiple lexemes with etymologies at different levels in the Timor-Alor-Pantar family that are also present in the Kawaimina languages. At the same time, we draw attention to the presence of lexemes and sub-lexical elements shared between either individual Maka and Kawaimina languages or sets of them. In some cases, the original source for these shared lexemes is impossible to determine and in others mutual borrowing from a third, unknown source language seems likely. Finally, we draw attention to evidence that there are many lexemes in the Kawaimina languages whose phonological shape points to recent borrowing through Makasae. Taken together, we suggest that the evidence indicates that the Maka and Kawaimina languages have more entwined histories and more complex patterns of borrowing and influence between them than has been previously made clear.

This paper is structured as follows: in section 2, we introduce the language groups involved in the Kawaimina-TAP contact situation. Section 3 presents a detailed discussion of the lexical entwinement of the Maka and Kawaimina languages at different levels. In section 4, we highlight the complexity of the contact situation by zooming in on the case of *-kai*, a suffix which has been attributed to the TAP language Makasae by Veloso (2016), but whose history appears much more complex. Section 5 concludes.

2 Language Setting

Timor-Leste is home to some 20 language varieties (Figure 6.2). The majority belong to the Austronesian family. The remaining handful of languages are part of the Papuan Timor-Alor-Pantar (TAP) family, a small group of some 30 languages spoken on Timor and the adjacent Indonesian islands (Schapper, Huber and Engelenhoven 2012; Schapper, Huber and Engelenhoven 2014). In this paper, we focus on the Maka languages, a low-level subgroup of TAP con-

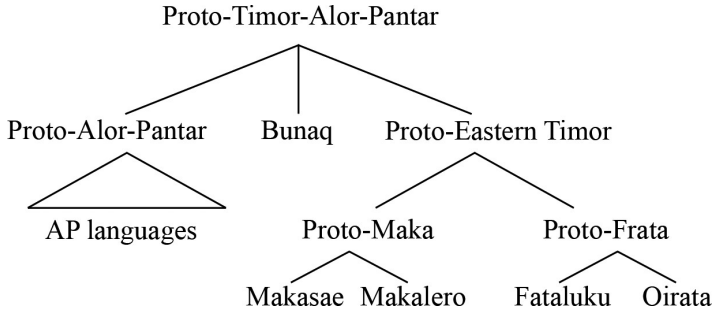


FIGURE 6.1 The relations of the Papuan languages of the Timor-Alor-Pantar family

sisting of two languages, Makalero and Makasae, spoken in the eastern part of East Timor. Figure 6.1 illustrates the position of the Maka group within TAP. Together with Fataluku on the island's eastern tip (Figure 6.2), and Oirata on Kisar island just to the north of Timor's eastern tip, they make up the Eastern Timor branch of the TAP family.

With some 130,000 speakers (General Directorate of Statistics 2015), Makasae (ISO 639-3 code: mkz) is the largest of the Eastern Timor languages, and indeed the largest TAP language. It is also Timor-Leste's third largest language. Spoken by a population of less than 8,700, its closest relative Makalero (ISO 639-3 code: mjb) is significantly smaller. Makalero was only assigned an ISO 639-3 code in the 2015 edition of *Ethnologue* (Lewis, Simons and Fennig 2015); in older sources, it is frequently treated as a dialect of Makasae. In fact, our knowledge of the extent of dialect differences within Makasae is still limited. For instance, a variety known as Sa'ani, which is spoken between Makalero and Makasae, is variably treated as a separate language or a Makasae dialect. Given that Sa'ani remains undescribed, either assessment must be considered arbitrary to a degree, and the same may be true of many other Makasae dialects (cf. Huber 2017: 269).

The Kawaimina languages are a group of closely related Austronesian varieties spoken to the west of the Maka languages. The term Kawaimina is an acronym coined in Hull (1998: 102) as a cover term to refer to four varieties that subgroup together: Kairui, Waima'a, Midiki, and Naueti (Figure 6.2). With a total of 21,227 speakers (General Directorate of Statistics 2015), Waima'a (ISO 639-3 code: wmh) is the largest of the group, followed by Naueti (ISO 639-3 code: nxa). Midiki and Kairui (ISO 639-3 code: krd), the latter with less than 4,000 speakers, are the smallest of these languages. The Kawaimina languages can be tentatively assigned to the hypothesised Timor-Babar subgroup (Edwards 2018, 2021). The Timor-Babar subgroup includes most of the Aus-

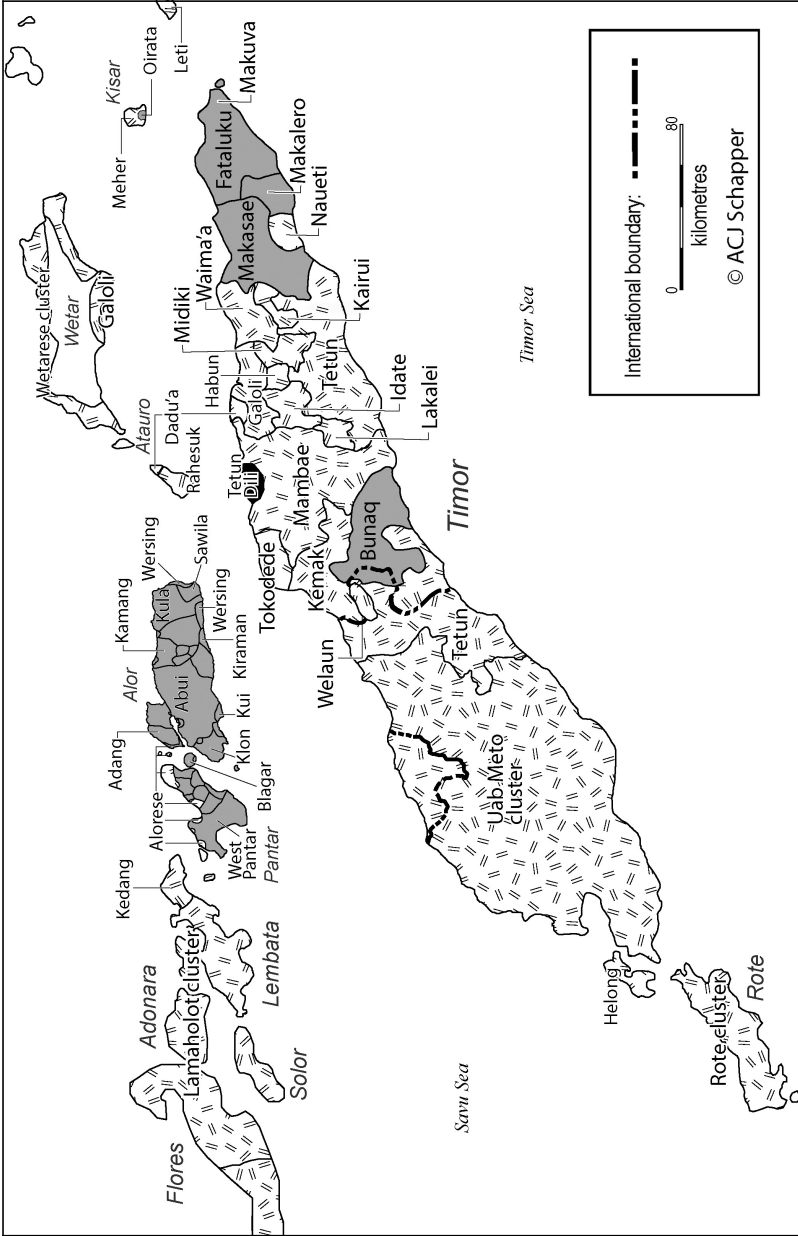


FIGURE 6.2 Linguistic map of Timor and surrounds
ADAPTED FROM SCHAPPER 2020B

tronesian languages of Timor as well as those of Wetar and the islands to the east up until Babar. Within Timor-Babar, the Kawaimina languages, along with Tetun, Habun, Galolen and Lakalei, make up the East Timor group (Edwards 2018: 88).

The existing documentation on Waima'a and Naueti shows that most speakers of these languages are highly multilingual, and the same can be assumed for Kairui and Midiki speakers. Knowledge of Timor-Leste's lingua franca Tetun is widespread. In 2006, Hajek and Himmelmann (2006: 10–11) reported that Waima'a was under increasing pressure from that language as many parents chose to speak to their young children in Tetun rather than Waima'a. Makasae, a vital and important regional language, also plays a role: according to Correia (2011: 388, cf. 6), speakers of Waima'a, Naueti, Midiki and Makasae living close to the language boundary "usually have a good command of each other's vernacular". Hajek and Himmelmann (2006: 10) confirm that knowledge of Makasae is widespread in Waima'a-speaking areas inside or on the edge of the region's major urban centre, Baucau. However, they find little knowledge of Makasae in Caisido, a Waima'a-speaking village less than 10 kilometres to the west of the city (2006: 10). They note that this is unexpected given the long-standing close contact between the two groups and speculate that "Makasae-Waima'a bilingualism must have been much more widespread in the past". In the Naueti language area, Veloso (2016: 5–6) reports that most men over 20 in the Uatolari subdistrict have "at least excellent negotiation skills in Makasae", and notes that the Naueti dialect of that region is characterized by a comparatively strong presence of Makasae loans.

For this paper, we made use of all available sources on the Kawaimina and Maka languages. Among the Kawaimina languages, we focus mostly on Waima'a and Naueti since they are the best documented. For Waima'a, we looked at the sketch grammar (Bowden et al. 2006), the Waima'a-English-Tetun-Malay glossary (Belo et al. 2005), and the Waima'a Toolbox files, all of which are accessible in the DoBeS archive. For Naueti, we used the lexical data published in Arnaud and Campagnolo (1998), Saunders (2003) and, most recently, Veloso (2016). Kairui and Midiki are both poorly documented. We only had access to the word lists of Dawson (2014) available in the PARADISEC archive, and the comparative Kawaimina Swadesh list provided in Veloso (2016). For Makalero we used Huber (2011) and Pinto (2004). Makasae sources differ depending on the dialect. For Makasae Ossu, we used Brotherson (2003), Huber (2005, 2008) and Jessé Fogaça (pers. comm), for Makasae Baucau Fogaça (2015), Hull (2004) and Huber (fieldnotes), for Makasae Laga Correia (2011), for Makasae Fatumaka Arnaud and Campagnolo (1998), Náchér (2012) and Ribeiro (2005), and for Makasae Ossorua Sarmiento (2005). Finally, reconstructed Pro-

to Malayo-Polynesian (PMP) forms are taken from Blust and Trussel's (2020) online *Austronesian comparative dictionary* (ACD).

3 Lexicon Shared between Kawaimina and Eastern Timor Languages

As noted in section 1, some authors have commented on the conspicuous absence of lexical borrowings from Papuan languages in particular Kawaimina languages, while others have asserted the presence of a strong Papuan lexical element. In this section, using our ongoing historical work on the TAP family (e.g., Schapper, Huber and Engelenhoven 2012, 2014, Usher and Schapper 2022), we re-examine the question of the Papuan lexical element in the Kawaimina languages.

The need for closer study of this question became apparent to us when we were conducting a detailed study of Austronesian borrowings in the Eastern Timor languages (preliminary results reported in Schapper and Huber 2019; the whole study is being prepared for publication elsewhere). In examining the Kawaimina languages we noted, on the one hand, multiple lexemes with an apparent TAP origin and, on the other hand, multiple lexemes shared between Maka and Kawaimina languages for which the directionality of borrowing was not readily apparent. We also became aware of the situation whereby Austronesian etyma were borrowed into Maka languages and then back into Kawaimina languages from Maka languages (examples first mentioned in Schapper forthcoming). In this paper, we limit ourselves to discussing lexemes that fit into these categories. It is beyond the scope of this paper to discuss lexemes that appear widely in both the Austronesian and Papuan languages of the Timor region. Most of the cases of this kind represent Austronesian borrowings into Papuan languages, but some can be analysed as early Papuan borrowings whose reflexes then became widely dispersed in Austronesian languages (see Schapper forthcoming for some potential examples).

The remainder of this section is structured around the level of reconstructability within the TAP family shown by lexemes appearing in Kawaimina languages. Section 3.1 considers borrowings of TAP etyma in Kawaimina languages, while section 3.2 looks at borrowings of Eastern Timor (ET) etyma in Kawaimina languages. Section 3.3 discusses lexical form-meaning pairings shared between Maka and Kawaimina languages, while Section 3.4 considers lexicon shared between Makasae and one or more Kawaimina languages. In these last two sections, we draw attention to the complexity of the contact situation by highlighting that in many cases the direction of the borrowing is unclear. Additionally, we show that just because a lexeme has an Austronesian etymology, it

should not be assumed that the immediate direction of borrowing is from an Austronesian language into a Papuan one.

Throughout this section morpheme boundaries in lexemes that we mark reflect our own, often historical, analysis. Makasae data is provided with the dialect name. Where no dialect is identified, the word is pan-dialectal.

3.1 TAP Etyma in Kawaimina Languages

The least problematic lexemes to identify as borrowings from TAP languages are those which have established Proto Timor-Alor-Pantar (PTAP) etymologies. Thus far we have identified 7 PTAP etyma that have been borrowed into Kawaimina languages. These are set out in Table 6.1. PTAP etyma and their supporting reflexes are drawn from Usher and Schapper (2022) and Schapper (in preparation). Makasae and Makalero reflexes of each PTAP reconstruction are presented in a separate column for ease of comparison with the forms in Kawaimina languages. In the remainder of this section, we discuss each of the borrowings in turn.

TABLE 6.1 TAP etyma in Kawaimina languages

Source	Makasae-Makalero	Kaiwamina
PTAP * <i>muni</i> ‘smell, emit a smell’ > Fataluku <i>mini-k</i> ‘nose’, Teiwa <i>mun</i> ‘smell, stink’, Nedebang <i>-aminni</i> ‘stink, smell bad’, Klon <i>muin</i> ‘nose, smell’, Wersing <i>-muiŋ</i> , Sawila <i>-muni</i> ‘smell, stink’, Kula <i>-muni</i> ‘fragrance’ (Schapper in prep.)	Makasae Ossu, Fatumaka <i>muni</i> ‘kiss’ Makasae Laga, Baucau <i>ate-muni</i> ‘sandalwood’ Makasae Fatumaka, Baucau <i>muni-ri</i> ‘smell good’ Makalero <i>muni-?</i> ‘kiss, smell at’, <i>ate-muni</i> ‘sandalwood’	Waima’a <i>muni</i> ‘kiss’, <i>wau-muni</i> ‘fragrant’, <i>daka-muni</i> ‘k.o. basil’, <i>hae-wau-muni</i> ‘citronella (<i>Cymbopogon citratus</i>)’ Naueti <i>muni</i> ‘kiss’, <i>wou-muni</i> ‘fragrant’, <i>kai-wou-muni</i> ‘sandalwood (<i>Santalum</i> sp.)’, <i>hae-wou-muni</i> ‘lemongrass (<i>Cymbopogon</i> sp.)’
PTAP * <i>kaku</i> ‘younger relative’ > Fataluku <i>kaʔu-sila</i> ‘be small’, <i>kaʔu-kisa</i> ‘small’, Bunaq <i>kau?</i> , Blagar <i>kaku</i> ‘sibling of same gender, friend’, Reta <i>kaku</i> ‘friend’, Kamang <i>-kak</i> , Wersing <i>kaku</i> , Sawila <i>ka:ku</i> ‘younger sibling’ (Schapper in prep.)	Makasae Baucau, Laga <i>kaʔu</i> ‘small’ Makasae Laga <i>kaʔu-kaʔu</i> ‘very small’ Makasae Fatumaka <i>kau</i> ‘small’ Makalero <i>kaʔu</i> ‘small’	Waima’a <i>kaʔu</i> ~ <i>kaʔu-n</i> ‘small’

TABLE 6.1 TAP etyma in Kawaimina languages (*cont.*)

Source	Makasae-Makalero	Kaiwamina
PTAP *an[u,i]ŋ ‘person’ < Bunaq <i>en</i> , Kui <i>anin</i> ‘person’, Kamang <i>anij</i> ‘human numeral classifier’, Wersing <i>anij</i> , Sawila <i>anij</i> ‘person’ (Schapper in prep.)	Makasae <i>anu</i> ‘person’	Waima’a <i>anu-atu</i> ~ <i>anu-uta</i> ‘female, woman, wife’ Naueti <i>ona-ata</i> ‘female, woman’ Midiki <i>anu-wata</i> ‘woman’ Kairui <i>anu-ota</i> ‘female, woman’
PTAP *[s,t]abur ‘crab’ > Fataluku <i>capu-ku</i> ~ <i>capu-ke</i> , Bunaq <i>sawar</i> , Teiwa <i>tafar</i> , Nedebang <i>tafi</i> , Reta <i>tubal</i> , Blagar <i>tubar</i> , Klön <i>tbur</i> , Kui <i>tabui</i> , Abui <i>tafui</i> , Kafoa <i>tafoi</i> , Kamang <i>tapui</i> , Sawila <i>sapar</i> ‘crab’ (Schapper in prep.)	Makasae Laga <i>sabi</i> ‘crab’ Makasae Baucau <i>sabi-kai</i> , <i>sabi-leki</i> ‘crab’ Makasae Fatumaka <i>sabi-li</i> , <i>sabi-lai</i> ‘crab’	Waima’a <i>sabu</i> ‘crab’ Naueti <i>sabu</i> , <i>sabu-luki</i> ‘crab’
PTAP *ina ‘eye’ > Makasae <i>ina</i> , Makalero <i>ina</i> , Fataluku <i>ina</i> , Oirata <i>ina</i> , Kamang <i>-ŋ</i> , Abui <i>-ieŋ</i> , Kafoa <i>-eŋ</i> , Kafoa <i>-eŋ</i> , Kui <i>-en</i> , Klön <i>-en</i> , Blagar <i>-eŋ</i> ‘eye’ (Usher and Schapper 2022)	Makasae Laga, Baucau, Ossu <i>kina</i> ‘show’	Naueti <i>kina</i> ‘show’ Waima’a <i>kine</i> ‘show, demonstrate’
PTAP *iri ‘urine’ > Makasae <i>iri</i> , Fataluku <i>iri</i> , Oirata <i>iri</i> , Blagar <i>ir</i> , Western Pantar <i>jir kaka</i> ‘urine’ (Schapper in prep.)	Makasae Laga, Baucau, Ossorua <i>kiri</i> ‘urinate’	Naueti <i>kiri</i> ‘urinate’
PTAP *madel ‘bat, flying fox’ > Fataluku <i>maca</i> , Oirata <i>maŋa</i> , Teiwa <i>maŋi</i> , Nedebang <i>marra</i> ‘bat’, Western Pantar <i>madde</i> ‘k.o. small bat’, Klön <i>mdel</i> , Kafoa <i>marel</i> , Abui <i>marel</i> , Kamang <i>matei</i> ‘bat’ (Usher and Schapper 2022)	–	Waima’a <i>mada</i> ‘bat’ Naueti <i>mada</i> ‘bat’ Midiki <i>mada</i> ‘bat’

The most straightforward example of borrowing of a TAP etymon into the Kawaimina languages involves PTAP *muni ‘smell, emit a smell’.² A reflex of this verb has been borrowed into Waima’a and Naueti as an independent verb with the sense ‘kiss’ and in compounds with the sense ‘fragrant, having smell’.³ Maka languages have a phonologically matching form *muni* with near-identical semantics, i.e., a transitive verbal use meaning ‘kiss’, plus uses in compounds—particularly in reference to plants—with the meaning ‘fragrant’. The other Papuan languages of Timor either have no reflex of this item, as is the case with Bunaq, or do not offer good phonological and semantic matches. Taken together, this strongly points to Maka languages being the immediate source of the borrowing of the TAP forms. A similar form-meaning pairing is widespread in Austronesian languages in the region and appears to reflect *mən̄ji(R) ‘fragrant’ (Edwards 2021). A relationship between this form and PTAP *muni seems possible. However, the *muni* forms discussed here in Kawaimina languages cannot be accounted for as reflexes of *mən̄ji(R). The regular reflex of *ə in Waima’a and Naueti is *e*, not *u*; this is seen in Waima’a *kai-kmeni* ‘sandalwood’, the latter part of which does reflect *mən̄ji(R).

Reflexes of PTAP *kaku ‘younger relative’ have also been borrowed into the Kawaimina languages as *kaʔu* ‘small’. This meaning and form is consistent with borrowing from a member of the Eastern Timor subgroup of TAP; PTAP *k regularly becomes Proto Eastern Timor (PET) *ʔ intervocalically and the semantic shift ‘younger relative’ > ‘small’ is found in all the Papuan languages of the Eastern Timor subgroup. It is notable that there are parallel borrowings of reflexes of PTAP *kaku to be found in several Austronesian languages of the Central Timor subgroup including Kemak *kaʔu* ‘young (of a baby)’, Mambae *kau* ‘younger sibling’. Schapper (forthcoming) argues that these forms are likely borrowings from a no-longer extant TAP relative of the nearby TAP language Bunaq, which has *kauʔ* ‘younger sibling’. Because Central Timor languages are not closely related to Kawaimina languages, parallel borrowings from different TAP languages provides the best explanation of the appearance of these forms in these disparate Austronesian languages.

Kawaimina languages have a word for woman in which a form *anu* ~ *ona* is compounded with *ata* ‘slave’. The first part of this compound is speculated

2 Note that this PTAP form exists alongside several other reconstructions with related meanings that share initial *mun. Most widespread is PTAP *muna ‘smell, fragrant’ > Sawila *muna* ‘fragrant, scent’, Kamang *mun* ‘smell, fragrant’, Kafoa *-mun* ‘smell’, Klon *mun* ‘perfumed’ (Schapper in prep.). In some cases, it is not possible due to segment loss to definitely assign a reflex to any one reconstruction.

3 Verbs of smelling often extend to kissing in Southeast Asian languages, see Schapper (2019).

here to be from Makasae *anu* ‘person’, reflecting PTAP *an[i,u]ŋ ‘person’; this Makasae form provides an exact phonological match to most of the Kawaimina forms.⁴ The semantic shift from ‘person’ in Makasae to ‘woman’ in Kawaimina is not large. The fact that the Makasae compounds with forms meaning ‘slave’ may reflect that women were considered subordinate or bonded to men in some way. That a word from Makasae is used as part of the compound for ‘woman’ may also suggest that women were traditionally sourced by Kawaimina-speaking groups from the Makasae and the word for them was imported alongside them. Similar to what already was observed with PTAP *kaku, a reflex of PTAP *an[i,u]ŋ ‘person’ has also been borrowed into the Central Timor language Welaun as *anu* ‘person’ (form from Edwards 2019:52). The form of this item also suggests that borrowing was from a no-longer extant TAP relative of the nearby TAP language, Bunaq, which has *en* ‘person’ (suggesting < pre-Bunaq **ani).

Kawaimina languages have also borrowed a reflex of PTAP *[s,t]abur ‘crab’. Makasae is unlikely to be the direct source of this borrowing, as the Kawaimina forms contain a final /u/, whereas the final segment is /i/ in Makasae. As seen in the previous example of Maka *anu* ‘person’, final /u/ in Maka languages would be expected to be borrowed as /u/ in Kawaimina languages. While /u/ is found in the final syllable of Makalero *dapuk* ‘crab’, the form does not match in other respects and is almost certainly a Fataluku borrowing. This suggests that the Kawaimina forms are either borrowed from pre-Maka before the change *u > i occurred in Makasae, or from another, now no longer extant TAP language which retained PTAP *u as u in this lexeme.

In two cases, TAP etyma are borrowed into Kawaimina languages from Makasae with an apparently verbalizing prefix *k-*. The roots of the lexemes ultimately go back to the PTAP nouns *ina ‘eye’ and *iri ‘urine’. The initial *k-* appears to derive verbs from these nominal roots in Makasae, but it is neither productive nor known from any other roots in the language. A derivational prefix *k-* is also not known from any other TAP language, though fossilized derivational suffixes are found in a number of TAP languages. For example, Makalero uses a suffix *iri-ʔ* ‘urinate’, cf. Fataluku *iris(-e)* ‘urinate’, Oirata *iris(-e)* ‘urinate’. In short, while the ultimate origin of the Makasae *k-* is unclear, the roots on which it appears are solid TAP etyma and they must have been borrowed into the Kawaimina languages.

4 The initial part of Naueti *ona-ata* shows lowering of *u* to *o* and metathesis of *a* and *o*. It is unclear why these changes occurred, but they do not obscure the obvious relationship between the Naueti form and the other Kawaimina languages, which have *anu-*. The irregularity of Waima'a *anu-atu* (in place of expected **anu-ata) appears to be a case of contamination from the first part of the compound.

Borrowing from an earlier or now lost language is also posited here for the Kawaimina lexemes for ‘bat’, *mada*. These represent borrowings of reflexes of PTAP **madel* ‘bat’. The immediate Papuan neighbours of these languages, Makasae and Makalero, do not attest reflexes of this PTAP form. However, a reflex of this PTAP etymon was certainly present at an earlier stage in Timor, as it is continued in the Frata languages (i.e., Fataluku *matsa* ~ *maca*, Oirata *maṭa* ‘bat’), and therefore must have been borrowed from an ancestor of Proto Maka or a lost relative of it.

TABLE 6.2 ET etyma in Kawaimina languages

Source	Makasae-Makalero	Kaiwamina
PET * <i>liri</i> ‘sprinkle, drizzle, flutter’ > Fataluku <i>liri~liri</i> ‘drizzle’, Oirata <i>aja liri~liri</i> ‘drizzling rain’	Makasae Baucau <i>liri</i> ‘scatter’ Makasae Fatumaka <i>liri</i> ‘flutter’, <i>liri~liri</i> ‘drizzle’ Makalero <i>liri</i> ‘sprinkle, add a small amount’	Naueti <i>liri-kiki</i> ‘suddenly scatter around’ Waima’a <i>liri</i> ‘scatter’
PET * <i>larun</i> ‘milipede, centipede’ < Fataluku <i>larun</i> ‘centipede’, Oirata <i>larun</i> ‘milipede’	Makasae Ossu <i>laru-ke</i> ‘centipede’ Makalero <i>laru-pi:k</i> ‘milipede’	Waima’a <i>saa-laru-kee</i> ‘centipede’ Naueti <i>laru-ke</i> ‘centipede’
PFRATA * <i>keko</i> ‘lobster’ > Fataluku <i>keko</i> ‘lobster, a sea creature like a prawn but big and purple’, Oirata <i>ke:k</i> ‘lobster’	–	Naueti <i>kako-raka</i> ‘big brown shrimp’
PET * <i>bora</i> ‘wrap, wind’ > Fataluku <i>poro~poro</i> ‘wrap’, Oirata <i>horo</i> ‘wrap’	Makasae Ossu <i>bora</i> ‘wrap, wind’ Makasae Fatumaka <i>bora</i> ‘wrap’	Waima’a <i>bura</i> ‘encircle, coil’ Naueti <i>boro-goe</i> ‘form a circle’
PMAKA * <i>g-ue</i> ‘around’	Makalero <i>pora</i> ‘wrap, wind’ Makasae <i>goe</i> ‘around’ Makalero <i>kue</i> ‘around’	

3.2 ET Etyma in Kawaimina Languages

Four instances of borrowings of etyma that arguably go back to Proto Eastern Timor can be identified in Kawaimina languages. These are set out with their known reflexes in Table 6.2 and discussed each in turn in what follows.

PET *liri ‘sprinkle, drizzle, flutter’ is supported by regular reflexes in all its daughters. To our knowledge, similar form-meaning pairings do not appear in any Austronesian languages other than the Kaiwaimina ones listed in Table 6.2. Our inference from this is that Waima’a and Naueti borrowed their lexemes *liri* from an ET language, most likely a Maka language since the meanings associated with their reflexes of proto ET *liri provide better matches to the Kawaimina ones than those found synchronically in the Frata languages. In a similar manner, proto ET *larun ‘milipede, centipede’ is well supported by regular reflexes in 3 of the 4 Eastern Timor languages. We hypothesise that this reconstruction is the ultimate source of the *laru-* formatives in Waima’a and Naueti. This formative is also found in the shared first part of Makasae and Makalero lexemes.

The initial element *kako-* in Naueti *kako-raka* ‘big brown shrimp’ bears a striking similarity in form and meaning to Proto FRATA (PFRATA) *keko ‘lobster’. Given that the form has some history in the Papuan languages, but is not known to occur in any other Austronesian languages, we assume that the directionality of borrowing here is from Papuan to Austronesian. It is, however, unclear whether this represents a case similar to the situation already described in section 3.1 for Kawaimina borrowings of PTAP *madel ‘bat’ whereby borrowing has taken place from an earlier or now lost relative of the ET languages. Instead, this may just represent a documentary gap in our knowledge of Maka languages where no term for lobster is recorded in any of the sources that we have consulted. Naueti borrowing from Fataluku directly is a logical possibility, but it is not a contact scenario that has been reported on in the literature thus far.

A more complex borrowing situation is represented by the fourth set of forms in Table 6.2. The second part of the Naueti form *boro-goe* ‘form a circle’ appears to be a borrowing of the Makasae reflex of PMAKA *g-ue ‘around’. The initial *g* of the Naueti form represents an original 3rd person prefix which has become entirely frozen in Makasae but still shows some productivity in Makalero (see Schapper, Huber and van Engelenhoven 2014: 108–110 for further discussion and illustration of this morphological pattern). The initial element of Naueti *boro-goe* ‘form a circle’ and Waima’a *bura* ‘encircle, coil’ are almost certainly linked with PET *bora ‘wrap, wind’. All appear to ultimately go back to PMP *balun ‘bind, bundle, wrap in cloth; death shroud cloth(ing)’ (ACD). However, the Naueti and Waima’a forms are not regular reflexes of this item, as we would expect PMP *b and *l to be reflected as *w* and *l* in both. This indicates that these items are borrowed from another, most likely Austronesian, language where *l had become *r* and *a had metathesized with *u, but no such language has been identified in the area today. Given that the forms in the ET languages

are regular reflections of PET *bora ‘wrap, wind’, it appears that the contact which led to this borrowing lays a long way back in time. The related forms in the Kawaimina languages may be borrowed from the ET languages, but need not be. Indeed, the differences in vowels in the forms presented by the ET and Kawaimina languages could be taken to suggest separate borrowing events into the Papuan and Austronesian languages. Interestingly, Naueti *boro-goe* reverses the order which a compound of this kind would be expected to have in Maka languages; typically, the directional element of a verbal compound in Maka languages occurs as the first morpheme of the compound, thus giving Makasae *goe-bora* ‘to wrap around, be wrapped around’ (attested for Makasae Ossu in Brotherson 2003: 133).⁵ The fact that the Naueti form does not reflect the order that would be found in Makasae lends support to the idea that the borrowing of *bora ‘wrap, wind’ occurred independently in Maka and Kawaimina languages. If Naueti borrowed *goe* together with *boro* from Makasae, then we would expect the order of elements to match that of the Makasae compound.

The idea of parallel borrowings in the ET and Kawaimina languages from a third, unknown language is challenging, but it is not without wider support in the data. Schapper and Huber (2019), for example, draw attention to an innovative numeral #kafo ‘eight’ which is widely in evidence across languages in parts of eastern Timor and southern Maluku. Table 6.3 sets out the forms that appear to belong to this set. What is striking here is the apparent parallel borrowings of slightly different forms into the various low-level subgroups of the two families. For instance, the forms in Maka languages suggest PMAKA *afo ‘eight’, but for PFRATA *kafa ‘eight’ has to be reconstructed. The correspondence of PMAKA ∅ and PFRATA *k is irregular. Among the Austronesian languages we can observe a similar lack of correspondence: the Kawaimina languages look to go back to a form *kaha where *h normally would reflect PMP *p; the forms in Wetar languages look to reflect *kaw where *w normally reflects PMP *b; Kisar-Luangic languages have forms that appear to reflect earlier *aβa, where medial *β would normally reflect PMP *b. This is as in Wetar languages, but the initial *k found in Wetar is lost. These different forms seem to point to replacement of original *k and *f with approximate sounds as the numeral diffused into each subgroup of the two families.

5 Examples of such constructions are plentiful in both Maka languages, e.g., Makasae *goe-le?u* ‘wrap, coil around (something)’, *goe-ria* ‘run around (something)’, Makalero *kue-lor* ‘fly around (something)’, *kue-la?a* ‘go around (something)’. See, e.g., Brotherson (2003) and Huber (2017: 299–303) for further information on locational and directional constructions in the Maka languages.

TABLE 6.3 Selected reflexes of #kafo ‘eight’ across eastern Timor and southern Maluku

Kaiwaimina languages (Austronesian)	Waima'a <i>kai-kaha</i> Midiki <i>kai-kaha</i>
Maka languages (TAP)	Makasae Baucau <i>afo</i> Makasae Fatumaka <i>afu</i> Makasae Ossu, Ossorua <i>apo</i> Makalero <i>afo</i>
Wetar languages (Austronesian)	Erai <i>kau</i> Tugun <i>kau</i>
Frata languages (TAP)	Fataluku <i>kafa</i> Oirata <i>kapa</i>
Luangic languages (Austronesian)	Kisar <i>wo-aa</i> Leti <i>βo-αβa</i> Luang <i>wo-awa</i> Wetan <i>wo-awa</i>
Babar languages (Austronesian)	Tela-Masbuar <i>wo-afu</i> Central Masela <i>wo-a</i> Emplawas <i>wo-auw</i>

The issue of shared borrowings is taken up again in the following section.

3.3 Shared Maka-Kawaimina Lexicon

In our study we found nearly a dozen items shared exclusively between both Maka languages and one or more Kawaimina languages. These are presented in Table 6.4. The appearance of related forms in Waima'a and Naueti clearly indicates that borrowing has taken place, but for most the original source of the borrowing is not clear.

The lexical forms in the Maka languages are for each set regular and could warrant a reconstruction of the lexeme to PMAKA. For the first five sets in Table 6.4, however, there are phonemes that indicate that the lexical history of these items within the Papuan languages is not deep. Instances of medial *k*, *g* and *d* in Maka languages occur only in innovative vocabulary. PMAKA medial **k*, medial **g* and medial **d* are not continuations of PTAP phonemes; medial PTAP **g* and **k* merge as **ʔ* in all Eastern Timor languages, while medial PTAP **d* merges with PET **t* in Maka languages. At the same time, most of the lexemes in Table 6.4 are not found in Austronesian languages outside of Kawaimina languages, and so a situation of Austronesian to Papuan borrowing cannot be assumed.

TABLE 6.4 Shared Maka-Kawaimina lexicon

Makasae-Makalero	Kawaimina
Makasae <i>bada</i> 'friend, colleague, relative'	Waima'a <i>bada</i> 'friend'
Makalero <i>pada</i> 'friend'	Naueti <i>bada</i> 'friend'
Makasae Baucau <i>nogo-nogo</i> 'stupid'	Waima'a <i>nogo-nogo</i> 'mad'
Makasae Laga <i>nogo-nogo</i> 'mad, crazy'	
Makalero <i>noko-noko</i> 'mad, crazy'	
Makasae Laga <i>gugu</i> 'silent, quiet'	Waima'a <i>gugu</i> 'mute'
Makasae Fatumaka <i>gugu</i> 'silent, quiet, calm'	Naueti <i>gugu laku-laku</i> 'silent', <i>gugu-lai</i> 'dumb, mute'
Makalero <i>kuku</i> 'silent'	
Makasae Laga, Fatumaka <i>tagar</i> 'step on, walk'	Naueti <i>taga</i> 'step'
Makalero <i>takar</i> 'walk, step'	
Makasae <i>lilibaka</i> 'butterfly'	Waima'a <i>lilibaka</i> 'butterfly'
Makalero <i>lilipaka</i> 'butterfly'	Naueti <i>liliboka</i> 'butterfly'
Makasae Baucau <i>gene</i> 'touch'	Waima'a <i>gene</i> 'touch', <i>gene-la</i> 'concerning'
Makasae Ossorua, Ossu <i>gene</i> 'hit'	
Makalero <i>kene</i> 'strike, hit the target'	Naueti <i>gene</i> 'touch', <i>gene-la</i> 'about'
Makasae <i>gali</i> 'back, around'	Naueti <i>gali-hila</i> 'look back'
Makalero <i>kali</i> 'back and forth, all around, upside down'	
Makasae Fatumaka <i>lai-koro</i> 'back'	Naueti <i>lai-buu</i> 'back'
Makalero <i>lai-pun</i> 'back'	
Makasae Laga, Baucau <i>nanu</i> 'great-grandparents, ancestors'	Naueti <i>nanu</i> 'great-grandparents'
Makalero <i>nanu</i> 'great-grandparents'	
Makasae Baucau, Laga, Ossu <i>lari</i> 'slope'	Waima'a <i>lari</i> 'hill'
Makasae Fatumaka <i>lari</i> 'mountain'	Naueti <i>lari</i> 'hill'
Makalero <i>lari</i> 'aslant, crooked', <i>larin</i> 'mountain'	

For the two sets that do appear to have Austronesian etymologies, there are problems with assuming that the directionality of borrowing is from Kawaimina to Maka languages. Makasae *gene* 'strike, hit the target' and Makalero *kene* 'strike, hit the target' reflect PMAKA *gene, but this form is likely a borrowing of a reflex of PMP *kəna 'be ensnared, caught in a trap; suffer, undergo, be struck by something; be entrapped or deceived; hit the mark' (ACD) (cf. Tetun *kona* 'strike, afflict'). The Kaiwaimina forms with *gene* cannot easily be seen as the source for the Maka borrowing, as we expect PMP *k to be reflected as *k* in both, PMP *ə to be reflected as *e*, and final PMP *a to be reflected as *a* in Naueti and *a*, but with sporadic raising to *o* following *u* and *e* following *i* in Waima'a.

Given three out of the four proto phonemes are wrongly reflected in Kawaimina languages, a direct connection between PMP *kəna and Kawaimina *gene* must be regarded as spurious. In a similar manner, Makasae *gali* ‘back, around’ and Makalero *kali* ‘back and forth’ regularly reflect PMAKA *gali, but seem to ultimately be a borrowing of a reflex of PMP *balik ‘reverse, turn around’. The first element of the Naueti form *gali-hila* appears to have the same origin as the Maka forms. Yet, it is not a regular reflection of PMP *balik ‘reverse, turn around’, because we would expect PMP *b to be reflected as Naueti *w*. Given the irregularities of the forms in the Kawaimina languages, we hypothesise that these are borrowings from an unknown Austronesian language into PMAKA that were then borrowed from Makasae, where PMAKA *g is maintained as *g*, into Naueti.⁶

In another two cases we have what appear to be irregular correspondences between forms in Maka and Kawaimina languages. Naueti *taga* ‘step’ lacks the final *r* found on the forms in the Maka languages (which appear to reflect PMAKA *tagar ‘step on, walk’). This suggests that Naueti did not borrow this form from a Maka language, but from another language with a related form where final *r* was lost. Similarly, Naueti *lari* ‘hill’ has an unexpected initial voiceless liquid, whereas the related forms in both Waima’a and Maka languages have plain *l*. In all other lexemes with a liquid considered here and in Schapper and Huber (2019), Naueti *l* has corresponded to *l* in Waima’a and the Maka languages. Naueti *lari* ‘hill’ suggests pre-Naueti ***h-lari* (see Schapper 2020b: 402–403 and Schapper and Zobel forthcoming for suggested pathways for at least some instances of voiceless sonorants in Kawaimina languages). Again, the irregularity in the Naueti form indicates that this lexeme was not the result of borrowing from recent contact arising through widespread knowledge of Makasae among the Naueti, but that it was borrowed at an earlier stage.

3.4 *Shared Makasae-Kawaimina Lexemes*

There are a sizeable number of lexemes shared between Makasae and one or more of the neighbouring Kawaimina languages. Table 6.5 sets out almost a dozen lexical forms shared exclusively between Makasae and Waima’a and/or Naueti. In all cases, the similarity of these form-meaning pairings is striking.

6 It appears that the initial velar stop on this item goes back to a 3rd person prefix in Maka languages. This is suggested by the fact that in Makalero initial *k* on this item is “removable” in the same contexts as a prefix *k*-, e.g., *ta-ali-laʔa* (RECP-back.and.forth-go) ‘get all mixed up with one another’.

TABLE 6.5 Shared Makasae-Kawaimina lexicon

Makasae	Kawaimina
Makasae Baucau, Ossu <i>tuka</i> 'behind'	Waima'a <i>tuko</i> 'back, behind'
Makasae Laga <i>gi tuka isi</i> 'behind, at the back of'	Naueti <i>tuka</i> 'backside'
Makasae Baucau <i>togu</i> 'deep, valley', <i>togun-u</i> 'deep'	Waima'a <i>togu</i> 'valley'
Makasae Ossu <i>togun-u</i> 'concave'	Naueti <i>togu</i> 'deep', <i>baʔa togu</i> 'valley'
Makasae Baucau <i>au-raga</i> 'coral', <i>meti raga</i> 'reef'	Waima'a <i>au-raga</i> 'coral'
Makasae Ossorua <i>au-raga</i> 'seaside tree'	
Makasae Ossorua <i>boboraka</i> 'spider'	Waima'a <i>babaraka</i> 'spider'
Makasae Ossu <i>babaraka</i> 'spider'	Naueti <i>boboraka</i> 'spider'
Makasae <i>rakalele</i> 'cheer'	Waima'a <i>rakalele</i> 'cheer'
	Naueti Uatolari <i>rakalele</i> 'cheer, acclaim'
Makasae Baucau <i>beu</i> 'can, be allowed'	Waima'a <i>beʔu</i> 'be able, can,
Makasae Laga, Ossu, Ossorua <i>beʔu</i> 'can, be allowed' ^a	may'
Makasae Baucau <i>waʔa</i> 'pip'	Waima'a <i>waʔa</i> 'seed, grain'
Makasae Laga <i>waʔa</i> 'seed, grain, pip, berry, seed'	Naueti <i>waʔa</i> 'pip'
Makasae Ossorua <i>wa</i> 'seed'	
Makasae Baucau <i>tutu-keu</i> 'mushroom'	Waima'a <i>tutu-keu</i> 'mushroom'
	Naueti <i>titi-kou</i> 'mushroom'
Makasae Baucau <i>ko</i> 'fart'	Naueti <i>ku</i> 'fart'
Makasae Baucau, Ossu <i>iru</i> 'bow'	Naueti <i>iru</i> 'bow'
Makasae <i>tumamae</i> 'firefly'	Naueti <i>tumamae</i> 'firefly'
Makasae Ossorua <i>nunu-bete</i> 'dolphin'	Naueti <i>nunu-bete</i> 'dolphin'

- a There is much variation in our Makasae sources regarding the rendering of the glottal stop phoneme. These are not reliable indicator of a dialectal differences. In both Maka languages, the glottal stop is often pronounced very faintly (e.g., Huber 2017: 274), especially in casual speech, but is heard much more clearly in careful speech. In many words, VʔV sequences and VV sequences may alternate not only within the same dialect but also in the same speaker (cf. *beu* ~ *beʔu* 'can, be allowed' and *waʔa* ~ *wa*: 'seed' in Table 6.5; see also *kau* ~ *kaʔu* 'small' in Table 6.1).

The matches over multiple lexical items would seem to exclude chance resemblance as an explanation.

The first five Makasae forms in Table 6.5 have phonemes that make clear they must be quite recent terms in the languages. As already mentioned with respect to Makasae forms in Table 6.4, medial *k* and *g* in Makasae occur only

in innovative vocabulary. But as discussed in the previous section, most of the lexemes are not found in Austronesian languages outside of Kawaimina languages, and so a situation of Austronesian to Papuan borrowing cannot be assumed. This leaves us with very little evidence to go on and for most lexemes we are forced to the conclusion that the lexemes are shared between Makasae and Kawaimina, but are of ultimately unknown origin. In what follows we comment on only the forms for which further discussion is possible.

Waima'a *babaraka* and Naueti *boboraka* 'spider' are shared with the Ossu and Ossorua dialects of Makasae, spoken in areas bordering the Waima'a and Naueti language regions. The Laga and Fatumaka dialects, somewhat further removed from the Kawaimina languages towards the east of the Makasae language area, have a different etymon *labarake*, translated as 'spider' and 'spiderweb', respectively. *Boboraka* ~ *babaraka* from the Ossorua and Ossu dialects of Makasae and *labarake* 'spider, spiderweb' in the Laga and Fatumaka dialects appear to share the same second element, *raka* ~ *rake*. While the first element of *labarake*, *laba*, is an Austronesian borrowing (cf. Tetun *labadain* 'spider', Waima'a *laba-dai* 'spiderweb' < PMP **lawaq* 'spiderweb' (ACD)), the origin of *raka* ~ *rake* is unclear. It is noticeable, however, that many languages in the region have four-syllable or longer terms for 'spider' (e.g., Leolima Kemak *busarabak*, Dadu'a *kokorakak*, Ili'uun *jalenahuun*, Welaun *dabadadain*, Owen Edwards pers. comm.) suggesting that some kind of sound symbolism is at play. In any case, the limited distribution of *boboraka* ~ *babaraka* within Makasae may suggest Kawaimina as the source of this term; however, it is not found in other Austronesian languages and would have to be a Kawaimina innovation.

Veloso (2016: 5) characterizes the Uatolari dialect of Naueti as having a larger number of Makasae borrowings than the Uatocarbau-Baguaia dialect.⁷ He presents two examples of borrowing from Makasae: Naueti Uatolari *rakalele* 'cheer, acclaim' and Naueti Uatolari *rubalele* 'vine (*Uvaria rufa*)'. In the case of *rakalele*, there are no cognates in the other Eastern Timor TAP languages and it does not appear to be segmentable in Makasae. By contrast, we find possible cognates for the likely constituent parts of this lexeme in Austronesian languages: for the final element *-lele* (cf. Waima'a *p'aa-lala* 'shout', Tetun *hak-*

7 Veloso (2016: 5) writes: "... I have noted an asymmetric incidence of Makasae loans in Uatolari Naueti compared to the amount present in the Uatocarbau-Baguaia dialect. Another reading of this phenomenon is that the Uatocarbau-Baguaia dialect shows more continuity with Naueti's sister languages ..."

lalak ‘make a loud noise, shout to make a loud noise, to shout, to cry out (many people showing enthusiasm, liveliness, etc.)’). For the initial element *raka-*, there are a number of Waima’a lexemes which appear to refer to attitudes or states of mind, e.g., *raka-bira* ‘lazy’, *raka-solo* ‘glad’, *raka-tiki* ‘tidy’. While not conclusive, this suggests that borrowing from a Kawaimina language to Makasae is a realistic possibility. For *rubalele*, the presence of *p* in the proposed Makasae source *rupalele* suggests that it is characteristic of the Ossu or Ossorua dialects, which are spoken in the Ossu subdistrict bordering the Uatolari subdistrict where Naueti is spoken; the expected form in other Makasae dialects would be *rufalele*.⁸ However, neither form is contained in our Makasae materials, and at the time of writing we have not been able to confirm it with a Makasae speaker. Neither could we find obviously cognate forms in other TAP languages. The etymon also does not seem to be present in other Kawaimina languages or Austronesian languages of the region.⁹ Naueti *rubalele* and Makasae *rupalele* are thus one of the class of shared items whose etymology cannot be established and where the direction of borrowing is unclear.

While the preceding discussion shows that we should not be too quick to assume Makasae influence on Naueti without good reason, we should also not assume that just because an item has a known etymology, particularly from influential Austronesian languages like Tetun or Malay, the directionality of borrowing is from Austronesian to Papuan. In fact, we find many borrowings in Kawaimina from these languages that have been mediated through Makasae. The tell-tale sign for Makasae being the immediate source for a loanword in a Kawaimina language is the presence of an additional final vowel not normally present in the Austronesian forms that is identical to the penultimate vowel of the root. Examples are given in Table 6.6. In Makasae paragon of a vowel echoing the final vowel of the root is a productive morphophonological process that affects all consonant-final roots including assimilated loanwords. The echo vowel is dropped when the root hosts a suffix or enclitic. By contrast, echo vowels are not known in Naueti or Waima’a phonology and Makasae roots borrowed with the echo vowels do not allow the final vowel to be dropped.

8 See Huber (2017: 272, 274) for the distribution of /f/ and /p/ in Makasae dialects.

9 Note, however, Tetun *karlele* ‘a variety of wild bean vine’, which may perhaps suggest that there is an element *-lele* associated with names for vines in at least one other Austronesian language of Timor.

TABLE 6.6 Examples of etyma borrowed through Makasae

Source	Makasae	Kaiwamina
Tetun <i>manaan</i> < Malay <i>mənaŋ</i> 'win'	Makasae Laga, Fatumaka, Ossu <i>manan-a</i> 'win, pass'	Naueti <i>manana</i> 'win' (cf. Waima'a <i>manaan</i> < Tetun)
Tetun <i>susar</i> 'be poor, experience difficulty'	Makasae Laga, Ossu <i>susar-a</i> 'difficult, complicated'	Naueti <i>susara</i> 'be difficult'
Tetun <i>dapur</i> ~ <i>dabur</i> ~ <i>dafur</i> < Malay <i>dapur</i> 'kitchen'	Makasae Laga, Baucau <i>dabur-u</i> 'kitchen' Makasae Fatumaka <i>dapur-u</i> 'kitchen'	Naueti <i>dapuru</i> 'kitchen' (cf. Waima'a <i>dabur</i> 'kitchen' < Tetun)
Tetun <i>botil</i> 'bottle' < Dutch <i>bottel</i> 'bottle'	Makasae Baucau, Fatumaka <i>botil-i</i> 'bottle'	Naueti <i>botili</i> 'bottle'
Tetun <i>dadur</i> 'imprison, hold captive'	Makasae Baucau <i>dadur-u</i> 'imprison, hold captive'	Waima'a <i>daduru</i> 'inmate, prison'
Tetun <i>lenuk</i> 'turtle'	Makasae Ossorua <i>neluk-u</i> 'turtle' (unexplained metathesis of l and n)	Naueti <i>neluku</i> 'turtle'
Tetun <i>toman</i> 'be accustomed to'	Makasae Laga <i>toman-a</i> 'to get used to'	Naueti <i>tomana leba</i> 'usually'
Voice form of PMP * <i>kawil</i> 'fishhook' (cf. Blust's PWMP * <i>ma-ŋawil</i> 'to fish with hook and line', ACD)	Makasae Laga <i>nail-i</i> 'fishing line' Makasae Fatumaka <i>nail-i</i> 'to fish, to hook'	Naueti <i>naili</i> 'fish hook' (cf. Waima'a <i>nai</i> 'fish hook')
SW Maluku language such as Kisar <i>dadila</i> 'gong'; widespread Wanderwort in Maluku, e.g., Bonfia <i>daldala</i> , Dobel <i>dadala</i> , Kei <i>dada</i> 'gong' cf. Tetun <i>dadir</i> 'bell'	Makasae <i>dadil-i</i> 'gong' (also Makalero <i>dadil-i</i>)	Waima'a <i>dadili</i> 'bell' Naueti <i>dadili</i> 'bell'

4 Mixed Origins and the Problem of Directionality: The Case of *-kai*

Section 3 has shown that the direction of borrowing between Maka and Kawai-mina languages is not always what it seems at first glance: the mere fact that a given etymon originates in Austronesian does not exclude the possibility of it having been re-borrowed into Kawai-mina through a TAP language. In this section, we highlight this issue further by outlining the complex history of a

specific etymon: *-kai*, a suffix found in Makasae and in the Kawaimina languages. At first glance, the suffix would seem to have a clear Austronesian source, going back to PMP *kahiw ‘wood, tree’ (ACD), reflected in Waima’a, Naueti and Midiki as *kai* ‘wood, tree’. In Waima’a and Naueti, *kai* also has a classificatory function: in Waima’a, it is the “default classifier, used for counting anything other than things for which one of the other classifiers is used” (Bowden 2006: 14), and in Naueti, it is grammaticalized as an agreement prefix on numerals used to count non-human referents (Veloso 2016: 44–45). A closer look suggests that we are dealing with two homophonous suffixes which appear to have been borrowed and re-borrowed multiple times.

A search for lexical items containing a suffix *-kai* in our Makasae, Waima’a and Naueti lexicon sources results in a list of items clustering in a small number of semantic domains: body part terms, animals, plants, and a handful of terms referring to humans and kin relations. Most noticeably, the suffix is found in all languages on a partially overlapping set of body part terms (Table 6.7).¹⁰

The Waima’a body part terms predominantly refer to protruding and/or elongated, bony parts of the body, suggesting that the original function of the suffix on body part terms was a shape-based classificatory one. From there, the suffix appears to have been borrowed into Makasae, perhaps as part of the body part term *turukai*, which is found in the Fatumaka and Ossorua dialects of Makasae. During the borrowing process, *turukai* underwent a semantic shift from ‘nose’ in Waima’a to ‘mouth’ in Makasae. This shift may be motivated by the use of Waima’a *turukai* ‘nose’ in the compound *manu-turukai* ‘beak’ (lit., bird-nose, Belo et al. 2005).¹¹ In Makasae, the suffix gained some degree of productivity, being used with TAP etyma (e.g., *muri-kai* ‘nose’ < PTAP *muri, *muta-kai* ‘back’ < PTAP *mota ‘behind, back’) as well as with body part terms whose Waima’a counterparts do not contain the suffix (e.g., *fanu-kai* ‘face’ < PTAP *panu ‘face’).

We find the largest number of body part terms with *-kai* in Veloso’s (2016) Naueti word list, where the suffix is consistently labelled as a Makasae borrowing. Given its Austronesian etymology and the parallel uses in Waima’a, it seems

10 Throughout this section we hyphenate where we analyse there to be an historical morpheme boundary. The sources are inconsistent as to whether *-kai* is treated as a morpheme. For example, in Belo et al. (2005) some of the Waima’a body-part terms are hyphenated (*n²eo-kai* ‘nape of the neck’, *malu-kai* ‘collar bone’, and *lase-kai* ‘penis’), whereas others are not (*turukai* ‘nose’ and *wuokai* ‘sternum’).

11 Other languages of Timor use a compound with ‘mouth’ to convey the same meaning, e.g., Tetun *manu-ibun* ‘beak’, literally, ‘bird-mouth’ (Belo et al. 2005; cf. Morris 1984).

TABLE 6.7 Body part terms marked with *-kai* in Makasae, Waima'a and Naueti

Makasae	Kaiwamina
Makasae Laga, Baucau <i>muni-kai</i> 'nose'	Waima'a <i>туру-kai</i> 'nose'
Makasae Ossu, Ossorua <i>muri-kai</i> 'nose'	Naueti <i>iru-kai</i> 'nose'
	Midiki <i>tu-kai</i> 'nose'
Makasae Ossorua, Fatumaka <i>туру-kai</i> 'mouth, lips'	Naueti <i>nunu</i> ~ <i>nunu-kai</i> 'mouth'
Makasae Laga, Fatumaka, Ossu <i>dela-kai</i> 'chin'	Naueti <i>timu</i> ~ <i>timu-kai</i> 'chin'
Makasae Ossorua <i>dela</i> ~ <i>dela-kai</i> 'chin'	
Makasae Laga, Fatumaka <i>mani-kai</i> 'neck'	Waima'a <i>n²eo-kai</i> 'nape of the neck' [†]
Makasae Baucau, Ossorua, Ossu <i>mane-kai</i> 'neck'	Naueti <i>[?]neo</i> ~ <i>[?]neo-kai</i> 'neck'
	Midiki <i>kai</i> 'neck'
Makasae Baucau <i>lia-kai</i> 'wing'	Naueti <i>lia-kai</i> 'wing'
Makasae Ossorua <i>lia</i> ~ <i>lia-kai</i> 'wing'	
Makasae Baucau <i>biti-kai</i> 'forehead'	
Makasae Ossu <i>budi-kai</i> 'forehead'	
Makasae Laga, Baucau <i>fanu</i> ~ <i>fanu-kai</i> 'face'	
Makasae Ossorua, Ossu <i>panu</i> ~ <i>panu-kai</i> 'face'	
Makasae Baucau, Ossu <i>muta</i> ~ <i>muta-kai</i> 'back'	
	Naueti <i>ṅala</i> ~ <i>ṅala-kai</i> 'crown of the head'
	Naueti <i>[?]lero</i> ~ <i>[?]lero-kai</i> 'throat'
	Waima'a <i>malu-kai</i> 'collar bone, clavicle' [†]
	Waima'a <i>wuo-kai</i> 'sternum'
	Naueti <i>gara</i> ~ <i>gara-kai</i> 'ear'
	Naueti <i>ikutara-kai</i> 'pelvis, hip'
	Waima'a <i>lase-kai</i> 'penis'
	Naueti <i>haʔa-kai</i> 'thigh'
	Naueti <i>gate-kai</i> 'calf (of the leg)'

highly unlikely that Naueti *-kai* is a Makasae borrowing in any straightforward sense. However, it is possible that contact with Makasae played a role in the extension of the range of the suffix to a comparatively large and diverse set of body part terms in Naueti (cf. ‘mouth’, ‘chin’, ‘wing’).

In Makasae and Naueti, *-kai* is also found in a small number of nouns referring to animals, as seen in (1). It is likely that the suffix has a similar classificatory function in these cases as with body parts: in Naueti *kida-kai* ‘dragonfly’ and *ʔmala-kai* ‘grasshopper’, it can be hypothesized to relate to the elongated form and stick-like appearance of the insects in question. The animals referred to in Makasae as *taimani-kai* ‘heron’ and *sabi-kai* or *bora-kai* ‘crab’, on the other hand, both have characteristic protruding body parts. Given the fact that *-kai* does not appear to be common in animal names in either language and that there is neither a direct semantic nor a formal overlap, these can be assumed to be independent, language-internal developments.

- (1) a. Makasae Laga, Baucau *tai-mani-kai* ‘heron’
 Makasae Baucau *sabi-kai* ‘crab’
 Makasae Baucau *bora-kai* ‘crab’
 b. Naueti *kida-kai* ‘dragonfly’
 Naueti *ʔmala-kai* ‘grasshopper’

The suffix *-kai* is also found in a small set of plant names in Makasae, Waima’a and Naueti (2). The presence of *kai* ‘wood, tree’ in Kawaimina plant names is hardly surprising, and several have been borrowed into the Maka languages. Usually, however, *kai* is the first, rather than the last, element in Kawaimina plant names; Belo (2005) and Veloso (2016) include numerous examples, some of which are given in (3a) and (3b). Likewise, in native Makasae plant names, the generic noun *ate* ‘tree, plant, wood’ is the first element. At first glance, the position of *-kai* in the plant names in (2), at the end of the name, is thus unusual. Most likely it is in these cases not the generic plant noun, but rather the classifier *-kai* that we have seen in body parts as well as animals, referring to elongated, hard protruding parts characteristic of the plants in question.

- (2) a. Makasae Baucau *uru-kai* ‘pepper, chili’
 b. Waima’a *iludai-kai* ‘cassava’
 c. Naueti *kone-kai* ‘turmeric’
 Naueti *ua-kai* ‘rattan’
 Naueti *dare-kai* ‘corn cob flower’

- (3) a. Waima'a *kai-bubu* 'eucalypt'
 Waima'a *kai-dile* 'papaya tree (*Carica papaya*)'
 Waima'a *kai-dawa* 'Malay lac tree (*Schleichera oleosa*)'
 b. Naueti *kai-haku* 'quinine tree (*Cinchona* sp.)'
 Naueti *kai-dila* 'papaya tree (*Carica papaya*)'
 Naueti *kai-dawa* 'Malay lac tree (*Schleichera oleosa*)'
 c. Makasae Baucau, Laga *ate-muni* 'sandalwood (*Santalum* sp.)'
 Makasae Baucau *ate-ra?u* 'blackboard tree (*Alstonia scholaris*)'
 Makasae Baucau *ate-kaisuti* 'tree species (*Cassia timoriensis*)'

In Makasae, there is a second suffix *-kai*, a widely used diminutive suffix on personal names, e.g., *Edukai* < *Eduardo*, *Anakai* < *Ana*, *Makai* < *Maria* (Correia 2011: 55; Huber 2008: 13). According to Correia (2011: 97, fn. 121), this *-kai* derives from a common form of address *kakai*, a combination of the kinship term *kaka* 'older brother or sister' (itself an Austronesian borrowing, > PMP **kaka* 'elder sibling of the same sex') and the diminutive suffix *-i*, which is often reduced to *kai* in everyday language use. There is a further diminutive suffix, *-lai*, which is used not only with shortened personal names, but also with nicknames derived from common nouns and verbs (Correia 2011: 54–55). The existence of the formally similar suffix *-lai* may have helped along the grammaticalization of *-kai* as a diminutive suffix.¹² The Makasae diminutive *-kai* is also used on personal names in Naueti, e.g., *Libakai*, *Makai* (Menezes and Rosario Pires 2006; Veloso 2016), although it is unclear to what degree it is productive in that language. Thus, while the classifying suffix *-kai* has been borrowed into Makasae from Kawaimina, there is some evidence of Kawaimina languages in turn borrowing the Makasae diminutive *-kai*.

Finally, we also find *-kai* in a small set of common nouns referring to human beings (Table 6.8). The noun *asukai* 'man, husband' is not only shared across Makasae, Waima'a and Naueti, but is also found in Kairui and Midiki. According to Veloso (2016: 123) it is a Makasae loan in Naueti. However, within the TAP languages of Eastern Timor *asukai* is not found beyond Makasae; the other Eastern Timor languages use *nami* 'male, husband', suggesting *asukai* may instead have an Austronesian source. As noted in Hull (2000: 174), there is a striking similarity between *asukai* and Tetun *asuwain* 'hero', the initial *asu-* element of which most likely goes back to PMP **qasawa* 'spouse'.¹³ The *-kai* suffix in the

12 Note, for instance, the alternation in *sabi-kai* ~ *sabi-lai* 'crab'. It has been suggested above that *-kai* may have a classificatory function in *sabikai*. However, it may also be analyzed as a diminutive.

13 However, Hull's subsequent suggestion that the reconstructed root underwent metathesis to *asu* due to a semantic association with Tetun *asu* 'dog' appears more dubious.

TABLE 6.8 Human nouns with *-kai*

Makasae	Kawaimina
<i>asukai</i> ‘man, husband’	Waima’a <i>asukai</i> ‘man, husband’ Naueti <i>asukai</i> ‘man, husband’ Kairui <i>asukai</i> ‘man’ Midiki <i>asukai</i> ‘man’
	Waima’a <i>ine-kai</i> ‘mother’
	Waima’a <i>umo-kai</i> ‘nuclear family’
	Waima’a <i>naʔu-kai</i> ‘thief’
	Naueti <i>molu-kai</i> ‘stupid person’

Kawaimina languages and Makasae may relate to a botanic idiom used to describe family and alliance relations which is common in the Austronesian languages of the region (Fox 1980; Fox and Sather 2006), in reference to men’s cultural role in the founding and continuation of lineages. This botanic idiom likely also accounts for Waima’a *ine-kai* ‘mother’ and *umo-kai* ‘nuclear family’; in fact, the same suffix *-kai* is also found with kinship terms in languages outside of the Makasae-Kawaimina contact situation (cf. Welaun, Raklungu *ina-kai* ‘mother’, *ama-kai* ‘father’, Edwards 2019). Waima’a *naʔu-kai* ‘thief’ and Naueti *molu-kai* ‘stupid person’ remain unexplained for now.

In sum, *-kai* is ideally suited to illustrate the complexity of the linguistic diffusion which has taken place in the Kawaimina-Makasae contact situation: as we have shown in this section, there are actually two formally identical morphemes, which are found in all of the languages involved. The classificatory suffix found in body part terms is derived from Kawaimina *kai* ‘tree, wood’. It has been borrowed into Makasae, which in turn may have influenced its use in Naueti. The diminutive *-kai*, on the other hand, has spread to the Kawaimina languages from Makasae, although here it is not entirely clear how productive this is. Given that it is used most prominently in personal names, vocabulary lists do not give too much evidence. Interestingly, this suffix, too, originally goes back to an Austronesian root, PMP *kaka ‘elder sibling of the same sex’, which is truncated and combined with another Makasae diminutive, *-i*. In all cases, multiple borrowing events were likely involved in creating the distribution we see today.

5 Conclusion

This paper has presented a first study of the shared lexical histories of the Austronesian languages of the Kawaimina group and the Papuan languages of the Maka group spoken in East Timor. Rather than focus on the appearance of Austronesian lexemes in Papuan languages, we concentrated on the reverse scenario, one which is rarely addressed in the existing literature. In doing so, we have highlighted the complexity of the borrowing situation between Austronesian and Papuan languages in East Timor.

We have shown that today Kawaimina languages have around a dozen lexemes with Papuan sources going back to PTAP or PET. While the number of items identified is small at this stage, we still regard it as significant, since for the most part claims of innovative vocabulary borrowed from Papuan languages in the literature are not given a source in a particular Papuan language or family. For many of the borrowings identified as Papuan here, Maka languages cannot be the source of the borrowing and contact with other, related Papuan languages that are no longer extant is posited to have occurred. We regard the lexical documentation of the Papuan languages as a major limiting factor to the identification of further borrowings from the Papuan languages. While the Maka languages have for the most part well-described grammars, lexical materials for them are lacking in crucial domains such as plants and animals, in particular creepy-crawlies, where borrowings from Papuan languages seem to cluster.

The case of the Kawaimina and Maka languages also illustrates that it is important not to exclude a lexeme as a possible loan candidate just because it has a known Austronesian etymology. We saw that there were multiple instances where Austronesian lexemes were borrowed into the Maka languages or, more commonly, Makasae, and then from there borrowed again into one or more Kawaimina language. Assumptions about directionality of borrowing of such Austronesian items are likely to underpin some of the statements in the literature to the effect that Papuan borrowing is minimal in Kawaimina languages.

We have drawn attention to many lexemes which are shared exclusively between the Kawaimina and Maka languages. In these cases, the shared nature of the forms indicates that borrowing has taken place but the direction of the borrowing is impossible to determine. For many of these items, we noted that Maka languages have phonemes that were innovative in PMAKA and not continuations of phonemes from PTAP. This was taken to indicate that they were introductions into the phoneme inventory occasioned by borrowing. We speculated that parallel borrowing from a third source may account for some of

the shared lexemes in Kawaimina and Maka languages, but in most cases that scenario was more complex than positing borrowing into Maka and then to Kawaimina.

Finally, the case study of the *-kai* suffixes found in Makasae as well as the Kawaimina languages serves to highlight the degree to which the histories of the two language groups are entwined: we find evidence of borrowing at different historical stages, convergent as well as independent language-internal developments, and re-borrowing. Several Naueti forms containing *-kai* have been assumed to be borrowings from Makasae, making it one of the relatively few instances discussed in the literature of TAP lexical transfer into Timor's Austronesian languages. A classifier-like *-kai* was borrowed from a Kawaimina language into Makasae. There it gained some degree of productivity and may have reinforced its use in Naueti. By contrast, the diminutive *-kai* developed in Makasae, from where it spread to Naueti. The case study not only traces the complex history of the suffixes across language family boundaries, but also lends some support to previous claims of a stronger Makasae influence on Naueti as opposed to other Kawaimina languages.

The lexical entwinement of the Kawaimina and Maka languages set out in this chapter makes clear that historical linguists will need to reckon with bi-directionality in lexical borrowing between Papuan-Austronesian languages in the Timor area. It also shows that not only detailed documentary materials, but also a nuanced understanding of the diachrony of all languages involved are a prerequisite to accurately assess lexical transfer in language contact—something that is still lacking for many Papuan-Austronesian contact situations in the area. Further case studies carefully unpacking the lexical histories of these languages in contact are needed to shed light on the prehistorical dynamics between Papuan- and Austronesian-language speaking groups.

Acknowledgements

Many thanks to Owen Edwards for the detailed comments and extremely informative feedback he gave on an earlier version of this paper. We would also like to thank Alex Veloso, whose documentation and sketch of Naueti made this paper possible. Jessé Fogaça was also extremely helpful in checking various Makasae forms for us, while Thomas Mariano Pinto and Iliwatu Danabere provided information on Makalero forms. Schapper would also like to thank Erik Zobel for discussion of various topics. All errors are our own. Schapper's research was funded by the Netherlands Organisation for Scientific Research VENI project "The evolution of the lexicon. Explorations in lexical sta-

bility, semantic shift and borrowing in a Papuan language family”, the Volkswagen Stiftung DoBeS project “Aru languages documentation”, and the European Research Council “OUTOFPAPIA” project (Grant Agreement No. 848532).

Bibliography

- Arnaud, Véronique and Henri Campagnolo. 1997. *Lexique thématique plurilingue de trente-six langues et dialectes d'Asie du Sud-Est insulaire*, 2 vols. Paris: L'Harmattan.
- Belo, Maurício da Conceição Aparício, John Bowden, John Hajek, Nikolaus P. Himmelmann and Alexandre Vital Tilman. 2005. Glosáriu Waima'a Caisido. The Language Archive. https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0008_345A_5
- Blust, Robert and Stephen Trussel. 2020. Austronesian comparative dictionary. URL: <https://www.trussel2.com/acd/>
- Bowden, John. 2006. Notes on Waima'a grammar. The Language Archive. https://archive.mpi.nl/tla/islandora/object/tla:1839_00_0000_0000_0008_3448_A?asOfDateTime=2018-03-02T11:00:00.000Z
- Brotherson, Anna. 2003. 2003: *A spatial odyssey: Referring to space in Makasai*. Canberra: Australian National University Honours thesis.
- Correia, Adérito José Guterres. 2011. *Describing Makasae: A Trans-New Guinea Language of East Timor*. Sydney: University of Western Sydney PhD thesis.
- Dawson, Virginia. 2014. Midiki wordlist. PARADISEC. http://catalog.paradisec.org.au/repository/VD1/krd20140214_wordlist
- Edwards, Owen. 2018. Top-down historical phonology of Rote-Meto. *Journal of the Southeast Asian Linguistics Society* 11(1). 63–90.
- Edwards, Owen. 2019. Reintroducing Welaun. *Oceanic Linguistics* 58(1). 31–58.
- Edwards, Owen. 2021. *Rote-Meto comparative dictionary*. Canberra: ANU Press.
- Fogaça, Jessé Silveira. 2015. *Aspectos gramaticais da língua Makasae de Timor-Leste: Fonologia, morfologia e sintaxe*. Brasília: Universidade de Brasília PhD thesis.
- Fox, James J. 1980. Models and metaphors: Comparative research in Eastern Indonesia. In James J. Fox (ed.), *The flow of life: Essays on Eastern Indonesia*, 327–333. Cambridge, MA: Harvard University Press.
- Fox, James J. 2006. Introduction. In James J. Fox and Clifford Sather (eds.), 2006. *Origins, Ancestry and Alliance: Explorations in Austronesian Ethnography*. Canberra: ANU Press.
- General Directorate of Statistics. 2015. 2015 Census publications. <http://www.statistics.gov.tl/category/publications/census-publications/2015-census-publications/>
- Hajek, John and Nikolaus P. Himmelmann. 2006. Waima'a Caisido setting sketch. The

- Language Archive. https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_000_0_0000_0008_3440_9
- Huber, Juliette. 2008. *First steps towards a grammar of Makasae*. München: Lincom.
- Huber, Juliette. 2011. *A grammar of Makalero: A Papuan language of East Timor*. Utrecht: LOT.
- Huber, Juliette. 2017. Makalero and Makasae. In: Schapper, Antoinette (ed.), *The Papuan languages of Timor, Alor and Pantar: Sketch Grammars*, vol. 2, 267–352. Berlin: Mouton.
- Hull, Geoffrey. 1998. The Basic Lexical Affinities of Timor's Austronesian Languages: A Preliminary Investigation. *Studies in Languages and Cultures of East Timor* 1. 97–202.
- Hull, Geoffrey. 2000. Historical Phonology of Tetum. *Studies in Languages and Cultures of East Timor* 3. 158–212.
- Hull, Geoffrey. 2004. The Papuan Languages of East Timor. *Studies in Languages and Cultures of East Timor* 6. 23–99.
- Lewis, M. Paul, Gary F. Simons and Charles D. Fennig. 2015. *Ethnologue: Languages of the World*, 18th edition. Dallas, Texas: SIL International. <http://www.ethnologue.com>
- McWilliam, Andrew. 2007. Austronesians in linguistic disguise: Fataluku cultural fusion in East Timor. *Journal of Southeast Asian Studies* 38(2). 355–375.
- Menezes, António and Isabel do Rosario Pires. 2006. *Tetun-Nauweti-English*. Dili: Timor Loro Sa'e-Nippon Culture Center.
- Morris, Cliff. 1984. *Tetun-English dictionary*. Canberra: Pacific Linguistics.
- Nácher, Alfonso. 2012. *Lexico Fataluco-Portugues (Introdução e edição de Geoffrey Hull)* [Fataluku-Portuguese dictionary (Introduction and edition by Geoffrey Hull)]. Dili: Gráfica Patria.
- Pinto, Tomas Mariano. 2004. Makalero Dicionário. [Makalero Dictionary.] Dili: Timor Loro Sa'e-Nippon Culture Center.
- Ribeiro, Alvaro. 2005. Makasae Fatumaka. Language Documentation and Training Centre. University of Hawai'i at Mānoa. Partially retrievable under <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/956b04c5-3677-4a27-a221-af7aade124ca/content> (15 November 2022).
- Sarmiento, João da Silva. 2005. Makasae Osoroa. Language Documentation Training Center, University of Hawai'i at Mānoa. https://web.archive.org/web/2011206224450/http://www.ling.hawaii.edu/~uhdoc/makasae_osor/background.html (9 November, 2021)
- Saunders, George. 2003. Comparative vocabulary of the Naueti dialect. *Studies in Languages and Cultures of East Timor* 5. 79–106.
- Schapper, Antoinette. 2019. The ethno-linguistic relationship between smelling and kissing: A Southeast Asian case study. *Oceanic Linguistics* 58(1). 92–109.

- Schapper, Antoinette. 2020a. Introduction. In Antoinette Schapper (ed.), *The Papuan Languages of Timor, Alor and Pantar: Sketch Grammars*, vol. 3, 1–52. Berlin: Mouton.
- Schapper, Antoinette. 2020b. The origins of isolating word structure in eastern Timor. In David Gil and Antoinette Schapper (eds.), *Austronesian undressed: How and why languages become isolating*, 391–446. Amsterdam: John Benjamins.
- Schapper, Antoinette. Forthcoming. Papuan influence on the Malayo-Polynesian languages of Southeast Asia. In Alexander Adelaar and Antoinette Schapper (eds.), *The Oxford Guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: Oxford University Press.
- Schapper, Antoinette. In prep. The history and diversification of the Timor-Alor-Pantar languages. Manuscript, Perth.
- Schapper, Antoinette and Juliette Huber. 2019. The Austronesian-Papuan contact history of eastern Timor: What lexical borrowing can tell us. Presented at the 11th International Austronesian and Papuan Languages and Linguistics Conference (APLL11), 13–15 June, Leiden. DOI: 10.5281/zenodo.5641879
- Schapper, Antoinette, Juliette Huber and Aone van Engelenhoven. 2012. The historical relation of the Papuan languages of Timor and Kisar. In Harald Hammarström and Wilco van der Heuvel (eds.), *Contact and classification of Papuan languages* (Language and Linguistics in Melanesia Special Issue 2012), vol. 1, 194–242. Port Moresby: Linguistic Society of New Guinea.
- Schapper, Antoinette, Juliette Huber and Aone van Engelenhoven. 2014. The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology*, 91–147. Berlin: Language Science Press.
- Schapper, Antoinette and Erik Zobel. Forthcoming. Languages of Timor and Southern Maluku. In Alexander Adelaar and Antoinette Schapper (eds.), *Oxford Guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: Oxford University Press.
- Veloso, Alexandre. 2016. *A grammar sketch of Naueti, a language of East Timor*. Leiden: Leiden University MA thesis.
- Usher, Timothy and Antoinette Schapper. 2022. The Greater West Bomberai Family. *Oceanic Linguistics* 61.1: 469–527.

PART 2

Modern and Contemporary Contact



Detecting Papuan Loanwords in Alorese: Combining Quantitative and Qualitative Methods

Francesca R. Moro, Yunus Sulistyono and Gereon A. Kaiping

Introduction

In many parts of eastern Indonesia and Melanesia, speech communities often lack archaeological data and historical written sources, meaning that linguistic data is the only means by which to reconstruct past social interactions of speech communities (Ross 2013; Klamer 2015). Alorese, a language spoken in eastern Indonesia in a small-scale bi-/multilingual setting, is one such community. To reconstruct the sociolinguistic past of the Alorese, this paper analyses quantitative and qualitative patterns of lexical borrowing between Alorese and its neighboring languages.

Alorese is the only Austronesian language spoken on the coasts of the Alor and Pantar archipelago. On current accounts, it consists of 13 dialects or varieties corresponding to the main coastal villages where Alorese is spoken (see Figure 7.3 in § 2). The other languages spoken on those islands are the Alor-Pantar languages¹ (henceforth AP), which belong to the (Papuan) Timor-Alor-Pantar family (henceforth TAP see § 1).² As a point of contact between Austronesian surrounded by non-Austronesian languages, Alorese constitutes a special ‘natural laboratory’ for language contact studies. Since their arrival on the archipelago about 600 years ago, Alorese varieties have been in contact with the local AP languages. This long-term contact has affected the Alorese grammar, resulting in morphological simplification and a few structural borrowings (Klamer 2011; Moro 2018, 2019; Moro & Fricke 2020).

Interestingly, the two earlier publications on the topic (Klamer 2011; Robinson 2015) seem to indicate that Alorese lexicon is less affected by the long-term contact than the grammar. Both of these studies focus on a small part

¹ Note that despite the name, Alorese itself is not a Timor-Alor-Pantar language.

² The following abbreviations are used: AP = Alor-Pantar, PAL = Proto Alorese, PAP = Proto Alor Pantar, PFL = Proto Flores-Lembata, PMP = Proto Malayo-Polynesian, PTAP = Proto Timor-Alor-Pantar, PWL = Proto Western Lamholot, TAP = Timor-Alor-Pantar.

of the basic vocabulary (a Swadesh list), which might be more resistant to borrowing than the lexicon overall. A section in the short grammar of Alorese by Klamer (2011: 104–107) indicates an estimated percentage of 5.2%, while an article by Robinson (2015) discussing Austronesian borrowings into AP languages and AP borrowings into Austronesian languages finds about 3.8% AP loans in Alorese. These numbers are surprisingly small, considering the length of contact.

In this paper, we research whether the observation applies beyond the core vocabulary by extending the data to a 596-concept list, including all 13 Alorese dialects. Unlike other studies investigating Austronesian-Papuan borrowings (see among others Klamer's chapter in this volume), we did not pre-select the semantic domains to study, but investigated the entire dataset, and got the semantic domains of the loanwords inductively. In order to detect borrowing events, an algorithm was used to sifts loanwords out of a huge lexical pool: ~600 words × 13 Alorese dialects, × 55 Austronesian languages, × 42 TAP language varieties = approximately 66,000 word forms (see § 2). This pool is much larger than the dataset used in the previous research on AP borrowing in Alorese.

The present chapter, thus, illustrates an innovative methodological approach to the study of loanwords which uses an algorithm for automatic lexical similarity detection to study loans across two linguistic families. In this chapter, we describe the two-step procedure that was employed and how the results compare to work that has done this manually, to answer questions such as: does the size of a dataset make a difference when we investigate relative amount of borrowing? And does the percentage of borrowings increase when we investigate a large dataset, including highly borrowable concepts, compared to when we investigate a Swadesh list? Another innovative aspect of the chapter is that this is the only study in which 13 dialects of a minority language of Indonesia are compared. Comparing dialects on the patterns of lexical borrowing allows us to answer questions such as, do dialects of a language show differences in terms of their patterns of borrowing? Can this difference be related to their geographical location, their neighbours, or to the individual histories of the dialect communities?

A preliminary version of this research has been published in Chapter 6 of the PhD dissertation of Sulistyono (2022), in which lexical borrowings from and into Alorese and various languages including AP, Malay, Dutch and Portuguese are discussed. The present chapter has reconsidered the loan status of some words, excluding one concept, 'finished', and including four concepts 'dolphin', 'gravel', 'to breathe', and 'to hide'. Additionally, we provide an explanation to account for the limited lexical influence, and place our findings in

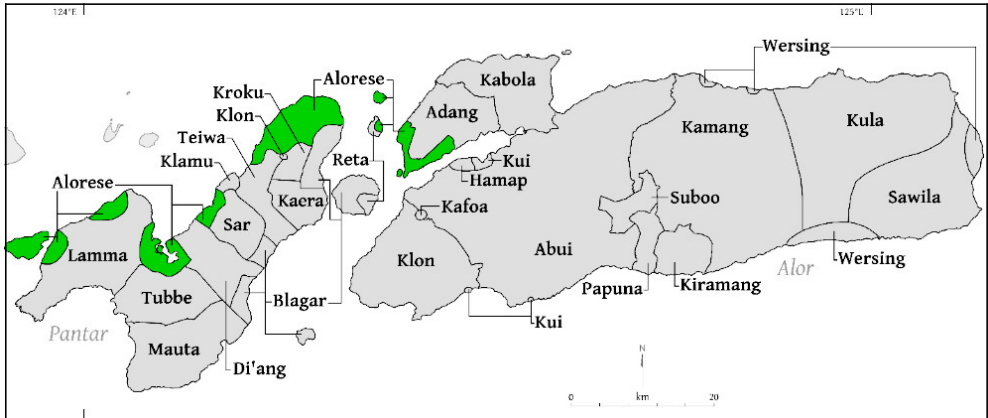


FIGURE 7.1 Alorese spoken on Alor and Pantar

a broader geographical perspective, relating our results to those of other studies in the present volume.

This chapter is organised as follows. As a background to this study, we begin by providing some basic information on Alorese and AP languages in §1; this is followed by §2 illustrating the research questions, the dataset, and the methodology of the present study. §3 presents the main findings, while §4 discusses the findings and gives some concluding remarks.

1 Alorese and the AP Languages

Alorese has approximately 25,000 speakers (Eberhard, Simons, & Fennig 2019). It is spoken along the coasts of Alor and Pantar, and on two small islands in the Alor-Pantar Strait in the Indonesian province of Nusa Tenggara Timur (see green areas on Figure 7.1 above). Besides Indonesian and the local Malay variety, Alorese is the only Austronesian language and is indigenous in the area.

The other languages spoken on those islands are roughly 25 Papuan languages of the Alor-Pantar (AP) subgroup, which belongs to the Timor-Alor-Pantar (TAP) family (Schapper, Huber, & van Engelenhoven 2017). There is evidence that the AP languages are spoken on the Alor archipelago since ~3,000 BP (Klamer 2017: 10), thus long before the arrival of the Alorese.

On Alor, Alorese is only spoken on the northern peninsula, alongside Adang; on Pantar, it is spoken alongside Krok, Teiwa, and Nedebang (Klamu), among others. The historical situation of Alorese as Austronesian language spoken amid a mosaic of AP languages continues to the present day.

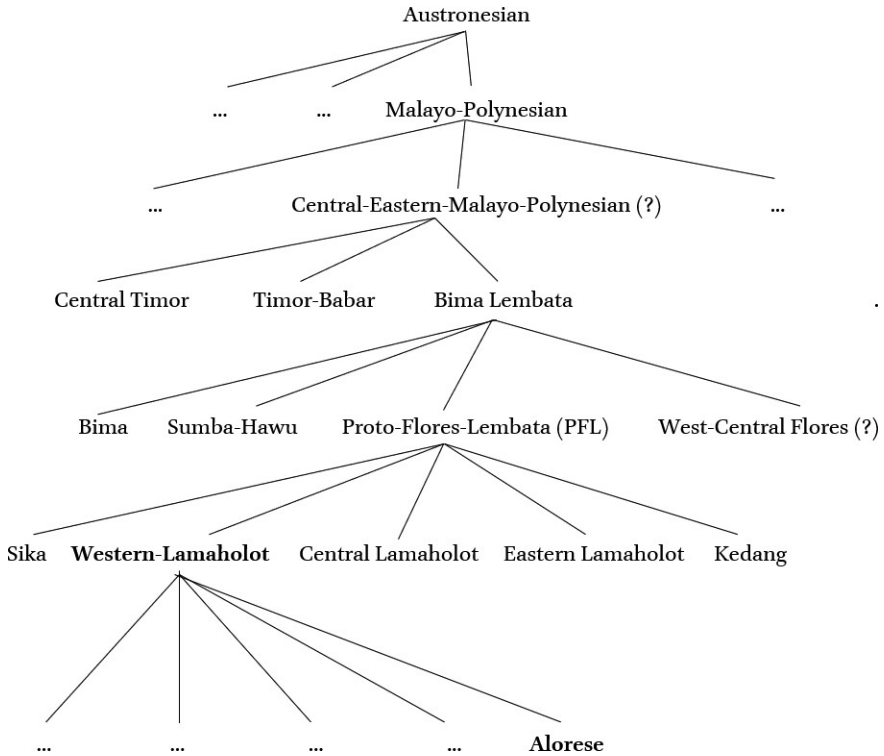


FIGURE 7.2 Genealogical classification of Alorese

Historically, Alorese speakers are descendants of groups migrating eastwards from the neighbouring island of Flores and its offshore islands (Klamer 2011: 8–15; Wellfelt 2016: 248–249; Sulistyono 2022). Historical linguistics indicate that the language spoken by these migrating groups was a western Lamaholot variety that later developed into what we today call ‘Alorese’. Therefore, from a genealogical perspective, the closest relatives of Alorese are western Lamaholot varieties (Doyle 2010: 30; Elias 2017; Fricke 2019; Sulistyono 2022). Alorese and western Lamaholot varieties belong to the Flores-Lembata subgroup of Malayo-Polynesian languages, which also includes the eastern and central Lamaholot varieties, Sika, and Kedang (Fernandez 1996; Fricke 2019). Figure 7.2 above shows the genealogical classification of Alorese (Sulistyono, 2022:144; Fricke, 2019:20).

According to Anonymous (1914: 77), the first Alorese settlers arrived “5 to 600 years ago”, meaning that they arrived around 1300–1400. Local oral history suggests that the northeastern Pantar area, in particular today’s villages of Pandai (see Figure 7.3 on the next page), was the first area settled by the Alorese in the 14th century. It was followed by the expansion to the Alor pen-

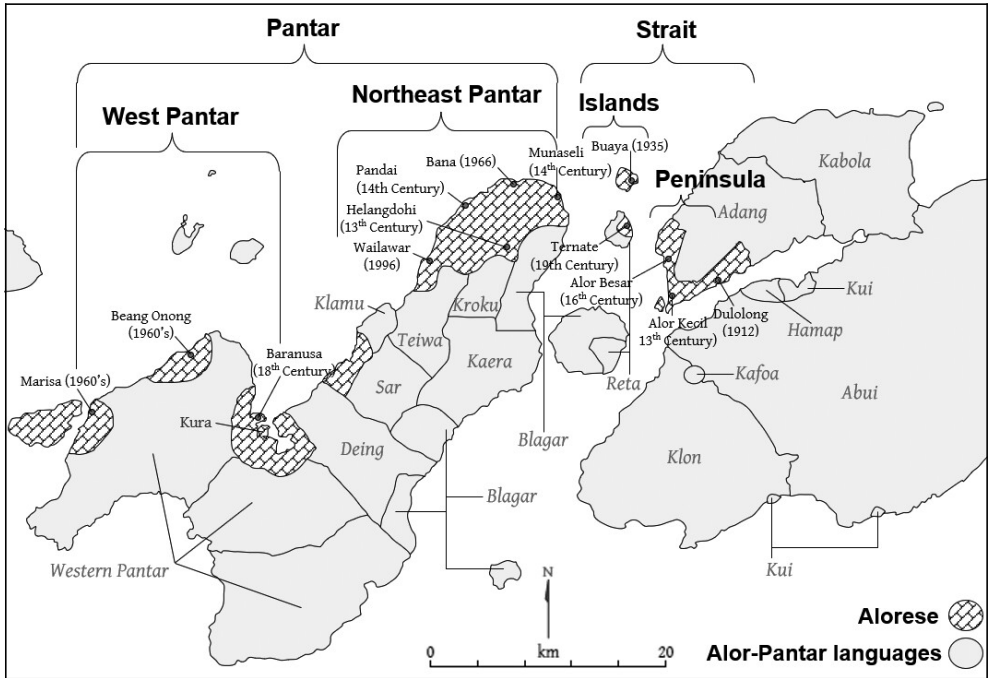


FIGURE 7.3 The 13 Alorese varieties spoken on Alor and Pantar, labeled according to village name

insula in the 16th century, and later expansions to the west and to the Strait starting in the 18th century (see Sulistyono 2022 for a detailed account on oral histories in this region). We will see in § 3.1 that this scenario is supported by the patterns of AP lexical borrowings in Alorese investigated in this paper. As a result, from a geographical and historical perspective, there are 13 Alorese varieties grouped in four main clusters: Northeast Pantar, Alor Peninsula, Strait, and West Pantar (from oldest to most recent). This geographical grouping is useful when determining the spread of loanwords in the varieties.

Traditionally, the Alorese practise exogamy with the neighbouring AP communities. In the past, exchanging women was a necessity for the Alorese, because their settlements only numbered about 200–300 people (Anonymous 1914: 89–90). Today, exogamy is still practised; however, the percentage of Papuan women has dropped considerably, as the Alorese settlements have become larger (approximately 1,500–2,000 inhabitants) and it has become easier to find a spouse within the same settlement. The settlement patterns are patri-virilocal, and the women generally move to the husband's village and are expected to learn Alorese (cf. also DuBois 1944: 85). Exogamy and patri-virilocal culture are inevitably linked to specific language acquisition patterns. In the past, Alorese villages must have been home to a continuous and considerable influx

of Papuan women who learned Alorese as a second language (L2), as well as bilingual children growing up learning both Alorese and an AP language (from their mother). These acquisition patterns are currently changing, as the local Malay variety and Indonesian are both gaining more ground.

Turning to the issue of language equality, at some point in the history of the Alorese, their language started to enjoy slightly more prestige over the AP languages, due to its role as lingua franca in the area of the Alor-Pantar Strait before Indonesian was introduced in the 1960s (Stokhof 1975: 8; DuBois 1944: 16). The status of Alorese as lingua franca arose due to the involvement of the Alorese in a Chinese-Muslim trade network bringing goods and slaves to Alor. Furthermore, during colonial times, the Alorese rulers acted as intermediaries between the inland Papuan population and the colonial governments (Stokhof 1984: 111). This situation must have led to the increase of asymmetric multilingualism, with Papuan speakers learning Alorese, but Alorese speakers remaining mostly monolingual.

2 The Present Study

This paper investigates Alor-Pantar (AP) loanwords in Alorese looking at a large lexical dataset. Using a two-step combination of automatic pre-screening and qualitative checks, we classify as candidate loanwords in Alorese all forms that are not inherited from the ancestor language (Proto Flores-Lembata), but that are formally similar to their semantic equivalents in one or more AP language, and check them individually.

2.1 *Dataset and Methodology*

In order to understand the patterns of loanwords in Alorese, we worked with word list data collected from field work and published sources aggregated in the online lexical database LexiRumah (Kaiping, Edwards, & Klamer 2019). We use version 1.0.0 of the database.³ The sources of the individual word lists and forms used here can be found on LexiRumah. The dataset contains between 104 and 756 forms (counting all synonyms, and counting polysemous words once for every meaning) associated to a list of 596 concepts. The concept list contains pronouns and numerals, and nouns and verbs relating to both basic human activities (e. g., 'knife', 'to pull', 'to work', 'fireplace ash'), as well as the

3 A more recent version 3.0.1 includes an expanded set of languages, which are mostly more distant Austronesian or other Papuan languages, and thus not relevant for the Alorese lexicon.

natural and cultural world of the region (e.g., ‘sun’, ‘island’, ‘mountain’, ‘dolphin’, ‘rice ear bug’, ‘chicken’, ‘to plant yam’, ‘to clear land by burning’).

The language dataset contains 13 Alorese varieties or dialects (one given in two sources), each of which displays between 450 and 756 forms for those concepts, 55 other Austronesian languages or varieties, and 42 TAP languages or varieties. For the Austronesian and TAP languages, the dataset contains between 104 and 756 forms (counting all synonyms, and counting polysemous words once for every meaning) associated to the list of concepts.

The first step we took was a data-mining process to discover potential loanword patterns in such a large dataset. So, we investigated lexical data from a quantitative perspective, by applying automatic lexical similarity detection. In the second step, we conducted a qualitative fine-grained analysis on the similarity sets whose patterns of distribution were compatible with borrowing event between Alorese and AP languages.

Borrowing between Alorese and AP languages would be visible in forms that are similar between Alorese and AP languages, and not explained otherwise: If forms in AP and Flores-Lembata languages are also similar, other explanations are assumed (e. g., borrowing before the genesis of Alorese, or widespread borrowing from Indonesian). Borrowed forms may have different meanings from the form in the donor language due to an originally general term being applied to a more specific foreign concept or due to subsequent semantic shift (Winter-Froemel 2013). In wordlist data, semantically different borrowed forms are hard to detect (List & Forkel 2021), and thus beyond the scope of our study. We thus focus on etymologically related forms within each concept.

In order to find candidates of etymologically related forms shared between AP languages and Alorese, we applied the automatic lexical similarity detection tool LexStat (List 2012), implemented in LingPy 2.6.5 (List et al. 2019). The LexStat algorithm uses a simplified ‘sound class’ representation (List 2012) of the forms in each language. Forms are matched with each other, and their sound class sequences are aligned with each other, giving a score that describes how many sounds in a form need to be changed to generate the corresponding form in a different language. Using stochastic methods, LexStat extracts the information whether the correspondence between different sounds is systematic or sporadic. LexStat’s cognacy score then describes how many effective changes, discounting systematic differences, are needed to transform one form into another—lower scores mean that two forms are likely cognate, higher scores point to a lack of etymological relation. All pairs forms that have a cognacy score more similar than a set threshold of 0.55 are then connected into a network. The resulting network of forms is then split into discrete cognate classes using a graph partition algorithm, such as Infomap (Rosvall, Axelsson

TABLE 7.1 Relevant patterns of distribution of lexically similar forms in languages of the region, and the corresponding borrowing or inheritance history of such a form

Hypothesis	AP languages	Alorese	Flores-Lembata	Indonesian	Other Austronesian	Explanation
1	Present	Absent	Absent	Absent	Absent	Inherited TAP vocabulary
2	Present	Present	Absent	Absent	Absent	Loan from AP into Alorese or vice versa, to be further inspected
3	Present	Present	(likely present)	Present	(likely present)	Indonesian loan into local languages
4	Present	Present	Present	Absent	Absent	Likely Alorese loan (inherited from PFL) into AP

& Bergstrom 2009). The algorithm creates classes which are strongly connected internally but have only weak connections between different classes.

While LexStat has been designed, and tested (Rama et al., 2018) for identifying cognate forms under systematic sound correspondence, the underlying similarity scoring is also promising for loan detection. While borrowing from one language into another does not follow systematic diachronic sound laws, phonological adaptation from the donor language to the recipient language may nonetheless introduce systematic changes (Uffmann 2015) and the general surface similarity should be picked up by LexStat's sequence alignment algorithm.

In order for an item to be an indication of borrowing between Alorese and an AP language, the lexically similar forms must be present in at least one Alorese dialect and at least one AP language. Different patterns of distribution of such forms outside Alorese and the AP languages indicate different hypotheses about the history of the word. The most important such hypotheses are summarized in Table 7.1. In this table, "present" means that the lexical similarity set contains a form for at least one language/dialect of that group, "absent" means that no language in that group has an attested form in that lexical similarity set.

As illustrated by hypothesis 2 in the previous table, an AP loan candidate in Alorese must be present in at least one Alorese variety, in at least one AP language, but in no other Austronesian language. To illustrate the automatic loan detection, an example is presented in Table 7.2, which shows the lexical similarity set for the concept 'to breathe'.

TABLE 7.2 Examples of a lexical similarity set associated with the concept 'to breathe' generated using automatic comparison

Concept	Language	Alignment	Form
to_breathe	Alorese-Munaseli	h ɔ - p a ŋ	hopang
to_breathe	Blagar-Bama	s o - p a ŋ	sopang
to_breathe	Blagar-Kulijai	h o - p a ŋ	hopang
to_breathe	Blagar-Nule	h o - p a ŋ	hopang
to_breathe	Blagar-Pura	h o - p a ŋ	hopang
to_breathe	Deing	- o - p a ŋ	opang
to_breathe	Kaera	s u ? p a ŋ	su'pang
to_breathe	Western Pantar-Tubbe	h o - p a ŋ	hopang
to_breathe	Reta-Pura	h o : - p a ŋ	hoopang
to_breathe	Reta-Ternate	h u - p a ŋ	hupang

The automatic comparison recognized that one Alorese variety, i.e., Alorese-Munaseli has the word *hopaŋ* 'to breathe' which is similar to forms attested in several AP languages. The AP forms are related and follow semi-regular sound changes (PAP initial *s > Kaera s, Blagar h, see Holton & Robinson 2017: 56). Therefore, this set potentially indicates a loanword from AP languages into Alorese (the Munaseli variety).

From the 596 concepts, the automatic detection filtered 167 sets of loan candidates, such as the one in Table 7.2 above. The resulting lexical similarity sets were inspected according to their pattern of distribution in the languages and dialects of the region. We manually checked the 167 loan candidates in more detail, to see whether the etymological relationship between the forms as hypothesized by LexStat makes sense beyond only the word lists. Of the potential 167 lexical similarity sets, 74 turned out to be erroneous, leaving us with 93 loan candidates. The erroneous cases include meaning mismatches, whereby, due to the different word order in Alorese and AP languages, the two aligned forms have formal similarity but are not semantically related. An example of an error due to a meaning mismatch is given in Table 7.3.

In Alorese, which is verb medial, the form for 'bite' is *gaki*, while *ata* means 'person' (*gaki ata* 'to bite someone'). In Blagar-Pura, which is verb final, the form for 'bite' is *adaŋ*, while *jabar* means 'dog'. The algorithm aligned Alorese *ata* 'person' with Blagar *adaŋ* 'bite' on the basis of formal similarity, but semantically the two forms are not related.

TABLE 7.3 Examples of a meaning mismatch due to word order

Concept	Language	Alignment	Form
to_bite	Alorese-Ternate	a t a -	gaki ata
to_bite	Blagar-Pura	a d a ŋ	jabar ing adang

The potential 93 loanwords were inspected more carefully to establish whether they are indeed AP loanwords in Alorese. Out of the 93 candidates, 28 turned out to be AP loanwords, the others were loanwords in the other direction (from Alorese into AP), or from Malay, or loanwords of unclear direction, or the resemblance was due to chance. In the following section, we present the 28 AP loanwords.

3 AP Loanwords in Alorese

In this section we present the AP loanwords organized by semantic fields, a choice shared by other contributions in this volume (e.g., Klamer, Edwards, and Schapper & Huber), to gain an additional perspective on the type of contact between the Alorese and the AP speakers. We assigned the AP loanwords to five semantic fields from the most prone to borrowing to the more resistant to borrowing: *Basic actions and technology* (§ 3.1), *Social and political relations* (§ 3.2), *Agriculture and vegetation* (§ 3.3), *The physical world and Animals* (§ 3.4), and a miscellaneous field including (*quantity, emotions, motion, kinship, the body, spatial relations, sense perception*) (§ 3.5). The semantic fields are those of Tadmor et al. (2012), but where slightly modified to be consistent to those of Edwards (this volume). *Basic actions and Technology* thus includes *Tools* as well as *Weapons*, and *The house*. The *Law and Religion and belief* were combined with the *Social and political relations* field. Unlike Edwards, we also combined *Animals* and *The Physical world*. Approximately half of the AP loanwords occur in the three most borrowable semantic fields (*Basic actions and technology, Social and political relations, and Agriculture and vegetation*).

In § 3.6 we will draw generalizations regarding their distribution among Alorese varieties, and their donor languages. All comparisons presented in this section were made with the tool EDICTOR, (etymological dictionary editor) at <https://digling.org/edictor/>. EDICTOR visualizes and allows to edit the cognate judgements in a lexical database. The tool also aligns similar sounds within the sets which helps to discover sound correspondences.

3.1 *Basic Actions and Technology*

3.1.1 'Fish trap'

The Alorese and AP forms for the concept 'fish trap' are presented in Table 7.4.

TABLE 7.4 Lexical similarity set associated with the concept 'fish trap'

Language	Alignment
Alorese-Pandai	k ε r
Blagar-Bakalang	v e r
Blagar-Bama	w e r
Blagar-Nule	k e r
Blagar-Tuntuli	v e r
Kula-Lantoka	g a r
Nedebang	tʃ a r
Teiwa	k ε : r
Wersing-Maritaing	- a r

The Alorese-Pandai word *ker* 'fish trap' is an innovation, different from the inherited form *puko?* < PFL *pukət 'fish trap' (Fricke 2019: 240) present in the other Alorese varieties. The source for this innovation are AP languages, which present similar forms. The sound changes among AP languages are semi-regular, because initial stops are usually retained among AP languages, so the change *k* > *v/w* in some Blagar varieties remains unexplained; final *-r is retained unchanged in all the AP languages, except Klamu. In Klamu, the PAP final *-r is expected to be lost (Holton et al. 2012: 94), but it is possibly irregular because a retention of *-r is also attested in PAP *dur > Klamu *dur* 'rat'. Due to the geographical spread of the word among AP languages, we consider this a loanword from AP languages, most likely Blagar-Nule or Teiwa, into Alorese-Pandai.

3.1.2 'Bed'

The Alorese and AP forms for the concept 'bed' are presented in Table 7.5.

Alorese-Pandai, Alorese-Munaseli and Alorese-Alor Besar have innovated the form *deki* for 'bed; raised platform'. This word is likely to be a loanword from AP languages which have similar forms that are related and reflect regular sound changes. Initial PAP *d is retained in all the languages. Medial *k is retained in Blagar, reflected as ? in Adang (Holton & Robinson 2017: 56). Among

the AP languages, Blagar, Reta, Kaera, and Western Pantar have the most similar form because the medial *k* is retained. From a geographical perspective, however, Blagar or Reta are likely the donor language(s) as this form is found in Alorese varieties spoken around the Alor-Pantar Strait.

TABLE 7.5 Lexical similarity set associated with the concept 'bed'

Language	Alignment
Alorese-Pandai	d e k i
Alorese-Munaseli	d e k i
Alorese-Alor Besar	d e k i
Western Pantar-Lamma	d e k i
Blagar	d e k i
Reta-Pura	d e k i
Teiwa	d e k
Adang	d e ?

3.1.3 'To fold'

The Alorese and AP forms for the concept 'to fold' are presented in Table 7.6.

TABLE 7.6 Lexical similarity set associated with the concept 'to fold'

Language	Alignment
Alorese-Pandai	l a k u k -
Kula-Lantoka	l a k u p -
Sawila	l a k u p i
Blagar	p i l i k u

For the concept 'to fold', the Alorese-Pandai variety in northeast Pantar uses both the inherited form *lepe* and an innovation *lakuk*. Among the AP languages, the most similar forms are *lakup(i)* in Kula and Sawila, *pəliku/piliku* in Blagar. It is unclear whether the AP words for 'to fold' presented in Table 7.6 are all related. In Sawila, *kupi* means 'to fold' (Kratochvíl 2014: 408), but the additional syllable *la-* is of unclear origin. In Blagar, *pi-* is an inalienable possessor for first

person plural inclusive (Steinhauer 2014: 182). Since at least some of the AP word seem to form a cognate set, while Alorese-Pandai is the only variety to use this form, we consider this a loanword from AP languages, and most likely Blagar into Alorese-Pandai. The change of the vowel from *i* to *a* (*liku* to *lakuk*) is also attested in other Blagar loanwords, such as Alorese *kalaki* ‘angry’ from Blagar *kilikil*, Alorese *reha* ‘monitor lizard’ from Blagar *rihi*, and Alorese *tera* ‘to close’ from Blagar *terij*.

3.1.4 ‘To pull’

The Alorese and AP forms for the concept ‘to pull’ are presented in Table 7.7.

TABLE 7.7 Lexical similarity set associated with the concept ‘to pull’

Language	Alignment
Alorese-Pandai	- w a k
Blagar-Kulijahi	a w a k
Blagar-Nule	a v a k
Abui	h a f i k
Klon	g ə b i k
Adang	? a b i ? i ŋ
Kabola	a p i ? i ŋ

For the concept ‘to pull’, Alorese-Pandai uses both the inherited form *vider* ‘to pull’ and an innovation *wak* ‘to pull’. For this concept, the majority of Alorese varieties use a Malay loan *tarek* ‘to pull’. The Alorese-Pandai word *wak* ‘to pull’ is possibly a Blagar loan, because a similar form *awak/avak* ‘to pull’ is attested in Blagar-Kulijahi and Blagar-Nule. The initial vowel *a-* is a prefix in Blagar indicating causative (Steinhauer 2014: 160, 194). This Blagar word seems to be related to the other AP words listed in the table. The sound changes are semiregular, as initial *b is reflected as *f* in Abui, and can be reflected as *v* if in intervocalic position (after the addition of a prefix) in Teiwa and Nedebang (Holton & Robinson 2017: 56), and in this case also in Blagar. The vowel *a* in Blagar remains difficult to explain, although Edictor found one other correspondence of Blagar *a* and, for instance, Klon *i*: Blagar-Nule *hava?* ‘house’ ~ Klon-Hopter *?əwi* ‘house’. The Blagar forms are formally the most similar, hence we identify Blagar as the donor language for this loan.

3.1.5 'To wash'

The Alorese and AP forms for the concept 'to wash' are presented in Table 7.8.

TABLE 7.8 Lexical similarity set associated with the concept 'to wash'

Language	Alignment
Alorese-Baranusa	--l a m i ŋ
Alorese-Munaseli	--l a m i ŋ
Alorese-Pandai	--l a m i ŋ
Adang-Lawahing	--l a: m --
Adang-Otvai	--l a m --
Reta	--l a: m i ŋ
Deing	--l a n a ŋ
Hamap	n a l a m --
Kabola	--l a m --
Kafoa	- u l a m --
Western Pantar-Lamma	--l a m i ŋ

For the concept 'to wash', some Alorese varieties on Pantar, have innovated the form *lamiŋ*, next to the inherited *bema* ('to wash' for clothes) and *hue* ('to wash' for dishes). The Alorese word *lamiŋ* 'to wash' in the varieties of Baranusa, Pandai, and Munaseli appears to be a loanword from an AP source. This form is an inherited AP form, with related forms in several AP languages, as can be seen in Table 7.8. Reta (*laamiŋ*) and Western Pantar-Lamma (*lamiŋ*) have the most similar forms to Alorese and both these languages are in contact with Alorese on Pantar; Reta is close to Munaseli and Pandai, while Western Pantar is close to Baranusa. Therefore, these are the most likely donor languages. This AP loanword is also mentioned by Klamer (2011: 105) and by Robinson (2015: 28), both pointing to Western Pantar as the donor language.

3.2 *Social and Political Relations*

3.2.1 'To pray'

The Alorese and AP forms for the concept 'to pray' are presented in Table 7.9.

TABLE 7.9 Lexical similarity set associated with the concept 'to pray'

Language	Alignment
Alorese-Munaseli	- g a m a r
Kaera	a ? - m u r
Western Pantar-Lamma	- h a m u r
Reta	- - a m u r
Teiwa	- h a m a r

The Alorese-Munaseli variety in northeast Pantar innovated *gamar apa* for 'to pray', while in the other Alorese varieties the more widely used term for 'to pray' is *sabeay* (< Malay loan *sambayang* [samba^hian]) 'to pray; to worship God'). The Munaseli form *gamar apa* comprises *gamar* (external origin) and *apa* (Alorese 'something'). In the set for the concept 'to pray', it seems that Alorese-Munaseli has borrowed *gamar* from a neighboring AP language, such as Teiwa, which has *hamar* for 'pray'. Since cognates of the Teiwa form for 'to pray' are attested across several AP languages, it is likely that this is an inherited AP form. Teiwa is very likely to be the donor because the vowels are identical to the Alorese-Munaseli word *gamar*. The initial *g* in the Alorese-Munaseli word may come from the Teiwa form *ga-hamar* 'pray for someone', whereby *ga-* is a third person singular pronoun (Klamer 2010: 55).

3.2.2 'Adultery'

The Alorese and AP forms for the concept 'adultery' are presented in Table 7.10.

TABLE 7.10 Lexical similarity set associated with the concept 'adultery'

Language	Alignment
Alorese-Alor Besar	b u h a
Alorese-Munaseli	b u h a

TABLE 7.10 Lexical similarity set associated with the concept ‘adultery’ (cont.)

Language	Alignment
Kaera	b u s -
Blagar-Pura	b u h a
Reta	b u h a
Teiwa	b u: s -

No similar forms to the Alorese-Alor Besar and Alorese-Munaseli word *buha* ‘adultery’ are attested in the near-by Flores-Lembata languages and no proto forms are available for this concept. Conversely, the AP forms are historically related and reflect regular sound changes. The initial PAP *b is expected to be retained unchanged in all the languages; the final PAP *s is expected to be retained regularly as s in Teiwa, Kaera, and Sar, and changed into h in Blagar and in Reta (Holton & Robinson 2017: 56). This sound change is attested in several Blagar words, such as PAP *mis > Blagar *mihi* ‘sit’ and PAP *bis > Blagar *bihī* ‘mat’, where an epenthetic vowel is added after the weakening of *s. Given the presence of the glottal fricative h and the vowel in Alorese varieties, we conclude that Alorese-Alor Besar and Alorese-Munaseli borrowed *buha* from either Blagar or Reta.

3.3 *Agriculture and Vegetation*

3.3.1 ‘Digging stick’

The Alorese and AP forms for the concept ‘digging stick’ are presented in Table 7.11.

TABLE 7.11 Lexical similarity set associated with the concept ‘digging stick’

Language	Alignment
Alorese-Bana	- - n o r u ?
Alorese-Helangdohi	- - n ɔ r u ?
Blagar-Nule	- - n o r u k
Reta	h a n o: r u k
Western Pantar	- - s o r u -

The table shows that the Alorese-Bana and Alorese-Helangdohi varieties in northeast Pantar innovated *noru?* ‘digging stick’. The other Alorese varieties use the word *kuaŋ* inherited from PWL (*nuan, Sulistyono 2022: 255).

Among the AP languages Blagar-Nule (*noruk*), Reta (*hanoruk*), and Western Pantar (*soru*) show related forms for the concept ‘digging stick’, these forms seem to go back to a form like #sVnoru(k) ‘digging stick’. It seems that a semantic change occurred in Western Pantar to the relatively close concept ‘stick; pole’. The sound changes are semiregular. The initial *s is regularly reflected as *h* in Reta and retained unchanged in Blagar and Western Pantar (Holton & Robinson, 2017: 56). The intervocalic *n is expected to be retained unchanged in all languages, however the sequence -Vn- is lost in Western Pantar (#sVnoru(k) > *soru*). The intervocalic *r shows irregular reflexes in Western Pantar because it is expected to be retained as *l*. Finally, the final *k is expected to be lost in Blagar and retained in Western Pantar (Holton & Robinson 2017: 56), but here we see the opposite pattern. Even though the sound correspondences among the AP languages are semi-regular, we consider the form inherited in AP languages. We, therefore, consider that the Alorese word *noru?* is a loanword and that the most likely donor for this concept is Blagar-Nule which has the form *noruk*, most similar to the Alorese form *noru?*.

3.3.2 ‘Garden’

The Alorese and AP forms for the concept ‘garden’ are presented in Table 7.12.

TABLE 7.12 Lexical similarity set associated with the concept ‘garden’

Language	Alignment
Alorese-Munaseli	b u t a ?
Adang-Lawahing	b u t u -
Adang-Otvai	b u t - -
Blagar-Warsalelang	b u t a x
Blagar-Tuntuli	b u t a q
Blagar-Pura	b u t a
Kabola	b u t u ?

The table shows that Alorese-Munaseli in northeast Pantar innovated *ekaŋ buta?* for ‘garden’. For the concept ‘garden’, the general Alorese term that goes back to PWL is *ekaŋ* ‘garden’ (Sulistyono 2022:255). However, the Munaseli vari-

ety uses a compound *ekaj buta?*, which comprises an inherited form *ekaj* (< PWL *eka ‘garden’) and the new form *buta?*, which is of external origin.

The AP words meaning ‘garden’ are clearly related and attested in AP languages spoken around the Alor-Pantar Strait. Possibly, a Proto Nuclear Alor-Pantar form *butVq ‘garden’ could be reconstructed based on this cognate set (Kaiping & Klamer 2019: 35). Alorese-Munaseli has borrowed the form *buta?* to form a compound *ekaj buta?* ‘garden’. The donor language is most likely a Blagar variety because they have the most similar forms and are geographically close to Munaseli. The lenition of final stop *x/q* in Blagar into a glottal stop in Munaseli is expected because Alorese does not allow final *x/q*.

3.3.3 ‘Rattan’

The Alorese and AP forms for the concept ‘rattan’ are presented in Table 7.13.

TABLE 7.13 Lexical similarity set associated with the concept ‘rattan’

Language	Alignment
Alorese-Munaseli	l u - a
Alorese-Pandai	l u - a
Blagar-Kulijahi	l i - a
Blagar-Nule	l i j a
Blagar-Pura	l i - a
Blagar-Bama	l e g
Blagar-Tuntuli	l e - g
Blagar-Warsalelang	l e : - g
Reta	l i - a g
Reta	l i j a g
Kaera-Abangiwang	l ε : - g
Kabola-Monbang	l o j o ?
Blagar-Bakalang	l i j a
Adang-Otvai	l e
Teiwa-Lebang	l i j a g
Kui-Labaing	l e
Adang-Lawahing	l ε ?
Deing	l i a x
Sar-Adiabang	l i j a h
Sar-Nule	l i j a g

We assume that the form *lua* ‘rattan’ in Alorese-Pandai and Alorese-Munaseli comes from an external source. This form is used alongside a different form *ue/uwe*, inherited from PFL **uay* ‘rattan’ (Fricke 2019: 477). For this concept, regular sound correspondences can be seen among AP languages, namely PAP initial **l* is retained unchanged in all languages, as expected. There is no PAP sound that is reflected as final *g*, but the synchronic correspondences are fairly regular: for instance, Teiwa *bag* ~ Deing *bax* ‘seed’, Teiwa *og* ~ Deing *ox* ‘hot’. Interestingly, no AP language shows the vowel combination *ua* found in the Alorese form *lua*, so we hypothesize that Alorese borrowed the form *lia* from Blagar, but it changed the diphthong from *ia* to *ua*, to be reminiscent of the inherited PFL form **uay* ‘rattan’.

3.3.4 ‘Root’

The Alorese and AP forms for the concept ‘root’ are presented in Table 7.14.

TABLE 7.14 Lexical similarity set associated with the concept ‘root’

Language	Alignment
Alorese-Alor Kecil	- a l i - - η
Alorese-Dulolong	- a l i - - η
Adang-Lawahing	- a l i ? i η
Adang-Otvai	- a l i ? a η
Hamap	- a l i - a η
Kabola	h a l i ? i η
Kafoa	- ɾ l i : k a η
Kamang	- a l i : - - - -
Abui	- a i - - - -

For the concept ‘root’, the majority of the Alorese varieties use the inherited form *ramu?* (< PMP **Ramut* ‘root’, see Sulistyono 2022: 265). The form *aliη* is an innovation in Alorese-Alor Kecil and Alorese-Dulolong in the Alor Peninsula and it is borrowed from AP languages. The PAP intervocalic **-l-* is retained unchanged in all modern-day AP languages (Holton et al. 2012: 94), except in Abui where it is lost. It seems that Alorese borrowed the word *aliη* ‘root’ from Adang, Kabola, or Hamap. These languages are located close to the Alor Peninsula varieties. Among these languages, Adang is most likely to be the donor. The Adang-Lawahing word *ali?iη* is the most similar to Alorese *aliη* ‘root’. It is

likely that Alorese borrowed the word from Adang-Lawahing and dropped the ʔ in the process, as Alorese varieties spoken on the Alor Peninsula and in the Strait do not have word medial glottal stop.

3.3.5 'Taro'

The Alorese and AP forms for the concept 'taro' are presented in Table 7.15.

TABLE 7.15 Lexical similarity set associated with the concept taro

Language	Alignment
Alorese-Alor Besar	a - g ɔ l
Alorese-Alor Kecil	g o l o -
Alorese-Dulolong	g ɔ l ɔ -
Alorese-Baranusa	g o l o -
Blagar-Pura	a u
Blagar-Bakalang	a w g o l
Blagar-Kulijahi	a w g o l
Blagar-Nule	a w g o l
Blagar-Pura	g o l - -
Adang-Otvai	a - g ɔ l
Hamap	a - k o l
Reta	a i g o l

The Alorese word for 'taro' is formed by two elements, the word *au/ai* is possibly from PFL *kayu 'tree' (see Fricke 2019: 521) and the word *gol/golo* which is likely an innovation borrowed from AP languages. Alorese-Dulolong has *kaɟo golo* with the form *kaɟo* 'tree; wood', also from PFL *kayu 'tree'.

Among the AP languages, the words *awgol* (Blagar-Bakalang), *au gol* (Blagar-Pura), *agɔl* (Adang-Otvai), and *ai gol* (Reta) are most similar to the Alorese forms. The AP forms are also formed by an element *au/ai/a* meaning 'tree' or 'tuber' and an element *gol/go/hol/ho* meaning 'taro'. This pattern is found also to refer to other tubers, as for instance Blagar-Pura *au benu* 'cassava', and *au kasi* 'sweet potato'. Languages where these form for 'taro' are used are located close to the Alor Peninsula, and the most likely donor seems to be Blagar or Adang. The fact that Alorese-Baranusa also uses a similar form, *au golo*, for 'taro' may indicate that this word was borrowed early on, although the fact that is absent in the northeastern Pantar varieties may also indicate a

later borrowing from Alorese varieties in the Alor Peninsula to the Baranusa variety.

3.4 *Animals and Physical World*

3.4.1 'Coral rock'

The Alorese and AP forms for the concept 'coral rock' are presented in Table 7.16.

TABLE 7.16 Lexical similarity set associated with the concept 'coral rock'

Language	Alignment
Alorese-Munaseli	kɔ - k a -
Blagar-Bakalang	k o - k a -
Blagar-Bama	k o - q a s
Kaera	q o ? q i s
Teiwa	q o - q a s
Adang-Otvai	? o ? - o i
Kabola	k o ? - o i

The Alorese-Munaseli innovation *koka* for the concept 'coral rock' is a loanword from AP languages, in which the forms for this concept are cognates. The sound changes are semi-regular: PAP initial and intervocalic *q is reflected as *k* in Blagar, *q* in Teiwa, and *?* in Adang (Holton & Robinson 2017: 56); Kaera is expected to have *x* but has *q*, and Adang is expected to have zero in intervocalic position, but here may have a glottal stop to avoid a sequence of two identical vowels. The correspondence of Teiwa *q* and Kabola *k* is regular, and attested in other words such as Teiwa *qab* ~ Kabola *kaba* 'spear' and Teiwa *qarnuk* ~ Kabola *karnu* 'ten'. The PAP final *s is retained in Teiwa and Kaera, and reflected as *h* in Blagar and Adang (Holton & Robinson 2017: 56), here however, Adang and Kabola have final *i* and Blagar-Bama has retained the final *s*. The correspondence *a* ~ *i* in Teiwa and Kaera is attested also in other forms such as Teiwa *saxa?* ~ Kaera *si?aq* 'chicken', Teiwa *hasak* ~ Kaera *is?ik* 'empty'. The most likely donor for this concept is Blagar-Bakalang which has the form *koka*, identical to the Alorese one.

3.4.2 'Mud'

The Alorese and AP forms for the concept 'mud' are presented in Table 7.17.

TABLE 7.17 Lexical similarity set associated with the concept 'mud'

Language	Alignment
Alorese-Alor Kecil	b a n a -
Alorese-Dulolong	b a n a -
Alorese-Ternate	b a n a -
Blagar-Pura	b a n a k u ŋ
Nedebang	b a n a q a
Reta	b a n a k u ŋ
Sar	b e n a : q
Abui	f a n a q

The Alorese word *bana* 'mud'⁴ has no similar forms in the neighboring Flores-Lembata languages or in the proto languages. This innovation is likely to come from AP languages. For the concept 'mud', it looks like that the AP forms go back to a form like #banak or #banaq. The sound changes are regular, as PAP initial *b is retained in all languages, but Abui where it is reflected as *f* (Holton & Robinson 2017: 56). The addition of a final syllable *-uŋ* in Blagar and Reta remains unclear, although similar additions of final syllables are found in the Strait Alorese varieties, where the final syllables *-uŋ*, *-iŋ*, and *-aŋ* are added to some words.⁵ Alorese varieties on Alor Peninsula and in the Strait apparently borrowed *bana* 'mud' from either Blagar or Reta, as these languages have the most similar form.

4 Note that the word *bana* in Alorese also means 'forest', from PMP *banua 'inhabited land, territory supporting the life of a community'. Robinson (2015: 22) considers the word for 'forest' an Alorese loan into AP languages: Alorese *banna* 'forest' (cf., Lamaholot (Ile Ape) *bənanawa* 'forest') > Retta *vana*, Adang *bana*, Kula *banan* 'forest'.

5 The Alorese varieties Ternate, Buaya, Alor Besar, Alor Kecil, and Dulolong, which are spoken in the Alor-Pantar Strait area form a subgroup that is based on the exclusively shared sound change of PAL *w > *f* in all positions, PAL *ai > *ei* in word-final position, and the addition of the syllables *-uŋ*, *-iŋ* and *-aŋ* in final position in some words (see Sulistyono 2022: 216).

3.4.3 'Gravel'

The Alorese and AP forms for the concept 'gravel' are presented in Table 7.18.

TABLE 7.18 Lexical similarity set associated with the concept 'gravel'

Language	Alignment
Alorese-Alor Besar	b a l ɔ f a -
Alorese-Alor Kecil	b - l o f a -
Alorese-Dulolong	b a l o f a -
Alorese-Munaseli	g - l o w a r
Blagar-Bakalang	g ə d o b a r
Blagar-Kulijahi	g ə d o w a r
Blagar-Nule	g ə n o v a r
Blagar-Warsalelang	d o l o w a r
Teiwa	d a l a w a r
Adang-Otvai	d a r o f e
Sar	- - - - - w a r
Deing	d a l a w i r

For the concept 'gravel' Alorese-Alor Besar, Alorese-Alor Kecil, and Alorese-Dulolong on the Alor Peninsula innovated the forms *balofa* and *blofa* respectively, while Alorese-Munaseli in northeast Pantar innovated *gelovar*. The other Alorese varieties have an inherited form similar to *vato kar:ik* 'gravel' which is constituted of *vato* (< PMP *batu 'stone') and *kar:ik* (< PMP *kədi 'small', Blust & Trussel 2020).

A number of similar forms for the concept 'gravel' are found in several AP languages. Although it is not clear whether all the AP forms are cognates, most of them are related and are likely to be inherited. The forms *dobar/dowar/nowar/lowar/lawar* are preceded by the syllable *gə* in some Blagar varieties, and by the syllable *do/da* in Blagar-Bama and Blagar-Warsalelang, and in Adang, Teiwa, Sar, and Deing. In Blagar, *do-* is a deictic morpheme that means 'up there' (Steinhauer 2014: 159). The correspondence of *d* and *n* in the two words *gedobar* (Blagar-Bakalang) and *genowar* (Blagar-Nule) is also seen in *kədumu* (Blagar-Bakalang) and *kənumu* (Blagar-Nule) 'to suck'.

The form *gelovar* 'gravel' in Alorese-Munaseli is possibly borrowed from several different sources or is a mixed of different forms. The initial syllable *ge-* is similar to Blagar-Bakalang, Kulijai and Nule; the part *-lowar* is similar to Blagar-Warsalelang *dolowar*. The word *balofa* on the Alor Peninsula is also similar to

the form in Blagar-Warsalelang *dolowar* 'gravel'. The sound change $w > f$ is regular in the varieties of the Alor Peninsula and attested also in other loanwords, such as *safa* 'rice field' from Malay *sawah*. It is unclear why Alorese Alor Besar and Alor Kecil added the initial syllable *ba-*, one possible explanation may be that *ba-* is a shortening of the Malay word *batu* 'stone'. The most likely donor seems to be Blagar-Warsalelang with the form *dolowar* 'gravel', although the differences in forms may point to independent borrowing events in the Alorese varieties.

3.4.4 'Dolphin'

The Alorese and AP forms for the concept 'dolphin' are presented in Table 7.19.

TABLE 7.19 Lexical similarity set associated with the concept 'dolphin'

Language	Alignment
Alorese-Alor Besar	k u dʒ a - e
Alorese-Alor Kecil	k u dʒ a - i
Alorese-Bana	k u j a - -
Alorese-Buaya	- u dʒ a - e
Alorese-Dulolong	k u dʒ a h i
Alorese-Helangdohi	k u dʒ a - -
Alorese-Munaseli	k u dʒ a - -
Alorese-Pandai	- u dʒ a - -
Alorese-Ternate	k u dʒ a - e
Alorese-Wailawar	k u dʒ a - -
Adang-Lawahing	- u s a h a
Adang-Otvai	- u s a h -
Blagar-Bakalang	k u dʒ - - a
Blagar-Bama	k u dʒ - - a
Blagar-Kulijahi	k u dʒ - - a
Blagar-Nule	k u dʒ - - a
Blagar-Tuntuli	k u dʒ a h -
Blagar-Warsalelang	k u dʒ - - a
Deing	k u - i - -
Kaera	x u j a - -
Blagar-Pura	k u j a - -
Sar-Adiabang	k u j a - -
Teiwa	k u j a ʔ -

The word for ‘dolphin’ is widely attested both in Alorese varieties and in the AP languages. In other Austronesian languages to the far north of Alor, similar forms are attested; *uas* in Geser-Gorom (south Maluku), and *kuraf* in Uruangnirin (spoken on west Papua), and a Proto Oceanic form **kuriap* ‘dolphin’ has been reconstructed (Blust & Trussel 2020). However, no similar form is attested in the closest relatives of Alorese, the near-by Flores-Lembata languages, and no PMP forms are available for this concept. For this reason, we consider this to be an innovation in Alorese, possibly a loan from AP languages.

In the AP languages, there are similar forms showing regular sound correspondences, which indicate shared ancestry. The sound correspondences enable the reconstruction of a possible early AP form **kujasi* ‘dolphin’. The initial **k* is retained in all AP languages, except Karea where it is reflected as *x* (*xuja*), and Adang where it is usually reflected as glottal stop, but here it is lost (*usaha*). The approximant **j* is retained in Teiwa, Kaera, Sar and Blagar-Pura, but lost or changed in others, such as Adang where we find *s*. Medial **s* is retained, except in Adang where it is regularly reflected as *h* (*usaha*).

In Blagar, the approximant [j] only occurs in the interjection *jo* ‘yes’ and a few borrowings, such as the recently adopted Christian name *Yohan* [johan] and the word *rayat* [rajat], borrowed from Indonesian *rakyat* ‘the people’. With this evidence, the most likely scenario for this concept seems to be that PAP, the ancestor of AP languages, borrowed the form from an Austronesian donor and when the Alorese arrived in the Alor archipelago, they re-borrowed the form from AP languages. The similarity between the Alorese *kuḏḏae* and the Blagar word *kuḏḏa* ‘dolphin’ may also indicate recent contact, with Blagar then re-borrowing the Alorese form more recently.

3.4.5 ‘Monitor lizard’

The Alorese and AP forms for the concept ‘monitor lizard’ are presented in Table 7.20.

TABLE 7.20 Lexical similarity set associated with the concept ‘monitor lizard’

Language	Alignment
Alorese-Alor Besar	----- r e h a
Alorese-Alor Kecil	----- r e h a
Alorese-Dulolong	----- r e h a

TABLE 7.20 Lexical similarity set associated
with the concept 'monitor lizard'
(*cont.*)

Language	Alignment
Blagar-Bakalang	-----rihi
Blagar-Bama	-i-ris-
Blagar-Kulijahi	-----rihi
Blagar-Nule	-----ri--
Blagar-Tuntuli	-i-ris-
Blagar-Warsalelang	-i-ris-
Blagar-Pura	-a-ri--
Deing	je-ris-
Kaera	-iʔris-
Kaera	tɛʔres-
Klon-Hopter	wə-rih-
Kui	-----ros-
Nedebang	-----lisi
Sar-Adiabang	ji-ris-
Sar-Nule	-----ris-
Teiwa	-----ris-

The Alorese varieties on the Alor Peninsula and in the Strait have innovated the word *reha* 'monitor lizard'. The Alorese varieties on Pantar retain the inherited form *eto/teto damar* (< PWL *eto 'monitor lizard' see Sulistyono 2022: 260; *damar* is of unknown origin).

From the distribution and the regular sound changes, it is evident that the forms found among AP languages are a cognate set and go back to a proto form (Robinson reconstructed PAP *IVsi 'monitor lizard', 2015: 29)

Based on the evidence presented in Table 7.20, we conclude that the donor language for the Alorese word *reha* is probably Blagar (Bakalang and Kulijahi) because it has the form *rihi*, which is the most similar to Alorese *reha*. Reasons for the change of the non-final vowel *i* to *e* remain unclear, but the Alorese final *a* from Blagar *i* in loanwords seems regular, as seen earlier in the Alorese *tera* 'to close' from Blagar *terij* 'to close', and Alorese *lakuk* 'to fold' from Blagar *liku* 'to close'.

3.5 *Miscellaneous: Quantity, Emotions, Motion, Kinship, the Body, Spatial Relations, Sense Perception*

3.5.1 'Ten'

The Alorese and AP forms for the concept 'ten' are presented in Table 7.21.

TABLE 7.21 Lexical similarity set associated with the concept 'ten'

Language	Alignment
Alorese-Alor Besar	k a - r - t o u -
Alorese-Alor Kecil	k a - r - t o u -
Alorese-Bana	k a - r - t o u -
Alorese-Baranusa	k a - r - t o u -
Alorese-Beang Onong	k a - r - t o u -
Alorese-Buaya	k a - r - t o u -
Alorese-Dulolong	k a - r - t o u -
Alorese-Helangdohi	k a - r - t o u -
Alorese-Kayang	k a - r - t o u -
Alorese-Munaseli	k ə - r - t o u -
Alorese-Pandai	k ə - r - t o u -
Alorese-Ternate	k a - r - t o u -
Alorese-Wailawar	k a - r - t o u -
Abui-Fuimelang	k a - r - - - - - - - -
Abui-Petleng	k a - r - n u k u
Abui-Takalelang	k a - r - n u k u
Abui-Ulaga	k a - r - n u k u
Adang-Lawahing	- a i r - n u - -
Adang-Otvai	? e - r - n u - -
Abui-Atimelang	k a - r - - - - - - - -
Blagar-Bakalang	- a - r - n u - -
Blagar-Bama	q a - r - n u k u
Blagar-Kulijahi	- a - r - - - - - - - -
Blagar-Nule	- a - r - n u - -
Blagar-Tuntuli	q a - r - n u k -
Blagar-Warsalelang	x a - r - - - - - - - -
Blagar-Pura	- a - r i n u - -
Deing	q a - r - n u k -
Hamap-Moru	- a i r - n u - -
Kabola	k a - r - n u - -

TABLE 7.21 Lexical similarity set associated with the concept 'ten' (*cont.*)

Language	Alignment
Kaera	x a - r - - - - - - - -
Kafoa	k a - r - n u k u
Kamang	k a - r - n ɔ k -
Klon-Bring	k a - r ə n ɔ k -
Klon-Hopter	k a - r - n u k -
Kiramang	k a - r - n u k u
Kui	k a - r - n u k u
Western Pantar-Lamma	k e - - a n u k u
Nedebang	k a - - - - - - - - - -
Reta	k a - r a n u - -
Sar-Adiabang	q a - r - n u k -
Sar-Nule	q a - r - n u k -
Teiwa	q a : - r - - - - - - - -
Proto Alor-Pantar	q a - r - - - - - - - -
Proto Alor-Pantar	q a - r - - - - - - - -
Proto Timor-Alor-Pantar	q a - r - - - - - - - -

The Alorese numeral *kartou* 'ten' is formed combining the decimal base *kar* 'tens' and the numeral *tou* 'one'. The form for the decimal base *kar* is a borrowing from AP languages, while the numeral 'one' *tou* is inherited (< PWL **tou*) 'one'. Besides the phonological material, Alorese also borrowed the pattern of forming 'ten' as 'ten-one' from AP languages (see Schapper & Klamer 2017 for an extensive description of numerals in Alor-Pantar languages). This is an innovation only present in Alorese, absent from the other Flores-Lembata languages, which all preserve reflexes of the Proto Austronesian form **puluq* for 'ten' (Schapper & Klamer 2017: 320 ff.). This loan has also been discussed in Klamer (2011), Robinson (2015), and Moro (2018).

The PAP word **qar-* 'tens' has been reconstructed by Holton et al. (2012: 115). As described above, it seems that Alorese only borrowed the part of the numeral that marks tens, *kar-*, but retained the PWL form **tou* 'one' (Sulistyono 2022: 428). Since the form is present in all Alorese varieties, it is likely to be an old loan (see § 3.1). The donor is likely to be one which has initial *k* (and most likely one which has the exact syllable *kar*) because Alorese varieties also have initial *kar-*. Among the AP languages that have *kar*, the donor is most likely

one which is spoken close to the coast or located around the Alor-Pantar Strait, such as Klón or Reta.

3.5.2 'Angry'

The Alorese and AP forms for the concept 'angry' are presented in Table 7.22.

TABLE 7.22 Lexical similarity set associated with the concept 'angry'

Language	Alignment
Alorese-Alor Besar	k a - l a k i -
Alorese-Alor Kecil	k a - l a k i -
Alorese-Baranusa	k - - l i k i l
Alorese-Buaya	k - - l e k i -
Alorese-Dulolong	k ə - l a k i -
Alorese-Kayang	k - - l i k i -
Alorese-Munaseli	k - - l i k i l
Alorese-Pandai	k - - l i k i l
Alorese-Ternate	k a - l a k i -
Alorese-Wailawar	k - - l i k i l
Blagar-Warsalelang	k i - l i k i l
Blagar-Manatang	- - a l i ? i l
Blagar Kulijahi	- - - - - - - l i l
Kaera	k e ? l i k i l
Klón-Hopter	k ə - l i k - -
Western Pantar-Lamma	k - - - - a k i ŋ
Sar-Adiabang	k - - - - a k a -
Teiwa	k ə - l e x e l

No similar forms to the Alorese word for 'angry' are attested in the near-by Flores-Lembata languages and no proto forms are available for this concept. In some Alorese varieties, the concept 'angry' is a compound consisting of an inherited root *onoŋ* 'inside' (< PFL *una 'house; inside; hole', Fricke 2019: 464) and the AP loanword *kelikil*. This word is likely a loan from AP languages which present similar forms, for the same concept, and which reflect semi-regular sound changes: PAP initial and medial *l is retained in all languages, with Western Pantar as an exception; PAP medial *k is retained unchanged, but reflected as x in Teiwa (Holton & Robinson 2017: 56). The correspondence of Blagar k

and Teiwa *x* is regular because it is also seen in other words, such as Blagar *tekil* ~ Teiwa *taxal* ‘thin’ and Blagar *sokil* ~ Teiwa *soxai* ‘to dance’. Some varieties of Blagar have weakened and eventually lost the intervocalic *k and have *ali?il*, as in Blagar-Manatang, and *lil* as in Blagar Kulijahi (possibly from a form like *kilikil* as in Blagar-Warsalelang). Weakening of intervocalic *k is found also in Kabola, for instance the Blagar *k* and Kabola *?* correspondence is regular as seen in other words, such as Blagar *trukinuk* ~ Kabola *ti?inu* ‘nine’ and Blagar *tatoku* ~ Kabola *ato?o* ‘stomach; belly’.

Given that the AP lexemes seem to form a historically related set, and that there are no similar forms attested in the other Flores-Lembata languages, we conclude that the Alorese varieties borrowed this form from AP languages, most likely from Blagar or Kaera. The Alorese varieties spoken on and around the Alorese peninsula (Alor Besar, Alor Kecil, Dulolong, and Ternate) have the form *kalaki*, whereby the vowels *i* have been changed into *a*. The change of the vowel from *i* to *a* is also attested in other Blagar loanwords, such as Alorese *tera* ‘to close’ from Blagar *terij*, and Alorese *reha* ‘monitor lizard’ from Blagar *rihi*, and Alorese *lakuk* ‘to fold’ from Blagar *liku*.

3.5.3 ‘Road’

The Alorese and AP forms for the concept ‘road’ are presented in Table 7.23.

TABLE 7.23 Lexical similarity set associated with the concept road

Language	Alignment
Alorese-Baranusa	-- t ɔ : r
Alorese-Munaseli	-- t ɔ r
Alorese-Beang Onong	-- t ɔ r
Deing	w u t o r
Deing	-- t o r
Kaera	-- t o r
Teiwa-Lebang	y i t a r
Western Pantar	y a t o r
Kafoa	y a -----
Kui	y a -----
Abui-Takalelang	j a -----

In some Alorese varieties on Pantar, Alorese-Baranusa, Beang Onong and Munaseli, we find the AP loan *tor* ‘road’. Klamer (2011: 105) and Robinson (2015: 28) list this form as an AP loanword into Alorese, from Western Pantar *ya tor* ‘road’. The form *ya tor* is widespread among AP languages. The part *ya* is the AP word for ‘road’ and *tor/tar* is found in Kaera, Deing, Teiwa, and Western Pantar. The *tor/tar* element is semantically related to the word for ‘tail’ in AP languages (PAP *ora ‘tail’, see Holton & Robinson 2017: 78), such as Teiwa *t-or* ‘tail; tail-bone’ and Klon *t-or* ‘bone’, both with the possessive prefix *t-*. We suggest that a semantic shift from ‘tail’ to ‘main road’ has taken place in some languages, probably due to the fact that a road with curves does resemble an animal’s tail. Western Pantar is the only language where the compound is still complete. The other languages have either lost the *ya* part or the *tor* part. However, it is also possible that the varieties that only have *ya*, like Kafoa, Abui and Kui, might never have had the compound *ya tor*. In Abui-Takalelang, *foqa* means ‘big’; thus, *ja foqa* means ‘big road; highway’. Western Pantar and Deing are the most likely donor for this loanword.

3.5.4 ‘Younger sibling’

The Alorese and AP forms for the concept ‘younger sibling’ are presented in Table 7.24.

TABLE 7.24 Lexical similarity set associated with the concept ‘younger sibling’

Language	Alignment
Alorese-Bana	-- k a u -
Alorese-Buaya	-- k a u -
Alorese-Munaseli	-- k a u -
Alorese-Pandai	-- k a u -
Blagar-Kulijahi	-- k a - w
Blagar-Warsalelang	-- k a - w
Blagar-Bakalang	-- k a - w
Blagar-Nule	n e k a - w
Blagar-Tuntuli	p i k a - w
Blagar-Pura	- e k a k u
Sawila	n i k a k u
Kula-Lantoka	ŋ a k a k u
Teiwa	n a k a ? a w
Wersing-Maritaing	n ε k a u k

TABLE 7.24 Lexical similarity set associated with the concept ‘younger sibling’ (cont.)

Language	Alignment
Wersing-Taramana	n e k a k u
Reta-Ternate	g a k a k u
Reta-Pura	- - k a k u
Western Pantar-Lamma	- i a k u

As for the concept of ‘younger sibling’, the form *kau* is quite widespread among the Alorese varieties. In Alorese-Kayang and Alorese-Wailawar, a medial glottal stop has been inserted. Another term for this concept in Alorese is *aring* ‘younger sibling’, which is related to the Lamholot forms *aring/arik* (Fricke 2019: 529).

The Alorese *kau* form shows similarities with several AP languages, in which the forms go back to PTAP **kaku* ‘younger relative’ (see Schapper & Huber, this volume). In some AP languages, the form is presented with a possessive prefix. Blagar, Kaera, and Teiwa are all possible donors.

Interestingly, it seems that this form is highly borrowable, as it listed by Schapper and Huber (this volume), among the TAP etyma into the Austronesian languages of Timor. Unlike on Alor and Pantar, where the form has been borrowed together with its original meaning, on Timor the form has undergone a semantic shift from PTAP **kaku* ‘younger relative’ to Makasae and Makalero ‘small’.

3.5.5 ‘To bury’

The Alorese and AP forms for the concept ‘to bury’ are presented in Table 7.25.

TABLE 7.25 Lexical similarity set associated with the concept ‘to bury’

Language	Alignment
Alorese-Alor Besar	t - - o - - u -
Alorese-Buaya	t - - o - - u -
Alorese-Dulolong	t - - u - h o
Alorese-Alor Kecil	t - - o - h u -

TABLE 7.25 Lexical similarity set associated with the concept 'to bury' (cont.)

Language	Alignment
Alorese-Ternate	t - - o - - u -
Blagar-Bakalang	t - - o - - - - w
Blagar-Bama	t - r o - k u -
Blagar-Kulijahi	t - r o - - u -
Blagar-Nule	t ə r o - - - - w
Blagar-Tuntuli	t o r o - k u -
Blagar-Warsalelang	t ə r o - k u -
Blagar-Pura	t a r o - - u -
Kaera	t - r a ? q o -
Makasae	t a r - - - - u -
Teiwa	t a r a - x a ?
Kamang	f o i u

The Alorese word *tou* 'to bury' is likely to be an AP loanword in Alorese varieties on the Alor Peninsula and in the Strait, because no similar forms are attested in the near-by Flores-Lembata languages and no proto forms are available for this concept. Conversely, the AP forms are historically related and reflect regular sound changes (Blagar-Bakalang *tou*, Blagar-Tuntuli *toroku*, Teiwa *taraħa?*, Kamang *fo.u*). Proto AP *tVroqu 'to bury' may be reconstructed because initial *t- is attested regularly in most of the AP languages and the intervocalic *-r- is also expected to appear unchanged in most of the languages. In Kamang, intervocalic *-r- is expected to change into *l*, but in one of the varieties in the Kamang cluster, namely Tiyei, it has changed into *ɹ*. As for the vowels, the correspondence Blagar *o* and Teiwa *a* is regular, and attested in other words such as Blagar-Tuntuli *bogori* 'yellow' ~ Teiwa *bahari* 'yellow'.

Among the AP languages that have a reflex of this form, Blagar-Bakalang has the most similar form to Alorese, suggesting that Alorese borrowed *tou* 'to bury' from this Blagar variety. In Alor Kecil, the addition of intervocalic *h*, as seen in *tohu* 'to bury' is also seen in other words, such as Alor Besar *tafeuy* ~ Alor Kecil *təfihuy* 'fog'.

3.5.6 'Heart'

The Alorese and AP forms for the concept 'heart' are presented in Table 7.26.

TABLE 7.26 Lexical similarity set associated with the concept 'heart'

Language	Alignment
Alorese-Alor Besar	----- k u b - a ŋ -
Alorese-Alor Kecil	----- k u b - a ŋ -
Alorese-Bana	----- k u b - a ŋ -
Alorese-Beang Onong	----- k u b - a ŋ -
Alorese-Buaya	----- u b - a ŋ -
Alorese-Dulolong	----- k u b - a ŋ -
Alorese-Kayang	----- k u b - a ŋ -
Alorese-Munaseli	t a p k u b - a ŋ -
Alorese-Pandai	----- k u b - a ŋ -
Alorese-Ternate	----- k u b - a ŋ -
Alorese-Wailawar	----- u b - a ŋ -
Blagar-Bakalang	----- k u b - a ŋ -
Blagar-Bama	----- k u b - a ŋ -
Blagar-Kulijahi	----- k u b - a ŋ -
Blagar-Nule	----- k u b - a ŋ -
Blagar-Tuntuli	----- k u b - a ŋ -
Blagar-Warsalelang	----- k u b - a ŋ -
Kui	----- k u b l a - i
Klon	----- k u b
Reta-Pura	----- k u m b a
Blagar-Pura	----- k u b - a ŋ -
Wersing-Maritaing	- u - k a b - a ŋ -
Wersing-Taramana	g e u k a b - a ŋ -

For the concept 'heart', Alorese varieties innovated the form (*tapo/tapo*) *kubaŋ*. The part *tapo/tapo* means 'coconut' and is inherited (< Lamaholot-Kedang #tapu, see Samely 1991; Sulistyono 2022: 242), while the part *kubaŋ* is borrowed from AP languages. The form for 'heart' in AP languages is often given with a possessive prefix (*ge-* in Wersing, *ta-/eta-* in Klon and Reta)).

Based on the AP cognates presented in the table, a tentative PAP form *kVbaŋ 'heart' may be reconstructed. The PAP initial *k- is regularly retained as *k* in all

languages. Even though the PAP intervocalic *-b- is expected to be reflected as *p* in Wersing, a similar retention of intervocalic *-b- happens also in PAP *-lebur > Wersing *jebur* ‘tongue’ (Holton et al. 2012: 115).

The Alorese form (*tapo/tapo*) *kubay* ‘heart’ is probably a loanword from Blagar, as this language has the form that is most similar to the Alorese form. As for the addition of the (*tapo/tapo*) ‘coconut’ part, a possibility may be that the Alorese have re-analyzed the first person plural inclusive or reciprocal prefix *tV-*, which is often attached to body parts, as the first syllable of the word *tapo* ‘coconut’, and hence have added this word to the concept for ‘heart’.

Robinson (2015: 24) proposed the opposite pattern, namely that this is an Alorese loanword into Blagar and Wersing. This proposal was based on the similar form *taʔ kubay* ‘heart’ found in Kedang. However, the collection of more AP forms, and the internal diversity among the AP languages, suggest that the form *kubay* is likely of AP origin, while the Kedang word *taʔ kubay* is a loanword from Alorese, or an AP loanword into Kedang via Alorese.

3.5.7 ‘To breathe’

The Alorese and AP forms for the concept ‘to breathe’ are presented in Table 7.27 (repeated from Table 7.2).

TABLE 7.27 Lexical similarity set associated with the concept ‘to breathe’

Language	Alignment
Alorese-Munaseli	h ɔ - p a ŋ
Blagar-Bama	s o - p a ŋ
Blagar-Kulijai	h o - p a ŋ
Blagar-Nule	h o - p a ŋ
Blagar-Pura	h o - p a ŋ
Deing	- o - p a ŋ
Kaera	s u ʔ p a ŋ
Western Pantar-Tubbe	h o - p a ŋ
Reta-Pura	h o : - p a ŋ
Reta-Ternate	h u - p a ŋ

Alorese-Munaseli has innovated the word *hopang* ‘to breathe’ which is similar to forms attested in several AP languages, such as Blagar-Kulijahi, Nule,

Pura *hopang*, Western Pantar-Tubbe *hopang*, Kaera *supang*. Since, the AP forms are related and follow semi-regular sound changes (PAP initial *s > Kaera s, Blagar h, see Holton & Robinson 2017: 56), we consider this a loanword from AP languages into Alorese-Munaseli, with the most likely donor being a Blagar variety.

3.5.8 'Small'

The Alorese and AP forms for the concept 'small' are presented in Table 7.28.

TABLE 7.28 Lexical similarity set associated with the concept 'small'

Language	Alignment
Alorese-Alor Besar	k a e
Alorese-Alor Kecil	k a e
Alorese-Dulolong	k a e
Alorese-Ternate	k a e
Alorese-Buaya	k a e
Hamap	k a ʔ i - -
Kabola	k a ʔ a - i
Adang	k a ʔ a - i
Kaera	k i k i -
Blagar	k i k i -

The Alorese varieties on Pantar use three different inherited forms for the concept 'small'. There are two forms of PMP origin: *anay* (< PMP *anak) 'small' and *kari* (< PMP *kædi 'small in size'), and one form which can only be traced back to PWL *kesi/*kisu 'small' > *kihu* 'small' (Sulistyono 2022: 264). The Alorese varieties on the Alor Peninsula and in the Strait, however, have innovated a new form *kae* which suggests an external source. We do not group *kae* 'small' together with *kari* 'small' because all the Peninsula and Strait varieties consistently use the form *kae*, in doing so they differ from the conservative Pandai variety which retains *kari* < PMP *kædi 'small in size'.

The innovative form *kae* 'small' may have been borrowed from Adang *kaʔai* 'small', with the loss of medial glottal stop, which Alorese varieties lack. Several AP languages have similar words. The change of *k into intervocalic -ʔ- in Adang is regular (Holton et al. 2012: 94). However, the change of -in to -ai in Adang and Kabola remains unexplained.

Robinson (2015: 23) holds a different view on *kae* ‘small’, which she considers a word of Austronesian origin due to the similarity of Alorese *kae* with Kedang *keke* and Tetun *kiʔik* ‘small’.

The relationship between Alorese *kae* and Kedang *keke* (and Tetun *kiʔik*) is weak, and it is more likely that Alorese varieties borrowed *kae* from Adang *kaʔai* ‘small’. About the origin of Kaera *kiki*, Blagar *kiki*, and Adang *kaʔai* ‘small’, we agree that the origin of the AP forms may ultimately be from an Austronesian language spoken in the area before the arrival of the Alorese.

3.5.9 ‘To close’

The Alorese and AP forms for the concept ‘to close’ are presented in Table 7.29.

TABLE 7.29 Lexical similarity set associated with the concept ‘to close’

Language	Alignment
Alorese-Alor Kecil	-----tera-
Alorese-Bana	-----teraʔ
Alorese-Baranusa	-----tera-
Alorese-Beang Onong	-----tera-
Alorese-Alor Besar	-----tera-
Alorese-Dulolong	-----tera-
Alorese-Helangdohi	-----teraʔ
Alorese-Kayang	-----taraʔ
Alorese-Munaseli	-----teraʔ
Alorese-Pandai	-----tera-
Alorese-Ternate	-----fera-
Alorese-Wailawar	-----tera-
Kula-Lantoka	-----tira-
Deing	-----tiar
Kaera	wanteriŋ
Tubbe	-----tiarIŋ
Sawila	-li'tira
Reta-Pura	u--tiali
Adang-Lawahing	watεε
Adang-Otvai	-----u'tel
Blagar-Bakalang	venteriŋ
Blagar-Bama	venteriŋ
Blagar-Nule	venteriŋ

TABLE 7.29 Lexical similarity set associated with the concept 'to close' (*cont.*)

Language	Alignment
Blagar-Tuntuli	v e n t e r i ŋ
Reta-Ternate	- - u t i e l i
Wersing-Maritaing	l e t e r
Wersing-Taramana	l e t e r
Klon-Hopter	? u ' t e : r
Kiramang	- u t e r
Kui-Labaing	- u t e r i
Kabola-Monbang	w h u ' t e l e
Proto Alor-Pantar	- t i a r i n

For the concept 'to close', Alorese varieties innovated the form *tera(?)* 'to close' which is different from the form #letu? found in Lamholot varieties (see Fricke, this volume, Table 5.9). It is unclear why Alorese-Ternate has *fera* with initial *f*. AP languages display similar forms, all reflexes of the PAP form *-tari(n) (see Holton & Robinson 2017: 78). In many AP languages the root is preceded by an applicative prefix or by another verb: in Kaera *wan* is a verb which means 'be; exist' and occurs in serial verb constructions with various functions (Klamer 2014: 137). Considering that in AP languages the form is inherited, the form *tera(?)* in Alorese looks like an AP loanword, whereby Alorese varieties have borrowed the root *teri* 'to close' from either Kui, Kaera or most likely Blagar. The change of the ultimate vowel from *i* to *a* in loans is also attested in other Blagar loanwords, such as Alorese *kalaki* 'angry' from Blagar *kilikil*, Alorese *reha* 'monitor lizard' from Blagar *rihi*, and Alorese *lakuk* 'to fold' from Blagar *liku*.

3.5.10 'To hide'

The Alorese and AP forms for the concept 'to hide' are presented in Table 7.30.

TABLE 7.30 Lexical similarity set associated with the concept 'to hide'

Language	Alignment
Alorese-Alor Besar	d a - f u - -
Alorese-Bana	d ə - w u - -

TABLE 7.30 Lexical similarity set associated with the concept 'to hide' (*cont.*)

Language	Alignment
Alorese-Beang Onong	d a - w: u - -
Alorese-Beang Onong	d a - - u - -
Alorese-Buaya	d a - f: u - -
Alorese-Dulolong	d a - f u - -
Alorese-Dulolong	d a - - u - -
Alorese-Helangdohi	d ə - w u k -
Alorese-Kayang	d a - w: u - -
Alorese-Munaseli	d ə - w u k -
Alorese-Pandai	d a - w u - -
Alorese-Ternate	d a - f: - u
Alorese-Wailawar	d e - w u - -
Abui-Takalelang	t a - b u - -
Adang	t a w u n i ŋ
Kabola	t ə w u n i
Reta-Pura	t a β u n i ŋ
Western Pantar	- - - - u n n i ŋ

For the concept 'to hide' all Alor varieties display the form *dawu/ dəwuk* (in Pantar) or *dafu* (on the Alor Peninsula and in the Strait) that is not attested in the near-by Flores-Lembata languages. The change of *v* into *f* in Alor Peninsula and in the Strait varieties is regular (Sulistyono 2022: 214). This form is an innovation, possibly an old one, as it is found in all Alor varieties. The source are AP languages, which present forms that are similar to the Alor varieties. According to Robinson (2015: 25), the AP forms were borrowed from Malay (*bunyi* 'hide'), or from another Austronesian language going back to PMP *buni 'to hide'.

Some of the AP languages attached the reciprocal prefix *tV-* to the root and obtained forms like *tabuniŋ* (Reta) or *təwuni* (Kabola). Since the Alor forms *dawu/ dəwuk* are more similar to the AP forms (with the *tV-* prefix) than they are to PMP *buni, we conclude that Alor borrowed this form from AP languages, rather than inheriting it from PMP.

3.5.11 'Dirty'

The Alor and AP forms for the concept 'dirty' are presented in Table 7.31.

TABLE 7.31 Lexical similarity set associated with the concept 'dirty'

Language	Alignment
Alorese-Alor Besar	k a l i t a -
Alorese-Alor Kecil	k - l i t a -
Alorese-Bana	k - l i t a ?
Alorese-Baranusa	k - l i t a k
Alorese-Beang Onong	k a l i t a -
Alorese-Buaya	k a l i t a -
Alorese-Dulolong	k a l i t a -
Alorese-Helangdohi	k - l i t a ?
Alorese-Kayang	k - l i t a ?
Alorese-Munaseli	k - l i t a ?
Alorese-Pandai	k - l i t a -
Alorese-Ternate	k a l i t a -
Alorese-Wailawar	k - l i t a ?
Blagar-Bakalang	k - l i t a k
Blagar-Kulijahi	k ə l i t a h
Blagar-Nule	k ə r i t a k
Blagar-Pura	k a r i t a -
Reta	k a r i t a -
Teiwa	k - l i t a ?

For the concept 'dirty', all Alorese varieties innovated a form like #k(a)lita(?/k), different from the remaining Flores-Lembata languages that use a form reconstructable to PWL *mila 'dirty' (Sulistyono 2022: 245, 401). Klamer (2011: 105) lists this form among the Alorese loanwords from AP languages. Robinson (2015: 24), on the contrary, assumes this to be an Alorese loanword into AP languages, because it has a potential cognate in nearby Austronesian languages: Alorese *kalita* (cf., Lamaholot (Ile Ape) *prita*).⁶ We agree with Klamer and consider this an AP loanword into Alorese varieties for two reasons: (i) the AP forms share regular sound changes, (ii) the relationship between Alorese *kalita* and Lamaholot (Lamatuka) *prita* is weak.

⁶ According to LexiRumah 3.0.0, which reports data from Keraf 1978, it is Lamaholot Lamatuka which has *prita* for 'dirty', and not Ile Ape, which has *milan* 'dirty'.

The reflexes of AP forms show regular sound correspondences. In Blagar medial *l* corresponds to Reta medial *r* as seen in several other words, such as Blagar *bulaŋ* ~ Reta *buran* ‘sky’ and Blagar *bulit* ~ Reta *kaburit* ‘arrow’. A similar form *kilaʔe* ‘dirty’ is attested in Fataluku (a Papuan/Timor-Alor-Pantar language spoken in east Timor), which strengthens the proposal that this set is of AP origin. In addition to that, a similar cognate set with different semantic meaning, namely ‘old; elderly (people)’ is attested across AP languages, namely Abui *kalieta/kaleita*, Kafoa *kalta*, Kiramang *kaleta*, and Kui *kakaleta* ‘old; elderly (people)’. A semantic change might have occurred within the AP languages from ‘old; elderly’ to ‘dirty’. Finally, a comparison showed no correspondences between Alorese initial *k* and Lamatuka *p*. Given this evidence, we consider this a loan in Alorese from AP languages.

3.6 *Distribution of Loanwords*

Not all 28 AP loanwords occur in all 13 Alorese varieties. Some loanwords occur in all Alorese varieties, while others have a more limited geographical spread. Table 7.32 presents the distribution of loanwords in five groups: all Alorese varieties, only the varieties in northeast Pantar, only northeast Pantar and Alor Peninsula, only in the Alor Peninsula and in the Strait, and finally only in Pantar. The distribution of loanwords is informative about the relative age of loanwords, because loanwords attested in all Alorese varieties as regularly inherited forms were borrowed very early on before Alorese spread on the coastal areas of Alor and Pantar. The second group of loanwords are also possibly quite old, as those are found in northeast Pantar varieties, the area that is considered to be the homeland of the Alorese (see Sulistyono 2022).

TABLE 7.32 Distribution of loanwords based on geographic groups

Geographic groups	Concepts
All Alorese varieties	‘heart’, ‘ten’, ‘younger sibling’, ‘angry’, ‘taro’, ‘to close’, ‘to hide’, ‘dolphin’, ‘dirty’
Northeast Pantar	‘rattan’, ‘garden’, ‘digging stick’, ‘fish trap’, ‘coral rock’, ‘to breathe’, ‘to fold’, ‘to pull’, ‘to pray’
Northeast Pantar and Alor Peninsula	‘bed’, ‘gravel’, ‘adultery’
Alor Peninsula and Strait	‘monitor lizard’, ‘small’, ‘to bury’, ‘mud’
Pantar	‘road’, ‘to wash’
Alor Peninsula	‘root’

The semantic fields add a perspective on the type of contact. There is a difference between the early AP loans and the more recent loans attested in, for instance, the Alor Peninsula varieties. On the one hand, the early loans contain more basic vocabulary, such as numerals ('tens'), a kinship term ('younger sibling'), emotions ('angry') and body parts ('heart'). On the other hand, the more recent loanwords mainly concern nouns, particularly relating to the physical world, such as 'mud', 'monitor lizard', and 'root'.

Not surprisingly, since northeast Pantar is likely the homeland of the Alorese, the Alorese varieties more prone to borrowing are Alorese-Munaseli with 19 loanwords, and Alorese-Pandai with 15 loanwords on northeast Pantar, followed by the Alorese varieties on the Alor Peninsula (Alor Kecil, Dulolong and Alor Besar) (see Table 7.33), which are the second oldest group after the varieties of Munaseli and Pandai.

TABLE 7.33 Alorese varieties with their number of loanwords

Variety	Number of loanwords	Concepts
Munaseli	19	Adultery, angry, bed, coral rock, dirty, dolphin, garden, gravel, heart, rattan, road, small, ten, to breathe, to close, to hide, to pray, to wash, younger sibling
Pandai	15	Angry, bed, dirty, dolphin, fish trap, heart, rattan, small, ten, to bury, to close, to hide, to pull, to wash, younger sibling
Alor Kecil	13	Angry, dirty, dolphin, gravel, heart, monitor lizard, mud, root, taro, ten, to bury, to close, to hide
Dulolong	13	Angry, dirty, dolphin, gravel, heart, monitor lizard, mud, root, taro, ten, to bury, to close, to hide
Alor Besar	13	Angry, adultery, bed, dirty, dolphin, gravel, heart, monitor lizard, taro, ten, to bury, to close, to hide
Ternate	9	Angry, dirty, dolphin, heart, mud, ten, to bury, to close, to hide
Buaya	8	Angry, dirty, dolphin, heart, ten, to bury, to hide, younger sibling
Bana	9	digging stick, dirty, dolphin, heart, small, ten, to close, to hide, younger sibling
Wailawar	8	Angry, dirty, dolphin, heart, small, ten, to close, to hide
Baranusa	8	Angry, dirty, road, taro, ten, to close, to hide, to wash
Helangdohi	7	Digging stick, dirty, dolphin, small, ten, to close, to hide

TABLE 7.33 Alorese varieties with their number of loanwords (*cont.*)

Variety	Number of loanwords	Concepts
Kayang	6	Angry, dirty, heart, ten, to close, to hide
Beang Onong	6	Dirty, heart, road, ten, to close, to hide

The varieties with the smallest number of loanwords are the most recent ones, such as Beang Onong, which was established in the early 1960's, and Bana and Wailawar established in 1966 and 1996 respectively (see §1).

As for the main donor language(s), the AP donor languages are mainly languages spoken around the Alor-Pantar Strait. Blagar had an important role as donor language. In fact, out of 28 loanwords, 20 are likely to come from Blagar (or at least have Blagar among the possible donors): 'fish trap', 'bed', 'to pull', 'to fold', 'adultery', 'digging stick', 'garden', 'rattan', 'taro', 'mud', 'coral rock', 'gravel', 'monitor lizard', 'dolphin', 'younger sibling', 'to close', 'to breathe', 'to bury', 'heart', and 'angry'. That Blagar is the dominant donor comes as no surprise, since Alorese and Blagar have a close, historical relationship. Both communities are bound in a century-old sociopolitical alliance, called *Galiyao Watang Lema* (see Sulistyono 2022: 15–16).

Other AP languages around the Alor-Pantar Strait that have also contributed AP loanwords to Alorese are Adang, Klou, and Kaera. The contribution of these languages varies according the Alorese subgroup in question. Adang is more likely to be the donor of loanwords found in the Alor Peninsula varieties, while Klou is more likely the donor for loanwords found both in northeast Pantar and Alor Peninsula varieties. Kaera probably had one of the earliest contacts with Alorese, because almost all loans from Kaera belong to the first group. Western Pantar and Deing are donors only for Alorese varieties spoken on Pantar.

4 Discussion

In the previous section, we have presented evidence for lexical borrowing from the AP languages into Alorese. After a close inspection, applying automatic lexical similarity detection, and subsequently a qualitative fine-grained analysis (see §2), we have detected 28 loanword events between Alorese and AP languages on a list of 596 items. The percentage of AP loanwords in Alorese is, thus, approximately 4.7%, confirming previous results of Klammer (2011) and Robin-

son (2015), which were based on smaller datasets. This result shows, on the one hand, that the percentage of AP loanwords in Alorese is indeed small, and on the other hand, that conducting loanword analysis on a Swadesh list, like in Klamer (2011), is likely to give a representative figure of the number of loanwords in a language. Having said this, the innovative use of automatic lexical similarity detection used in the present chapter looks promising and it deserves to be tested further in other studies, because it allows the screening of large datasets and a comparison across language families in a short amount of time. An obvious issue is that the first screening by distribution patterns can turn up false positives (forms marked as related which turn out not to be). These can be filtered out, but a small chance remains that few additional, actual loanwords are not found because they have a spurious similarity to e.g., other Austronesian forms, which makes them not pattern as loanwords. There might be other caveats that we are not aware of, which future studies using the same methodology may unravel.

The limited lexical influence from AP languages into Alorese is not so peculiar if seen in a broader geographical perspective. Similar findings are reported in two contributions of the present volume: Schapper and Huber (this volume), who focus on lexical borrowing from Papuan languages into Austronesian languages of Timor; and Klamer (this volume), who presents evidence for the opposite pattern, namely ancient Austronesian words attested in the lexicon of the TAP languages. However, the way these two studies compiled their dataset was very different from ours. An interesting result, that is shared by Schapper and Huber's, Klamer's and our contribution is that, despite the length of contact, the number of loanwords is relatively small: a dozen loanwords in Schapper and Huber, 14 ancient loanwords in Klamer, and 28 loanwords in our study. For Schapper and Huber (this volume), one possible explanation for the small number of loanwords is the lack of data from the Papuan languages of Timor, especially in the semantic field of plants and animals, which is a domain that attracts a considerable number of Papuan borrowings. According to Klamer (this volume), the limited and scattered lexical borrowing from Malayo-Polynesian languages into TAP languages points to a contact scenario involving relatively superficial contact in few socio-cultural domains such as trade or marriage, which does not require a community to be bilingual.

We now turn to the discussion of the 28 AP loanwords, that, despite the small number, can still be regarded as significant and informative about the type of influence that AP languages had on Alorese. AP lexical influence on Alorese is reflected by loans involving agriculture and vegetation (digging stick, garden, rattan, root, taro), the physical world (coral rock, mud, gravel), animals (dol-

phin, monitor lizard), and basic actions and technology (fish trap, bed, to fold, to pull, to wash). So, it seems that the AP languages mainly contributed with terms referring to the environment, or referring to tools and actions related to the environment. This is confirmed by the study of Schapper and Huber (this volume), who show that Papuan lexical influence on the Austronesian languages of Timor is mostly found in the domains of plants and animals. A similar result is also presented in Edwards (this volume) who found that possible loanwords or innovation in the regional and west Timor strata of the Rote-Meto lexicon are robustly attested in semantic spheres very prone to borrowing, such as *Tools* and *Vegetation*. This is especially interesting, if seen in contrast with the Austronesian lexical influence on the TAP languages (Klamer, this volume), which is reflected by loans involving textile technology (needle, to weave, to sew); societal structures (slave, king/ruler), subsistence and trade (salt, seed, maize, skin), and marriage (bride price).

This limited lexical influence does rule out an adult language-shift scenario in the Alor archipelago, because this is usually accompanied by the retention of (specialist) vocabulary from the heritage language (Ross 2013; see also Klamer, this volume).⁷ The wholesale adoption of a good amount of lexical items is very frequent when there is an unequal relation between the languages, such that one community shifts to another language, and in doing so, it retains parts of the L1's vocabulary,⁸ or one community adopts many words from a prestigious L2 (Muysken 2013). Neither of these scenarios applies in the case of Alor and Pantar. In the Alor archipelago, bilingualism involving Alorese and AP languages was long and stable, as is today, and never ended in a shift.

Evidence from contact-induced grammatical changes in Alorese show that Alorese was initially spoken in bilingual communities characterized by symmetric bilingualism, dense social networks, and low normativity, with many bilingual children who introduced new grammatical constructions in Alorese on the model of their AP languages (Moro 2018, Moro & Fricke 2020). After this period, which was relatively short, Alorese communities became larger, networks were looser, the language started to enjoy more prestige and became a lingua franca in the area (see §1). Consequently, Alorese was learned as an L2 by many adult AP speakers, and the outcome of this type of contact was severe simplification of morphology (Klamer 2012, 2020; Moro 2019). These acquisitional and socialisation patterns can still be observed today, as local people on Alor report that many Adang speakers can speak Alorese, but that

7 According to Ross (2013: 30), adult language shift appears to have been rare in Melanesia.

8 This shift scenario is hypothesized for Rote-Meto (Edwards, this volume).

Alorese people cannot speak Adang. Therefore, the asymmetric bilingual patterns that have started sometime in the past continue to the present day (see Moro 2021).

Two factors, thus, explain the relatively small amount of AP loanwords in Alorese. First, as discussed above, the bilingualism situation that led to grammatical borrowing did not last long enough, and when Alorese became more prestigious, the pattern became asymmetric. The fact that bilingualism in the AP language(s) was not reciprocated by the Alorese prevented the adoption of AP words in the Alorese language. Second, it is likely that in the exogamous Alorese community, the spouses came from different AP communities and thus spoke different AP languages, as we can still observe in Munaseli today. In a fieldwork trip conducted in 2016, Francesca Moro recorded 12 AP speakers who had married an Alorese spouse and had moved into the Alorese Munaseli community: they had six different L1s: Kroku (five speakers), Blagar (three speakers), Teiwa (one speaker), Sar (one speaker), Kaera (one speaker), Klamu (one speaker). So, a possible answer to the question “why did the Papuan mothers not introduce more of their native Papuan *lexicon* into the Alorese they used?” (cf. Klamer 2012: 104), is that the many different AP languages involved might have prevented heavy lexical borrowing from one specific AP language. A similar outcome is found in creoles, where the presence of several L1s interfering with each other prevents transfer from a single L1 (cf. Muysken 2013: 717). We can conclude that the bilingualism had more influence on the grammar of Alorese than on its lexicon, as the grammar usually falls below the threshold of consciousness, and the grammatical changes were either shared by almost all the L1s (presence of a plural word, converged give-constructions, see Moro 2018 and Moro & Fricke 2020), or they were simplification process independent of the L1s (loss of inflectional morphology, see Moro 2019).

Finally, Schapper and Huber (this volume) point out that “it is important not to exclude a lexeme as a possible loan candidate just because it has a known Austronesian etymology”. We agree with this observation, as we also report cases, such as ‘to hide’ or ‘dolphin’, where lexemes coming from an Austronesian source were borrowed into AP languages, and then from there borrowed again into Alorese.

To conclude, we inspected the whole available lexicon (~600 words) of 13 Alorese varieties and found that, despite the length of contact between Alorese and AP speakers, the presence of AP loanwords is ‘only’ 4.7%. The bilingualism scenario found in Alorese-AP communities had more influence on the grammar of Alorese than on its lexicon. This limited lexical influence is accounted for by the asymmetric bilingualism patterns and by the presence of several L1s interfering with each other. Yet, the AP loanwords can tell us that contact

between the Alorese and AP speakers revolved around agriculture and vegetation, the physical world, and basic actions and technology, and that Blagar had an important role as donor language, probably due to its position on Pantar and in the Strait.

Acknowledgements

We would like to thank the audience of the APLL11 conference, as well as Marian Klamer and George Saad for their valuable feedback on previous versions of this chapter. The collection and the analysis of the data were supported by the NWO VICI Project *Reconstructing the past through languages of the present: The Lesser Sunda Islands* led by Prof. Marian Klamer (project number 277-70-012).

References

- Anonymous (1914). De eilanden Alor en Pantar, Residentie Timor en Onderhoorigheden. *Tijdschrift van het Koninklijk Nederlandsch Aardrijkskundig Genootschap* 31, 70–102.
- Blust, Robert A. and Stephen Trussel (2020). *The Austronesian Comparative Dictionary, Web edition*. Available at: <http://www.trussel2.com/ACD>.
- Doyle, Matthew (2010). *Internal divisions of the Flores-Lembata subgroup of Central Malayo-Polynesian*. Master thesis, Leiden University.
- DuBois, Cora (1944). *The people of Alor: A social-psychological study of an East Indian Island*. Harvard University Press.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (2019). *Ethnologue: Languages of the world* (22nd ed.). SIL International. <http://www.ethnologue.com>
- Edwards, Owen (this volume). Substrate Transference in Rote-Meto.
- Elias, Alexander (2017). Subgrouping the Flores-Lembata languages using historical glottometry. Presentation on the *9th International Austronesian and Papuan languages and linguistics conference* (APLL-9) 21–23 June 2017, Paris.
- Fernandez, Inyo Yos (1996). *Relasi historis kekerabatan Bahasa Flores*. Penerbit Nusa Indah.
- Fricke, Hanna (2019). *Traces of language contact: The Flores-Lembata languages in eastern Indonesia*. Doctoral dissertation, Leiden University.
- Holton, Gary, Marian Klamer, Frantisek Kratochvíl, Laura C. Robinson, and Antoinette Schapper (2012). The historical relations of the Papuan languages of Alor and Pantar. *Oceanic linguistics* 51(1): 86–122.

- Holton, Gary and Laura C. Robinson (2017). The internal history of the Alor-Pantar language family. In Klamer, Marian (ed.), *The Alor-Pantar Languages: History and Typology*, pp. 49–91. Berlin: Language Science Press.
- Kaiping, Gereon and Marian Klamer (2019). Subgroupings of the Timor-Alor-Pantar languages using systematic Bayesian inference. *SocArXiv*: 1–48.
- Kaiping, Gereon, Owen Edwards, and Marian Klamer (eds.) (2019). *LexiRumah 3.0.1*. Leiden: Leiden University Centre for Linguistics. Available at: <http://www.model-ling.eu/lexirumah/>.
- Klamer, Marian (2010). *A Grammar of Teiwa*. Berlin: Mouton De Gruyter.
- Klamer, Marian (2011). *A Short Grammar of Alorese (Austronesian)*. München: Lincom Europa.
- Klamer, Marian (2012). Papuan-Austronesian language contact: Alorese from an a real perspective. In Nicholas, Evans and Marian Klamer (eds.), *Melanesian languages on the Edge of Asia: Challenges for the 21st Century*, pp. 72–108. University of Hawai'i Press.
- Klamer, Marian (2014). Kaera. In Schapper, Antoinette (ed.), *The Papuan Languages of Timor, Alor and Pantar: Volume 1*, pp. 97–146. Berlin: Mouton De Gruyter.
- Klamer, Marian (2015). *Language as a time machine*. Inaugural lecture. <https://scholarlypublications.universiteitleiden.nl/handle/1887/32306>
- Klamer, Marian (2017). The Alor-Pantar languages: Linguistic context, history and typology. In Klamer, Marian (ed.), *The Alor-Pantar languages: History and typology*, pp. 1–49. Berlin: Language Science Press.
- Klamer, Marian (2020). From Lamaholot to Alorese: Morphological loss in adult language contact. In Gil, David and Antoinette Schapper (eds.), *Austronesian Undressed: How and why languages become isolating*, pp. 339–368. Amsterdam: John Benjamins.
- Klamer, Marian (this volume). Traces of pre-modern contacts between Timor-Alor-Pantar and Austronesian speakers.
- Kratochvíl, Frantisek (2014). Sawila. In Schapper, Antoinette (ed.), *The Papuan Languages of Timor, Alor and Pantar: Volume 1*, pp. 351–438. Berlin: Mouton De Gruyter.
- List, Johann-Mattis (2012). sCA: Phonetic Alignment based on sound classes. *New Directions in Logic, Language and Computation*, pp. 32–51. Springer.
- List, Johann-Mattis, Simon J. Greenhill, Tiago Tresoldi and Robert Forkel (2019). *LingPy. A Python Library for Quantitative Tasks in Historical Linguistics*. Version 2.6.5. Zenodo. <http://doi.org/10.5281/zenodo.3482794>.
- List, Johann-Mattis and Robert Forkel (2021). Automated identification of borrowings in multilingual wordlists [version 2]. *Open Research Europe* 2021, 1:79.
- Moro, Francesca R. (2018). The plural word 'hire' in Alorese: contact-induced change from neighboring Alor-Pantar languages. *Oceanic linguistics* 57(1): 177–198.

- Moro, Francesca R. (2019). Loss of morphology in Alorese (Austronesian): Simplification in adult language contact. *Journal of Language Contact* 12(1): 378–403.
- Moro, Francesca R. (2021). Multilingualism in eastern Indonesia: Linguistic evidence of a shift from symmetric to asymmetric multilingualism. *International Journal of Bilingualism* 25(4):1102–1119.
- Moro, Francesca R. and Hanna Fricke (2020). Contact-induced change in Alorese give-constructions. *Oceanic Linguistics* 59(1/2): 116–147.
- Muysken, Pieter. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(4): 709–730.
- Rama, Taraka, Johann-Mattis List, Johannes Wahle and Gerhard Jäger (2018). Are Automatic Methods for Cognate Detection Good Enough for Phylogenetic Reconstruction in Historical Linguistics? *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 393–400. New Orleans: Association for Computational Linguistics.
- Robinson, Laura C. 2015. The Alor-Pantar (Papuan) languages and Austronesian contact in east Nusantara. In Ross, Malcom and I Wayan Arka (eds.), *The Asia-Pacific Linguistics: Language Change in Austronesian Languages*, pp. 19–33. Canberra, Australia: Pacific Linguistics.
- Ross, Malcom (2013). Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact* 6(1):5–47.
- Rosvall, Martin, Daniel Axelsson, and Carl T. Bergstrom (2009). The map equation. *The European Physical Journal Special Topics* 178(1): 13–23.
- Samely, Ursula (1991). Wordlist on Kedang-Leuwayang. In Kaiping, Gereon, Owen Edwards, and Marian Klamer (eds.), *LexiRumah 3.0.1*. Leiden: Leiden University Centre for Linguistics. Available at: <https://lexirumah.model-ling.eu/lexirumah/>.
- Schapper, Antoinette and Marian Klamer (2017). Numeral system in the Alor-Pantar languages. In Klamer, Marian (ed.), *The Alor Pantar Languages: History and Typology*, pp. 277–329. Berlin: Language Science Press.
- Schapper, Antoinette, Juliette Huber, and Aone van Engelenhoven (2017). The relatedness of Timor-Kisar and Alor-Pantar languages: A preliminary demonstration. In Klamer, Marian (ed.), *The Alor-Pantar languages: History and typology*, pp. 91–146. Berlin: Language Science Press.
- Schapper, Antoinette and Juliette Huber (this volume). Entwined histories: the lexicons of Kawaimina and Maka languages.
- Steinhauer, Hein (2014). Blagar. In Schapper, Antoinette (ed.), *The Papuan Languages of Alor and Pantar: Volume 1*, pp. 148–218. Berlin: Mouton De Gruyter.
- Stokhof, Willem A. (1975). *Preliminary Notes on the Alor and Pantar Languages (East Indonesia)*. Canberra: Pacific Linguistics.
- Stokhof, Willem A. (1984). *Annotations to a text in the Abui language (Alor)*. *Bijdragen tot de taal-, land-en volkenkunde* 140(1): 106–162.

- Sulistyono, Yunus (2022). *A history of Alorese (Austronesian): Combining linguistic and oral history*. Doctoral dissertation, Leiden University.
- Tadmor, Uri, Martin Haspelmath, and Bradley Taylor. (2012). Borrowability and the notion of basic vocabulary. In Wichmann, Søren and Anthony P. Grant (eds.), *Quantitative approaches to linguistic diversity*, pp. 35–55. John Benjamins Publishing Company.
- Uffmann, Christian (2015). Loanword Adaptation. In Honeybone, Patrick and Joseph Salmons (eds.), *The Oxford Handbook of Historical Phonology*, pp. 645–667. Oxford University Press.
- Wellfelt, Emilie (2016). *Historyscapes in Alor: Approaching Indigenous Histories in Eastern Indonesia*. Doctoral dissertation, Linnaeus University.
- Winter-Froemel, Esme (2013). Formal variance and semantic changes in borrowing: Integrating semasiology and onomasiology. In Zenner, Eline and Gitte Kristiansen (eds.), *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations*, pp. 65–100. Berlin, Boston: De Gruyter Mouton.

Multilateral Lexical Transfer among Four Papuan Language Families: Border, Nimboran, Sentani, and Sko

Claudia Gerstner-Link

1 Introduction

The Papuan language families of Border, Nimboran, Sentani, and Sko cover a geographically contiguous area in the north of the island of New Guinea. The Border and Sko families are mainly located in the east in Papua New Guinea, while the Nimboran and Sentani families are located in the west in Indonesia (see Figure 8.1). It seems that this political split had consequences for language research in that, so far, these four families have not been brought together in unified research that may detect mutual influences among them. Doing this, the present article breaks new ground, and will lead to new insights about the peoples, their languages, their interaction, and their ‘nomadic’ impetus over centuries, which only recently came to a halt due to the centralised political government in both modern states. The selection of the four families is further motivated by the aim to set Kilmeri and the Border languages in their wider linguistic and geographical context; as the author of a grammar of Kilmeri (Gerstner-Link 2018) it is an objective of mine to anchor this language in a broader research context.

When dealing with language contact in the geographical area of the Border, Nimboran, Sentani, and Sko families one has to distinguish two layers: (i) contact among local vernacular languages of the same family and across families; (ii) contact between Austronesian and Papuan languages; (iii) contact between local languages and the modern *linguae francae* (Papuan) Malay and Dutch as well as Tok Pisin and English. Needless to say, there are numerous loanwords from these *linguae francae* into the indigenous languages under examination. For the present study, contact with Austronesian languages, Malay, Tok Pisin, Dutch, and English is beyond the focus.

The article starts by outlining the historical and geographical settings of the language families and the people. There are no written native sources; the scanty data we have about the peoples’ history rely on oral tradition. In a few grammars, these oral accounts are very briefly documented. Wordlists started

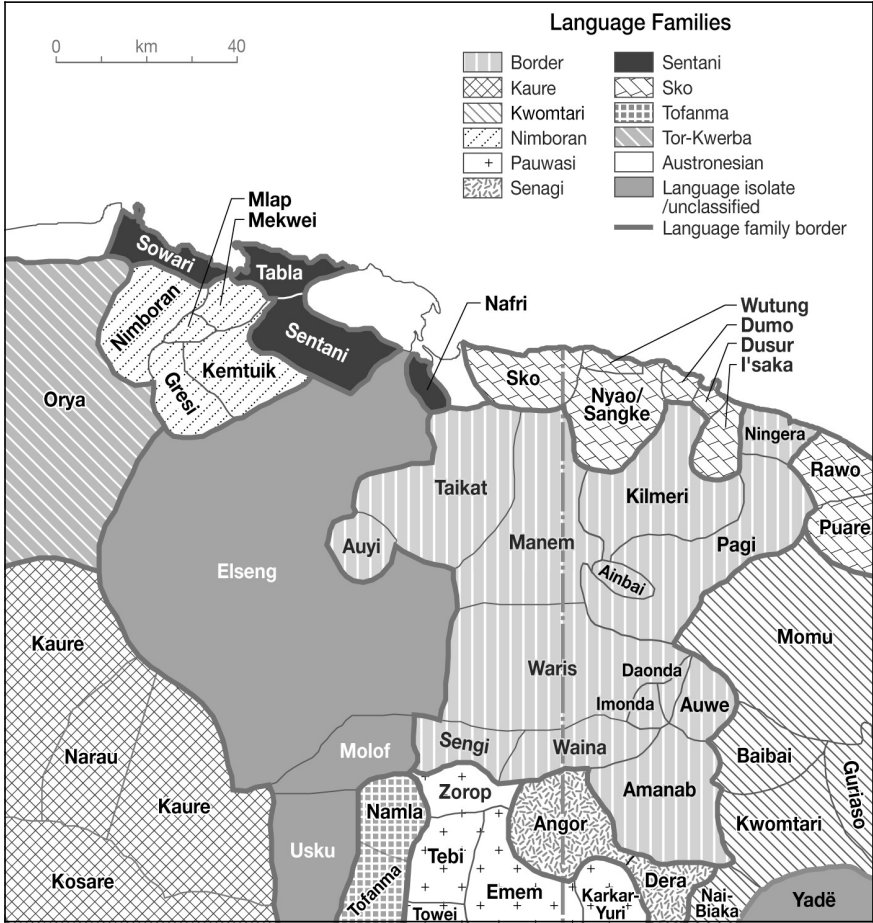


FIGURE 8.1 Language map

to be collected only in the 20th century. Section 3 reflects on this research situation and discusses some methodological considerations on which the vocabulary comparison and the recognition of loanwords are based. In the following three sections the Border languages are lexically compared with the Nimboran family, the Sentani family, and the Sko family. The putative transfers are listed and commented on one by one followed by a short summary concluding each of the three sections. These summaries provide information about number, word class, and semantics of transferred items, the directionality of transfers, the phonological integration into the recipient language, replacement or co-existence with an inherited word, and, if possible, about the relative age of the transfers. But these findings are not sufficient to propose concrete scenarios of contact in the sense of, say, Muysken's (2010:271–278) scenarios. The only case

TABLE 8.1 The Border family including Elseng^a

Border Family			
Bewani branch	Waris branch	Taikat branch	Elseng branch
Ainbai	Amanab (Minch 1992)	Auyi	Elseng (Menanti 2005)
Kilmeri (Gerstner-Link 2018)	Auwe [Simog]	Taikat (Smits & Voorhoeve 1994)	
Ningera	Daonda		
Pagi (Gerstner-Link 2000)	Imonda (Seiler 1985)		
	Manem		
	Sengi [Viid]		
	Waina [Sowanda]		
	Waris [Walsa] (Brown & Wai 1986)		

- a Elseng is claimed to be an isolate (Foley 2018:435–438). Based on the comparative method, there is good lexical and some paradigmatic evidence for its inclusion into the Border family (Gerstner-Link 2020, Ross 2005; Timothy Usher p.c.).

in which a certain scenario is quite probable is discussed in Section 7: it deals with *wanderwörter* whose spread was facilitated through extensive bird of paradise hunting in the area for trade outside New Guinea. Finally, Section 8 summarises the lexical transfers and reflects on their low number, which, however, is compensated to a small degree by a few patterns of structural convergence. The section ends with a discussion of putative migrations of the peoples, in particular the Kilmeri. At the same time, a hypothesis about the original homeland of the Border people and their languages is developed.

2 Historical and Geographical Settings

The Border languages (Table 8.1) cover a geographically contiguous area stretching from the Border Mountains and their foothills in the south to the valleys and plains north of the Bewani Mountains. The Bewani range is not inhabited. Nowadays, the people speaking Border languages live in three areas:

north of the Bewani Mountains in the Puwani-Pual river basin and on the northern coast east of Vanimo; south of the Bewani Mountains and north-east of the Border Mountains in swampy hills and small creek systems as well as in the *Wasengla valley* (Waris) that stretches south-east along the headwaters of the Bapi river; thirdly west of the Bewani watershed and in the Tami and Bewani valleys. The Sengi (Waris branch) live further south and west of the Border Mountains.

The literature provides evidence that several linguistic groups of the Border people have migrated to their current locations a number of generations ago. For the Imonda, Seiler (1985:1) states that “[t]he Imonda trace their history to an area [,] to the north-west”. Regarding the Waris people Brown (1990:8) says that their self-designation *Walsa* “seems to refer to them as the ones who successfully overcame the previous people to live in the area”. The area in question is the Wasengla valley, and the Waris speaking clans may have pushed the Umeda group of the Waina speaking people southwards in a less favourable location in the north-eastern foothills of the Border Mountains (Gell 1992:153–154). Another or additional scenario may be that the Waris expelled some clans that spoke languages of the Kwomtari family (see language map), whose descendants may now live in the hot and swampy lowlands to the east (Donohue and Crowther 2004:173). Regarding a group of the Amanab speaking people anthropologist Juillerat suggests “[that] the Border Mountains seem to have been populated, at least in part, from the west or northwest, and the cultures found there contrast sharply with those of the nearby plain.” (Juillerat 1996: xxi)¹ Finally, for the Kilmeri located north of the Bewani Mountains Gerstner-Link (2018:17–19) provides evidence that the people arrived at their current locations ten generations ago; the clan leader/s appropriated the land.

The Nimboran and Sentani families (Tables 8.2 and 8.3) we have fewer clues regarding their places of origin. According to their own oral tradition, the Nimboran came from the south to their current location: “Nimboran people say that their ancestors, along with those of the related ethnolinguistic groups of Kemtuik, Kwansu and Gresu, spread out into the Grimi River valley from a location named Singgi or *hngni* in the hills to the south. Today nearly all of the Nimboran people live to the north-west of the River Nembu.” (May 1997:3). Anceaux gathered his data on Nimboran between 1954–1957 in Jayapura (Hollandia) and during periodical visits to some Nimboran villages (1965:2–3). At this time, the

1 The “nearby plain cultures” belong to the vast cultural area of the Upper-Sepik and its tributaries whose western-most fringes they form (Craig 1980:2, 7).

TABLE 8.2 The Nimboran family

Nimboran family			
Nimboran branch	Kemtuik-Gresi-Mlap-Mekwei branch		
Nimboran (Anceaux 1965) (May 1997)	Mekwei branch	Mlap-Kemtuik-Gresi branch	
	Mekwei	Mlap branch	Kemtuik-Gresi branch
		Mlap [Kwansu]	Kemtuik (van der Wilden & van der Wilden 1975, 1976) (Smits & Voorhoeve 1994) Gresi (Smits & Voorhoeve 1994)

FOLEY 2018:446

Nimboran language was in full use. Unfortunately, Anceaux provides no clues about the history of Nimboran settlements.

Some groups of speakers of Sentani languages (Table 8.3) originate in a location that nowadays is populated by Border speakers. They trace their ancestors to the east. Chief Asareu tells that some ancestors originated from the earth, while others stem from Mount Fanim in the east. The settlement on the island of Osei in Lake Sentani was the first to be populated by migrants from the east. (Wirz 1934:257; 260) A Sentani myth says that a snake carrying a young man on its back swam across the Tami River towards the sea—the former Humboldt Bay—and finally reached the current location of Nafri (Table 8.3). The Tami and Bewani rivers flow through the current area of the Manem and Taikat people, who speak Border languages.

TABLE 8.3 The Sentani family

Sentani family	
Sowari branch	Tabla-Sentani-Nafri branch
Sowari	Tabla (Gregerson & Hartzler 1987)

TABLE 8.3 The Sentani family (*cont.*)

Sentani family
Sentani (Cowan 1965, Gregerson & Hartzler 1987) Nafri

FOLEY 2018:438

The Skou people themselves, as well as the other speakers of the Sko family languages, also look back at repeated movements of clans or groups of men. Donohue states that the speakers of Proto Macro-Sko originally lived along the middle Puwani-Pual River area (2004:5). This is exactly the area where nowadays the Kilmeri live, and Donohue conjectures that these people were displaced by the intrusion of speakers of the Bewani branch of the Border languages (Table 8.1). The expelled Macro-Sko speakers migrated to the west and to the east and spread along the coast. For the eastern-most Sko speakers, the Barupu, Corris provides a quite detailed description of their putative migration and later arrival at their present location near the Sissano lagoon. When the ancestors of the modern Barupu left the Puwani-Pual area, some of them may have headed east, reaching the lagoon from inland; others are said to have come along the coast (Corris 2005:3–8).

In sum, all these accounts provide evidence that the speakers of the four language families have a history of migration. According to oral tradition, Kilmeri clans migrated about 250 years ago. For the other groups, migration may have stretched over decades or even centuries and, at a time, comprised groups of clan size. See Section 8 for further discussion.

3 Method and Terminology

As a precondition for vocabulary comparison, we need reliable data sources that allow us to compare a sizable amount of the vocabularies of the languages concerned. This demand restricts the languages that can be thoroughly compared to those for which a lexicon and/or a grammar is available. For the Border languages, only Kilmeri, Waris, Imonda, and, to a lesser degree, Amanab fulfill this condition (see Table 8.1). For Taikat and Auyi, only (unsystematic) wordlists are published. Regarding the Nimboran languages, Anceaux's (1965) and May's (1997) grammars of Nimboran are good sources. The Sentani languages are lex-

TABLE 8.4 The Sko family

Sko family						
I'saka branch	Piore River-Serra Hills-Inner Sko branch					
I'saka (Donohue & San Roque 2004)	Piore River branch	Serra Hills branch	Inner Sko branch			
	Barupu (Corris 2005)	Womo	Skou branch	Eastern Sko branch		
	Ramo	Rawo	Skou (Donohue 2004; 2002)	Leitre branch	Wutung-Sangke-Dumo- Dusur branch	
	Sumo [Bouni] (Miller 2017)	Puare		Leitre	Wutung-Sangke branch	Dumo- Dusur branch
				Sangke Wutung (Marmion 2010)	Dumo Dusur (Ross 1980)	

FOLEY 2018:399; DONOHUE 2004:16; 18

ically represented by Cowan's grammar (1965) and supplemented by articles on Tabla and Sentani (Hartzler 1976; Gregerson and Hartzler 1987). Among the Sko family, good lexical sources are available for Skou, Wutung, Dumo, Dusur, I'saka, and Barupu (see Table 8.4).²

In reconstructing the contact scenario, I take the Kilmeri lexicon as a point of departure because (i) Kilmeri's documented lexicon is the most compre-

2 Throughout the article, I use the following notational conventions: The vocabulary items are presented in Standard IPA. In doing that, the orthography of the original sources is transcribed into IPA in accord to each author's spelling conventions. Morpheme boundaries are indicated by a hyphen. For composite lexemes I use the underscore to represent the boundaries between the parts. A consonant or vowel in round brackets represents an optional sound that is only realised in some languages of a family. Curly brackets indicate that a morpheme of a complex lexeme is not taken into account for comparison. A slash indicates lexeme variants within a language or a language family. The notation of tone in Skou and Wutung follows the conventions in Donohue (2004) and Marmion (2010).

hensive among the Border languages (Gerstner-Link 2021) and (ii) it is the language that the author knows best (Gerstner-Link 2018). The Kilmeri lexicon contains certain words that distinguish it from the other (well-documented) Border languages. Where do these vocabulary items come from when they are not inherited?³ Which lexical items of Kilmeri can be found in its neighbouring languages? This approach is restrictive in that it only allows for the discovery of a subset of mutual transfers or loans among the languages in question, namely those transfers that involve Kilmeri and the other Border languages. Transfers from, for instance, the Sko family to the Sentani family or vice versa, can more reliably be detected by researchers having first-hand knowledge of these families or single languages thereof. As we will see, lexical transfer among the above-mentioned languages and language families took place in multilateral directions: all families are both donors and recipients.

Turning to the question how to determine whether a lexical item is of foreign origin in a certain language I pursue the following path. If genetically related languages show cognate forms for a certain concept, then the lexeme in question is regarded as inherited. If a word is not attested in two branches of the same family but in only one, and it is also attested in another family, then I take it to be borrowed across the family borders. The fact that a word does not have an intra-family etymology is not an entirely conclusive sign of its loan status, since it might have been lost in the other branches of the family (Haspelmath 2009:44). However, without this working hypothesis there would not be any plausible reasoning to identify certain words as transfers or loanwords in the present context.

Regarding the Border family, no reconstruction has so far been done of a proto phoneme inventory accompanied by a (small) proto lexicon. For Waris and Kilmeri—representing two branches of the Border family—sound correspondences and cognates have been established by the author (see Appendix). Within the (putative) Bewani branch of the Border family, cognate sets for Kilmeri and Pagi have been uncovered (Gerstner-Link 2018:31–37). Based on these two sets of cognate pairs, I compiled a small triple set of cognate forms (see Appendix). These findings can count as a basis for inheritance within the Border languages. The sound changes involved could indicate the relative age of loans, insofar as these did or did not participate in a given change. For the Sentani family, the Proto Tabla-Sentani phonology has been reconstructed by Gregerson and Hartzler (1987); it serves as a basis for judgments about inheri-

3 In a few cases, Waris and Taikat reveal themselves as the recipient of foreign vocabulary.

ance in this family. Donohue (2002) describes structural phonological borrowing accompanied by the rearrangement of the phoneme systems of the Inner-Sko branch of the Sko family, whose phoneme inventory he reconstructed. This again allows us to recognise Inner-Sko inheritance. Isaka is an outlier genetically, but currently an immediate neighbour of Kilmeri. Shared and similar lexical items between these two languages are due to recent contact (Gerstner-Link 2018:45–47). By contrast, for the Nimboran family no comparative work is available. Due to this stage of research, the etymological background of the compared Nimboran words must remain a matter of informed guesses but not of proof.

The procedure I used to assemble semantically and phonologically similar forms across language families can be described as follows. The starting point is a shared inter-family concept in the lexicon. The next step is to compare segments and syllable structure of the assumed loan with its counterpart in the assumed donor language. For example, Kilmeri and Nimboran share the concept ‘old’, and we have Kilmeri *bepi* and Nimboran *bedí*. Although three segments of the words are identical, I don’t regard the forms as resemblant, because /p/ and /d/ in position 3 cannot be related. Both languages possess these phonemes in their inventories, and there is no reason that one of them should have been replaced by the other for phonemic adaptation. I regard the similarity as coincidental. By contrast, the concept ‘wallaby’ is realised as Kilmeri *emei* and Sentani proper *eme*. Here three segments are nearly identical in substance and order. Kilmeri could have taken over the form *eme* and have diphthongised the last vowel. Diphthongisation is a phonetic variation that can often be observed within Kilmeri when different speakers pronounce a word ending in /e/ or /o/; it can also be applied on words of foreign origin. See Section 5 below. Furthermore, if the phoneme inventories and/or the phonotactic rules of the languages in question differ, phonological adaptation has to be taken into account in order to establish segmental resemblance between forms.

A general problem with the languages concerned is the shortness of forms that are compared. This could be seen as causing a serious methodological weakness of the paper. Many forms I am dealing with are monosyllabic; sometimes they have only two segments. In this case there is the possibility of chance similarity. I can never exclude this possibility entirely, but I hope to present arguments that support the putative transfer. These arguments are based on word forms and their degree of similarity including phonological adaptation, on semantics including meaning shift (Blank 1997; Aikhenvald 2000), as well as on structural properties of the lexicon such as, in particular, co-existence of two terms for one concept. These terms may be nearly synonymous or the

new term may add a finer lexical distinction.⁴ Word forms consisting of only one segment are excluded as candidates for transfer. The concepts ‘father’ and ‘mother’ are also excluded, since they are frequently realised as nursery forms which nothing should be inferred from.

My terminology follows Matras (2009), Haspelmath and Tadmor (2009), and Haspelmath (2009). In their Loanword Typology project, Haspelmath and Tadmor use the following definition: “We define a loanword as a lexeme that has been transferred from one lect into another and is used as a word (rather than as an affix, for example) in the recipient language.” (2009:13) Essential in lexical transfer and borrowing are also the notions of donor vs. recipient language (Matras 2009); for Haspelmath (2009:44), the identification of a plausible source word and a donor language is key for recognising a certain word as loanword. In most cases discussed below, the donor language (or the donating language family) can be identified; yet there are also cases of transfer in which the direction of borrowing remains unknown. In principle, both languages involved can each be either the donor or the recipient. Transfer is plausible in particular when the putative loanword shows signs of phonological adaptation from the source language into the recipient language; thus, phonological adaptation is indicative of the direction of borrowing (Haspelmath 2009:45). Secondly, phonological and morphological adaptation are criterial to distinguish loanwords from code switching (Matras 2009:41). Contrary to code switching, loanwords should be used conventionally as parts of the recipient language (Haspelmath 2009:40). This criterion of conventionality is certainly important for the final loan status of a word, but cannot be checked for the languages under consideration here (but see footnote 11 below). I simply assume it to hold.

4 Lexeme Resemblances between Border and Nimboran

The lexical comparison between the Border family and the Nimboran family is primarily based on the vocabularies of Kilmeri, of Waris (Brown and Wai 1986) and of the single language of Nimboran. After the compilation of an alphabetic wordlist of Nimboran based on Anceaux’s grammar (1965), 337 pairs of words from Kilmeri and Nimboran designating the same concept could be compared.

The Nimboran terms are given with their lexically determined word accent in accord with their notation by Anceaux, as for instance, *méndu*. Regarding the

4 Gasser (2019:673) also considers synonyms as a guide to detect borrowed forms.

syllable structure of Nimboran, there are no word-final consonant sequences (Anceaux 1965:31; May 1997:13), while word-initial consonant sequences appear regularly. The constraints on syllable structure in Kilmeri are similar, yet consonant clusters are rarer than in Nimboran. Note that Kilmeri seems to show, in word-initial position, the development from nasals to plosives, namely /m/ > /^(m)b / and /n/ > /⁽ⁿ⁾d/, which sets it apart from the Waris branch of the Border family (see Appendix).

4.1 Nouns

We find 14 instances of lexical transfer of nouns between the two language families. In two cases the original family affiliation of the source lexeme remains unknown ('buttocks', 'neck; beak'). The terms are discussed roughly in the order of lexical fields.

'garden' The Waris branch of the Border family shows a common stem for 'garden', which takes the following forms: Waris *oso*, Manem *os*, Imonda *ov* (Seiler 1985), Amanab *aso* (Minch 1992:126). Pagi employs the very similar form *os*. This stem is also present in Nimboran, Kemtuik, Gresi, and Mlap as *usu* and in Mekwei as *asu* (May 1997:122; 126; Smits and Voorhoeve 1994:102). But three languages of two different branches of the Border family show entirely different words: Taikat has *manta* 'garden', Auyi *mu* has 'garden', and Kilmeri has *sele* 'garden'. Thus it seems plausible to me that the word originated in the Nimboran family and spread into the Border family.

'taro' There is a common Border word with Waris *saβa* (Brown and Wai 1986:96), Imonda *safa*, Manem *saf*, and Taikat *saf* referring to the indigenous taro plants (Smits and Voorhoeve 1994:108). Kilmeri, however, shows the form *wip* as the generic term for taro. This is borrowed from Nimboran *wip* (May 1997:18); cf. also Mlap *wip* and Kemtuik *wep* (1994:108).

'child' In Kilmeri the word for child shows the sex-neutral form *ruri*. In Waris it appears as {*mu*}-*tundis* 'girls' and *tuendis* 'boys'; the sound correspondences are regular. Imonda has the form *toand* 'boy, son', which is very close to Waris. Taikat has {*ma*}-*ntu* (Smits and Voorhoeve 1994:79). So we have a common, inherited word for the Border languages. In Nimboran and Gresi, 'child' is monosyllabic *du* (Anceaux 1965:15; Smits and Voorhoeve 1994:80). In Kemtuik 'child' appears as *do* [d^wo] (van der Wilden and van der Wilden 1975:37), in Mekwei as *do* (Smits and Voorhoeve 1994:80). Thus, the Nimboran family also shares the stem for 'child'. The transfer must have taken place between the families and before the Border internal sound change from Waris/Imonda /t,d/ to Kilmeri /r/. I argue

for the direction from Nimboran to Border, because in the Border languages the original stem became expanded into bi-/poly-morphemic words.

‘great-grandfather/parent’ For Kilmeri and Nimboran a lexeme is attested that refers to the generation above the grandparents; in Kilmeri it is sex-neutral, while in Nimboran it seems to designate males. The Kilmeri form is *básp*, and the Nimboran form is *babuásii* with stress on the penultimate syllable (May 1997:18). The bisyllabic structure of the Kilmeri term results from the loss of the second syllable of the Nimboran term, which precedes the syllable bearing the main stress. The Nimboran vowel sequence *ii* can be realised as [ik] (1997:18). In Kilmeri, syllable closures with /k/ are rare, and, if they occur, preferable have the form /ak/ or /(u)ok/; the coda /ik/ isn’t attested at all. So Nimboran [ik] is likely transferred as [ip], and Kilmeri is the recipient language.

‘sound, word, speech, story, language’ In the Border languages, the common inherited word referring to meanings like ‘sound’, ‘word’, ‘speech’, ‘story’, ‘language’ has the form *bɔ* (Kilmeri) or *mɔa/mv/mo* (Waris, Imonda, Amanab); the sound correspondence is regular. In Nimboran, the complex words *ne-mbwo* ‘word, language, speech, matter’ (May 1997:83) and *ne-mbwo-pem* ‘story’ (1997:53) are attested. Both expressions contain the morpheme *mbwo*, which is similar to Border *bɔ/mo*. The other Nimboran languages resemble Nimboran *ne-mbwo* closely (Smits and Voorhoeve 1994:254; van der Wilden and van der Wilden 1975:35). I argue for the direction from Border to Nimboran, because in the Nimboran languages the original stem became expanded into bi-/poly-morphemic words.

One might think that *bɔ* is a potential onomatopoeic form. However, Kilmeri has *muli/mui*.SG ‘say, speak’, *molije*.PL ‘say, speak’, and *muɛli* ‘talk to sb’ with Recipient object agreement, and I doubt that all these grammatically distinct forms are onomatopoeia.

‘tongue’ The word for ‘tongue’ is *ber* in Kilmeri, *meki* in Pagi, *mindɛ* in Waris, and *mɛnde* in Imonda; the forms are related via regular sound correspondences (see Appendix; Gerstner-Link 2018:31–41). A similar form we find in Nimboran with *méndu* (Anceaux 1965:18), but here it denotes ‘mouth’. The meaning shift from tongue to mouth is semantically plausible via (physical) contiguity (cf. Blank 1997:238–240), thus we can argue that the Nimboran word is a loan from Border. Probably it is taken from the Waris branch, since it shows the same consonantal phonemes. Kemtuik has the unrelated form [nr^mblɛn] ‘tongue’ (van der Wilden and van der Wilden 1975:37); this fact supports the direction of borrowing from Border to Nimboran.

‘behind, buttocks; faeces’ In Kilmeri, the word for buttocks is *eku*. In Waris, resemblant *akoko* is attested for ‘faeces’, while ‘buttocks’ is designated by an entirely different form in the Taikat and Waris branches of the Border family (Smits and Voorhoeve 1994:40–41). Yet in Nimboran proper (Anceaux 1965:22) we find *íaku* ‘buttocks’, which is formally similar to *eku* and *akoko*; trisyllabic *akoko* may be a partly reduplicated form. Since the other Nimboran languages have no forms designating ‘buttocks’ that can be related to those forms, one can assume that among the two language families there is an island consisting of the three resemblant forms above. A transfer between Waris/Kilmeri and Nimboran proper seems plausible including the meaning shift; but the direction of borrowing remains unknown.

‘hornbill, parrot’ The Kilmeri word referring to hornbills is *íwan*, while in Waris we find the unrelated form *peila* ‘hornbill’ (Brown 1986:78). Yet *íwan* is formally similar to *íway* ‘parrot’ of Kemtuik and Mlap (Smits and Voorhoeve 1994:130), and in Nimboran *ueiáy* ‘kind of small parrot, lory’ is attested (Anceaux 1965:30). Kilmeri lacks /ŋ/, while the Nimboran languages have both /n/ and /ŋ/ and could have taken over the word without adaptive change of the coda. Thus I conclude that Kilmeri borrowed the term from Nimboran and adapted it to its own consonant inventory. The meaning shift took place on the basis of the shared feature of a strong, curved bill.

‘kind of pigeon’ Kilmeri and Nimboran seem to share a term designating a certain type of pigeon (other than the crowned pigeon): Nimboran *imúo* and Kilmeri *imalo*. The referential property of pigeon-like birds holds for both languages. Formally, both languages show a trisyllabic word, nearly identical segments, and share the second-syllable stress. Nimboran’s only lateral is realised as retroflexed flapped lateral (May 1997:28; he subsumes it under the plosive series), while Kilmeri /l/ is a lateral approximant. When taking over Kilmeri *imalo*, the intervocalic approximant must have been dropped. A loan relationship with Kilmeri as the donor language is possible. The concept is not attested in other Nimboran and Border languages; therefore intra-family comparisons don’t work towards clarifying the direction of transfer.

‘neck; beak’ Kilmeri possesses several terms designating body parts of various animals. One of them is *besi* ‘beak’. The concept is not attested in other Border languages. In Nimboran we find *besí* ‘neck’ (Anceaux 1965:19), which resembles the Kilmeri word closely. On the assumption that Nimboran *besí* may also refer to a bird’s neck, transfer between the two languages is possible. The meaning shift seems plausible in either dir-

ection, since beak and neck are contiguous body parts of a bird, in front of the head and below the head.

‘mosquito; termite’ The Border family and Nimboran formally share a term that denotes various kinds of insects like ‘mosquito’ and ‘termite’ as well as unspecified ones. The Border languages have Waris *kles* ‘very tiny biting insects’ (Brown and Wai 1986:37), Imonda and Sengi *kles* ‘mosquito’, Kilmeri *kles* ‘mosquito’, and Pagi *eles* ‘mosquito’. This stem is not shared by Taikat, Auyi, and Manem (Smits and Voorhoeve 1994:136). In Nimboran we find *klesu* ‘termite’ (May 1997:124), while the Nimboran family forms for ‘mosquito’ are related to those of the Tor family (Smits and Voorhoeve 1994:137). It seems plausible that Nimboran borrowed the term *klesu* from the Border family and then shifted its meaning to ‘termite’.

‘mussel; bead’ Kilmeri *sájo* ‘fresh water mussel’ seems to appear in Nimboran {*uan*}*sáia* ‘kind of white bead’; the phoneme sequence is almost identical and the stress pattern is the same. Kilmeri also employs *sajo pul* ‘bead’ (lit. ‘mussel seed’), which would have supported the meaning shift from ‘mussel’ to ‘bead’. I assume Nimboran borrowed *sáia* from Kilmeri. For all the other Border and Nimboran languages, the concept ‘mussel’ is not attested.

‘sago grub, sago beetle’ In Kilmeri, sago grubs form a faunal class. Their classifying element is *bɛ(r)-* (Gerstner-Link 2018:646). In Nimboran we have *bre* ‘sago beetle’ (Anceaux 1965:11). The terms attested for Waris are *mɛŋemb* ‘beetle that produces edible grubs in sago’ (Brown and Wai 1986:50) and *nə_mbəl* ‘edible grubs’. The first element *nə* of *nə_mbəl* designates “the forest and its useful products” (Brown and Wai 1986:61). Pagi employs the same structure with *na_mpeɭ*. Thus we arrive at a common Border stem *ber/mbəl/mpɛɭ*, which was borrowed by Nimboran as *bre* due to the constraint that word/syllable-final /r/ is not allowed, while /r/ in consonant sequences is common (1965:31–35).

‘(vertical or horizontal) post in a house’ This meaning is only attested in two Border languages: in Kilmeri we have *jali* ‘supporting horizontal post’ and in Amanab *sumur* ‘housepost’ (Minch 1992:132). In Nimboran we find *jatɯ* ‘post’ (May 1997:37). The judgment of formal similarity between *jali* and *jatɯ* takes into account that Kilmeri lacks /t/. A transfer from Nimboran to Kilmeri is possible; then Nimboran /t/ would have been adapted as /l/. This adaptation is supported by the fact that word forms of the Waris branch with syllable-final /t/ appear with /l/ in Kilmeri: Waris *atxa* > Kilmeri *ɛɭ* ‘sugarcane’, Imonda *at* > Kilmeri *al* ‘leech’. Waris /x/ is lost in Kilmeri (see Appendix); so *ɛɭ* shows intervocalic /l/ like *jali*.

In addition, Kilmeri possesses *lɔɔs* ‘(vertical) housepost’ that designates the posts that are erected first on the ground when building a house. It may be that *jali* was taken over as a second term that would have allowed to distinguish between different kinds of post necessary for house building. Then we would have a case of “co-existence with the native word” (Haspelmath 2009:49), yet with specialisation of meaning.

4.2 *Verbs, an Adverb, and a Numeral*

Six verbs, one adverb, and one numeral are indicative of language contact between the families in question. In one instance a Border language (Kilmeri?) turns out to be the donor, in seven instances Kilmeri is the borrowing language.

‘go there, go thither’ Kilmeri possesses an inherently deictic verb *nɛ* ‘go thither’ (Gerstner-Link 2018:822; 837–840). In Nimboran ‘to go’ is a zero root (Anceaux 1965:158; May 1997:105), but there is a directional suffix *-ne* ‘from here to the end’ (May 1997:74), which has a similar deictic value as Kilmeri *nɛ*. Compare also the Nimboran postposition *ne* ‘to’ which expresses ‘motion towards’ as substitute of a verb (May 1997:121). For Waris *dəm* ‘going over there’ is attested, with no formal relationship to the Kilmeri word. For Imonda no verb designating the concept in question is attested. So it seems plausible to conclude that Kilmeri borrowed its inherently deictic ‘go’-verb *nɛ* from Nimboran’s directional suffix *-ne*.

The unmarked verb ‘to go’ in Kilmeri is *lɛ*; with the loan *nɛ* ‘go thither’ an inherently deictic verb was added to the motion verbs. We have a case of “co-existence with the native word” (Haspelmath 2009:49).

‘stand’ In the Border languages ‘stand’ can be regarded as an existential-postural verb. All these verbs have a singular and a suppletive plural form: in Waris *loxβ.SG/loβaxβ.PL* and in Imonda *lvh.SG/læfah.PL*. But the Kilmeri forms deviate from this shape, instead we find *nɛki.SG* ‘stand; erect’ and *pɔje.PL* ‘several stand’. The verb can be used both intransitively and transitively. In Waris we have *nəŋ* ‘think’, which appears in Kilmeri as *umul nɛki* ‘to think’, lit. ‘erect heart’. The sound correspondence between *nəŋ* and *nɛki* is regular (see Appendix). Nimboran has *nij-* ‘stand’ (May 1997:82), which is plausibly related to Waris *nəŋ* and Kilmeri *nɛki*. In Kemtuik, *ip* ‘to stand’ is attested (van der Wilden and van der Wilden 1975:33; 39; 55), which is different from the formally related Border and Nimboran forms. There seems to be an $nV(i,ɛ)C_{velar}$ island denoting ‘to stand’ formed by the languages of Kilmeri, Waris, and Nimboran, but the direction of transfer remains unknown.

‘distribute, share food’ In Kilmeri we find the rarely used collocation *ɪ pi* ‘to share freshly butchered meat’, while the default verb for sharing food with somebody is *ripɛi* with recipient/dative agreement (Gerstner-Link 2018:386) or *ripɛi.SG/rupɔpi.PL* ‘to distribute food among several persons’. The main verb *ɪ* of the collocation *ɪ pi* relates to Nimboran *iú* ‘to distribute’ (Anceaux 1965:28; May 1997:87). This Nimboran verb is construed with recipient/dative agreement. Because of the light verb construction in Kilmeri one can plausibly assume that Kilmeri borrowed the word from Nimboran.⁵ This is one more case of a loan that co-exists with the native verb, resulting in a semantic distinction, which is not attested in the other well-documented Border languages. In Waris we find *pɔa.SG/pɔaβul.PL* ‘to distribute food’, which does not relate to any of the Kilmeri forms.

‘hit, shoot, kill’ The Border languages share a common stem *lu/lo/lw* denoting the above meanings. Waris has *l-β/lu-β.SG* and *welxa-β.PL* ‘to shoot pigs with arrow’; Imonda has *lw.SG/lo.PL.A/lapi.PL.O* ‘to shoot’, and Kilmeri has *lui* ‘to kill, to hit’ (without a suppletive plural form, the object plural is indicated by a special quantificational suffix (Gerstner-Link 2018:347; 357)). Nimboran employs *luu-* ‘to hit’ (Foley 2018:450), also translated ‘to seize’ (May 1997:31). This verb does not formally relate to Kemtuik /pú.ik/ ‘to shoot’ (van der Wilden and van der Wilden 1975:49). Thus I conclude that Nimboran *luu* is borrowed from the Border languages, probably directly from Kilmeri.

‘be sick’ Kilmeri has the verb *mari.SG/marmarpi.PL* ‘to be sick’ denoting sickness of any kind; severe illness is indicated by the augmented form *nɔ-mari*. A verb with this meaning is attested neither in Waris and Imonda nor in any other Border language. But looking at Nimboran, we find *máre* ‘unconscious’ (Anceaux 1965:12; 24), and it makes sense to relate this word to Kilmeri *mari*. Nimboran *máre* may be an adjective or a stative verb; either way, it could have been borrowed across word class boundaries. I take it to be a verb, and the meaning shift from ‘unconscious’ in Nimboran to ‘being sick’ in Kilmeri is straightforward.

Anceaux mentions the possibility to form verbs from adjectives by use of verbal morphology (1965:120–121) and describes the infinitive—the root morpheme—as quasi-adjective that may combine with nouns (1965:112).

5 In Kilmeri, light verb constructions are normally used with adjectives and nouns in order to verbalise them. Plausibly, the same strategy was formerly used to distinguish borrowed verbs from formally (almost) identical native verbs, here *ɪ* ‘to recede’. In current Kilmeri LV constructions are used to integrate Tok Pisin verbs.

This supports the assumption that Kilmeri originally borrowed the verb from Nimboran.

‘answer’ Kilmeri possesses several verbs of speaking including *wui-* ‘to answer’. This verb shows obligatory agreement with the recipient/dative object (Gerstner-Link 2018:386). In Nimboran the respective verb is *uú-* ‘to answer’ (Anceaux 1965:124); it is construed with obligatory agreement of the recipient/dative argument like its Kilmeri counterpart. There are no data for ‘answer’ in the other languages of the Nimboran family. The formal and structural parallelism of the Kilmeri and Nimboran word makes a transfer probable. In Kilmeri, *wui-* seems to be an old word which is in the process of being replaced by the serial verb *dɔri_muɛli* ‘turn back_talk to sb’, a more frequently used verb.

‘before, formerly’ Kilmeri *kimikɛ* ‘before, formerly, in former times’ seems to be an isolated form in the Border family. For this meaning data are available only in a few languages: in Waris we find *dɔara* ‘before, previously’ and *nəmat* ‘a long time ago’ (Brown and Wai 1986), in Imonda *iauwnam* ‘in earlier days’ (Seiler 1985:27), and in Amanab *autunam* ‘long time ago’ (Minch 1992:120). None of the three words shows any similarity with the Kilmeri word. Yet in Nimboran we have *mínje* ‘before’ that can be related to Kilmeri, which has also the (less frequently used) short form *mikɛ*. Most probably, the Nimboran word was borrowed and phonemically adapted. Both words $\{ki\}$ *mikɛ* and *mínje* might also contain *kié* ‘time’ (Anceaux 1965:28); when taken over by Kilmeri, the Nimboran term must already have been fused.

‘two’ The numeral ‘two’ shows similar forms in Kilmeri *dupua* and Nimboran *namuán* (May (1997:50) spells *namwan*). Intervocally Kilmeri has an *n* as well, as shown by the form *rɔ-dupua* EMPH-two, which is realised as *rɔ-nupua*. ‘Four’ is *rɔdupua rɔdupua* in Kilmeri, typically realised as *rɔn-pua rɔnpua* (Gerstner-Link 2018:123). Note also the free variation of the onset in different Kilmeri speaking villages (cf. Brown 1991): Ilup *nɔpwa* and Isi 1 *nupwa* with a nasal versus Osol *dupwa* with the occlusion /ⁿd/ like Ossima *dupua*. We also find the same type of variation with labials: ‘sister’ is *muri* in Osol, but *bɔri* in Ossima and Oup (cf. Brown 1991). This might also account for the word-medial difference of Kilmeri /p/ versus Nimboran /m/.

The other Border languages use a different stem for ‘two’: Waris, Imonda, and Pagi have *sabla*, Amanab has *sabaga*, while for Taikat the two (unrelated) forms *sember* and *nanjer* are attested. Clearly, Taikat *sember* relates to *sabla* via metathesis of the liquid. Obviously, Kilmeri doesn’t fit in here, and I assume the language acquired *dupua* ‘two’ from Nimboran.

Kemtuik has *namuan* like Nimboran proper, Gresi has *namwan*, and Mekwei *naman* (Smits and Voorhoeve 1994:212). Despite the fact that the currently observable free variation of nasals and homorganic plosives in Kilmeri may not explain the consonant change in an old loan, I think that the hypothesis of transfer is the best account for its deviating form *dupua* ‘two’, which otherwise would stand as an entirely isolated form.

4.3 Summary

The comparison of Border/Kilmeri and Nimboran vocabularies results in 22 instances of lexical transfer between single languages and between families: 14 nouns, six verbs, one adverb, and one numeral. The transfer of nouns is symmetrical; the transferred nouns are related to nature and environment, kinship, body parts, natural kinds, and material culture. The transfer of verbs goes from Nimboran to Kilmeri in most cases. Since Kilmeri lacks certain consonants, in the direction from Nimboran to Kilmeri phonological adaptation of the loans is required: /ŋ/ > /n/ syllable/word-finally (*ueián* > *ivan*) and /ŋ(g)/ > /k/ intervocally (*nin* > *neki*, *mínje* > {*ki*}-*mike*); /t/ > /l/ intervocally (*jatú* > *jali*). In the opposite direction from Kilmeri to Nimboran we find metathesis to prevent final /r/, which isn't permitted phonotactically: *ber* > *bre*. Co-existence with inherited lexemes occurs four times with three verbs and a noun: *i pi* ‘to share butchered meat’, *ne* ‘go thither’, *wui* ‘to answer’, *jali* ‘post’. The verbs borrowed from Nimboran into Kilmeri illustrate different strategies of integration: (i) We find direct insertion of the stem/word (Wohlgemuth 2009:87–89); (ii) We find the citation form plus a light verb as in *i pi* ‘to share’ from Nimboran *iú* ‘to distribute’ (2009:102–109); (iii) We find the re-analysis of a directional suffix in a Nimboran zero stem verb as a verb: *-ne* ‘from here to the end’ becomes *ne* ‘go thither’ in Kilmeri.

5 Lexeme Resemblances between Border and Sentani

The vocabularies of Kilmeri and Sentani are compared on the basis of Cowan's grammar whose vocabulary list provides about 500 entries (1965:75–88). But only six pairs of words designating the same concept qualify as instances of putative lexical transfer; they belong to different word classes and are now presented one by one.

Note that four of the proposed loans into Kilmeri are either used infrequently (‘wallaby’, ‘place’), restricted to a very narrow context (‘like’), or add a special meaning in a certain grammatical domain (NEG). It is the semantic constraints on ‘like’ and NEG that may also account for their relatively infre-

quent use. They all co-exist with inherited forms of Kilmeri, which suggests more or less deliberate expansion of the vocabulary and reduces chance similarity in favour of contact-related transfer. It may well be that the “new” term for ‘wallaby’ may have designated a particular kangaroo species in Kilmeri, a distinction lost today.

‘wallaby, tree kangaroo’ The ordinary Kilmeri terms for ‘wallaby’ and ‘tree kangaroo’ are *bi_sem* and *bi_puel*; the first element *bi* is the classifying element, still used as an independent noun meaning ‘pig, terrestrial animal’. Yet there is a less frequently used term *emei* ‘wallaby’ in Kilmeri. This is clearly related to Sentani *eme/emeho* ‘forest kangaroo’ and borrowed from this language (Cowan 1965:78). In Waris, by contrast, we find the lexeme *pind* ‘marsupial’ which is cognate to Kilmeri {*bi*}_*per* ‘possum’ via the regular sound correspondence /d/ < > /r/ (see Appendix). Pagi has *som* ‘wallaby’ which resembles {*bi*}_*sem* of Kilmeri.

‘village, place’ In addition to the inherited lexeme *jilau* ‘village’ (< *jip_lau* ‘house_place’), Kilmeri has the word *jo* ‘place’. It is not frequently used, but once in a while it occurs in texts and in spontaneous discourse. It appears to be a transfer of Sentani *jo* ‘village’; Tabla also has *jo* ‘village’ (Gregerson and Hartzler 1987:14). By contrast, in the other Border languages we find Pagi *ji_tau* ‘village’, Imonda *la* ‘village’, Waris *la* ‘nest of bird or pig or insect’ (Brown and Wai 1986:41), which are cognate with Kilmeri *ji_lau* ‘village’.

‘sit, stay, live, settle, dwell, remain’ The Kilmeri existential-postural verb for singular/dual animate referents *nake* ‘to sit, to stay, to live’ has no cognate counterpart in the Border languages. Imonda has *afv* ‘to sit’ and the singular/plural pair *ale/a-fia* ‘to stay, to remain’; Waris has *aβ.SG/æβuβ.PL* ‘to sit’ (Brown and Wai 1986). Imonda *afv* and Waris *aβ/æβuβ* are cognates. It may be that Kilmeri *mapε* sit.PL is also etymologically related to these forms. For Taikat *amber* and *amramrap* are attested; the latter form might be a plural form because of its reduplicational structure (Smits and Voorhoeve 1994). Presumably, the Taikat words are cognate with the Waris lexemes. By comparison, Kilmeri *nake* ‘sit.SG’ is entirely different. Thus, the transfer of this verb form from Sentani *nəkə* ‘sit down, settle, dwell, stay, remain’ (Cowan 1965:85) into Kilmeri is likely.⁶

6 This loan relationship may be indicative of the fact that, some centuries ago, the Kilmeri people were forest dwellers who lived on hunting and only gradually developed a horticultural life style as the Sentani practise around Lake Sentani.

'like' Kilmeri has the special verb *kina* 'to like' without an etymology in the Border family. It is probably transferred from Sentani and Tabla *kəna* 'to like' (Gregerson and Hartzler 1987:13). In Kilmeri, *kina* co-exists with the inherited verb *muli* 'to want, to like' (Gerstner-Link 2018:490) and only appears as first component verb in verb serialisations with perceptive verbs denoting positive perceptions.

Distal deixis The common Border stem for distal deixis is *di/ri*, and it denotes spatial distance. In Kilmeri we have *ri-ɔ* 'there, that', consisting of the deictic stem plus a local suffix. In Waris we have *di* 'over there' (Brown and Wai 1986). The local distal deictic in Imonda is *ed* 'there' (1985:45), cognate with Kilmeri distance-neutral *εε* 'this, that'. In Sentani we find the following forms: *dika* 'that, those, yonder' as local deictic (Cowan 1965); Gregerson and Hartzler (1987:11) have Central Sentani *ndi* 'that' and East Sentani *ri(ki)* 'that', while for Tabla *di* 'that' is attested. These forms contrast with *dakə* 'this, these': *di-* denotes distality, while *da-* denotes proximity.

The distal deictic forms of Kilmeri and Waris relate to the distal stem *di-* of Sentani and Tabla, while their cognate proximal stems (Kilmeri *ɔ*, Imonda *ph*, Waris *honi*) are different from Sentani *da-*. This is an argument for the direction of borrowing: The form of the distal deictic was borrowed from Sentani. Because of the onset variation *di/ri* in the Border languages it is an old loan that was transferred before the intra-Border sound change /d/ > /r/ emerged.

Negative particle In addition to the normal verbal negation *ar* 'not', Kilmeri employs a special emphatic verbal negation *ba* (Gerstner-Link 2018:633). Pagi has a similar form *bam* 'no, nothing' (Gerstner-Link 2000). Kilmeri *ar* is cognate with Imonda *at*, which renders a sentential negation 'it is not the case' (Seiler 1985:171). The origin of *ba/bam* is less clear. In the Waris languages the narrow-scope verbal negation appears as *mas VERB-mo* in Amanab (Minch 1992:147), while Waris itself has a probable cognate form in the verb-final negative suffix *-moa* (Brown 1990: 11,21). Imonda shows discontinuous *sə VERB-m*, and, in addition, has a form *bal* that is suffixed by *-m* and serves as negation of verbless clauses. Seiler (1985:171–172) calls *bal* a "dummy element".

Could all these *ba(C)* forms of negation known in the Border family be related to Sentani *bam*, whose (quite broad) meaning is given as 'not, hardly; without; no good, bad' (Cowan 1965)? It seems plausible to assume that Kilmeri took the negative particle from Sentani as a pronounced second verbal negation despite of its more general negative function in that language (cf. Sentani *fə bam wali bam* 'without fear (and) without

life', i.e., 'impudent and careless' (1965:79)). Pagi took over the negative particle, too, but with a slightly different meaning.

In sum: The lexical transfer between the Sentani and the Border languages is unidirectional; the latter are the borrowing languages in all six instances. Given the lexical data that can be compared, this is a very low number of loanwords. The lexical entries in Cowan's grammar (1965) number about 600; among these are roughly 500 concepts for which a Kilmeri counterpart is known. From about 500 compared lexical items only six or 1% are shared. The borrowed negative particle co-exists with the inherited negative particle in Kilmeri, Pagi, and Imonda.

Deictics and negation/negative markers appear to be rarely borrowed; they are not listed in Matras's frequency-based hierarchy of borrowed function words (Matras 2009:157, 2007:32–36). In the case of their transfer from Sentani to Kilmeri/Border these borrowings served to expand a certain grammatical domain. The transfer of the deictic *ri* made a distinction possible that didn't exist before in the deictic system of Kilmeri. The inherited Border distal *ere* became restricted to questions containing a deictic, and it acquired the temporal meaning 'now', which is never attested with the proximal stem *o* (Gerstner-Link 2018:795–797). The new, borrowed distal took over the general distal function in Kilmeri's deictic system (Gerstner-Link 2018:797–801).

6 Lexeme Resemblances between Border and Skou

In this section I deal primarily with the single language called Skou, but other languages of the Sko family will also be taken into account if they may shed light on a certain question. These languages are *Isaka*, *Barupu*, *Wutung*, and *Dusur*; they are chosen because their grammars also provide vocabulary lists. Regarding loanwords in Skou, Donohue says the following: "In addition to this native lexicon, we can recognise a number of loans from languages with which Skou has been in contact. [...] There are probably also a number of words that find their origin in the languages related to *Mbo* (Kilmeri), *Elseng* (Morwap), *Tobati* and *Sentani*, but since lexical materials on these languages are scarce little can be said for that possible connection." (2004:31) Indeed, Kilmeri can be shown to provide a few source lexemes for Skou. The Skou and *Wutung* lexemes are given with their tones according to Donohue: *a* low, *á* high, *à* falling pitch (2004:99, 524–573) and *Marmion*: *á* high, *à* low, *â* highlow pitch (2010:93).

6.1 Nouns

'hole, hollow; empty' Skou *bí* 'empty' (Donohue 2004:524) can be related to Kilmeri *bi* 'hole, hollow' which represents the common Border form $C(b,m)V(i,ie,a)$. The phrase *bi solɔ* 'hollow only' means 'empty' in Kilmeri; an empty house is referred to by *jip bi solɔ* 'house hollow only'. This phrase shows the syntagmatic contiguity of 'house' and 'hollow' in Kilmeri. The Skou form *bí* has three additional meanings, namely, 'floor', 'shell, plating', and 'tree with air roots'. Donohue seems to interpret this form-meaning correlations as a quadruple homophony of *bí* instead of polysemy (2004:524). At first sight, homophony of four lexical entries of the form *bí* seems to make sense, since the four meanings appear to be quite different and unrelated. But there is a common seme of these meanings, namely 'hollowness'. This type of space can only be defined in terms of a surrounding structure delimiting the cavity enclosed by it. In particular: The word 'empty' calls up the concept of container defining an empty space. 'Floor' circumscribes the space beneath a house (downward direction) and beneath its roofing (upward direction). Often the floor is the only planar, extended confinement of a Papuan house (especially with regard to cooking houses). 'Shell' designates the "house" of mussels; they live in a cavity confined by the shell material. A tree with air roots—e.g., a Banyan tree (*Ficus Benghalensis*)—can also be conceived as creating a cavity that can be entered; one may feel like being in a "house" confined by a set of (more or less densely) hanging air roots.

Thus the semantic transition from Kilmeri *bi* 'hollow' to Dusur *bí* 'house' based on the seme of cavity is not too far-fetched; it relates to the concept of INTERIORICITY (Aikhenvald 2000: 277; 289), which is a well-known concept for establishing noun classes (other such concepts are, inter alia, SHAPE, SIZE, POSITION, DIMENSIONALITY, CONSISTENCY (2000:275–293)). In view of this, the meaning shift from 'hole, hollow; empty' to 'house' is quite plausible semantically.⁷ I conclude that Skou and Dusur borrowed the word *bí* from Kilmeri. This is supported by the following lexical findings: In Wutung we find *péy* 'house' (Marmion 2010:374) as well as *lóng* 'hole, opening' (2010:372). Skou has *pá* 'house' (Donohue 2004:528) and *i* 'hole' (2004:526). Likewise, I'saka and Barupu show no similarity between their terms denoting 'hole' or 'empty' and Skou/Dusur *bí*.

7 Of course, noun classification and meaning shift in language history are two fields. But a conceptual overlap should not be excluded a priori. Note that Kilmeri 'rice' is *dipsu* from *dipi_su* 'ant_egg', which is clearly a calque based on the shape of rice grains, that is, on the concept of SHAPE.

Furthermore, Donohue's cognate set for 'house' is not entirely convincing. Proto Skou *a can correspond to either *a* or *i* only in Leitre, while it is retained as *a* in all other Inner-Sko languages (2002:183; 188); compare the set for 'hair' (2002:189). In addition to the regular correspondence sets for Proto Skou vowels Donohue gives irregular sets (2002:189), and 'house' would also be an instance of it. In the form *bí* it is only the plosive that shows a regular change from *p > b. Donohue states: "Some unproblematic correspondence sets are found for vowels, but in addition to the cases summarized in table 2.1., there are many awkward correspondence sets, which probably reflect a long period of intense interaction and multiple reborrowings of words back and forth." (2002:188) Presumably, Donohue means intra-Sko family borrowings—but it could as well be that external borrowings are involved in the irregular picture he describes.

'sago (jelly), portioned sago' Skou possesses the word *ná* 'sago package' (Donohue 2004:527), which probably denotes portioned sago wrapped in a leaf. Other terms relating to sago are *hòe* 'sago palm', *hòe è* 'sago porridge', and *kóe* 'sago pancake' (Donohue 2004). Clearly, the forms *hòe* and *ná* cannot be related. In Kilmeri we have *due* 'sago palm' and *ja* 'sago jelly', in Waris *na* 'sago palm' and *jes* 'sago jelly'; in Taikat 'sago' is also *na* (Smits and Voorhoeve 1994). Waris/Taikat *na* and Skou *ná* are formally most similar. It seems possible that Skou took over the word from one of these languages by shifting and specialising its meaning, adding a new word to its own repertoire of expressions relating to sago.

'burn; fire' The Border languages share a stem $C(t,r)V(a,\varepsilon)$ 'to burn' as intransitive verb. Waris *ta-* is said to refer to the situations of the kind 'fire is burning' or 'food is cooking' (1986:112). Kilmeri *rɛ* 'burn' can be rendered as 'fire is blazing' or 'food is cooking/done'. In both cases the verb denotes the process of burning *and* the visible event of a fire. But the languages also possess a special word for fire, viz., *sue/sv*. However, in Skou we find *ra* 'fire' and *rà li* 'burn' with *li* 'do' in a light verb construction (Donohue 2004:529). This word is similar to the Border stem for 'burn', especially to Kilmeri *rɛ*, if we take into account the Skou rule "There is a consistent pattern in which mid open vowels lower in Skou following an *h or in a falling tone syllable." (Donohue 2002:188) Then it seems possible to conclude that Skou borrowed the word *rà* 'fire' directly from Kilmeri. There is also the compound *rá rí* 'burning wood' (lit. 'fire_tree', 2004:235). The meaning shift involved is plausible.

'bush knife' In the Border languages we find Kilmeri *nebi* 'bush knife', Waris *nabe* 'chopper, machete' (Smits and Voorhoeve 1994), and Taikat *nabej* 'chopper, machete' (Smits and Voorhoeve 1994). Wutung has *nápè* 'bush

knife' (Marmion 2010:94, 100; 373). The Wutung lexeme is a clear resemblance to the stem present in all three branches of the Border family; because of closest formal similarity it is probably borrowed from Taikat or Waris. Skou *anábí* 'machete' may also be taken from Border, while *táng* 'machete' (Donohue 2004:534) is certainly an old Macro-Skou word.

6.2 Other Lexemes

'shoot, hit' The verb denoting the hunting activities of shooting and hitting has already been discussed in Section 3.3. The Border languages share the common stem *lu/lo/lv* with these meanings, and the Skou form is *lú* 'shoot' (Donohue 2004:527). Thus I conclude that not only Nimboran, but also Skou borrowed this verb from the Border family, probably directly from Kilmeri *lui* because of the vowel quality.⁸

'good' Kilmeri employs the lexeme *maki* 'good, of best quality' that, at first sight, is hard to relate to an adjective form of the Border languages with (roughly) this meaning. Yet we have Ainbai *man̄gri* 'good' (Brown 1991) and Waris *maka-l* 'mature, big fruit' (Brown 1986). These three forms share the stem *man̄-/mak-*; the meaning shift involving Waris is plausible. Thus we can say that there is a common Border form with the meaning of 'good, big' found in two branches of the family. Skou possesses—except for the suprasegmental feature of tone—a formally identical adjective *máki* with the meaning 'big' (Donohue 2004:527). Wutung, which is adjacent to Skou, has *húwúrti* 'big' (Marmion 2010:371). The eastern-most Sko language Barupu has *pako* 'big, be big' (Corris 2005:383).⁹ Despite of the meaning shift towards size only I conclude that Skou *máki* 'big' is a loan from Kilmeri.

'well, then' Skou *so* 'well, then' is of "(highly) suspected non-Skou origin[s]" because the *s* cannot be assumed to be an allophone of one of the Skou phonemes (Donohue 2004:35). A possible solution regarding the foreign origin of this particle can be found in Kilmeri *sɔ* and/or *sɔ sɔlv*, which

8 A look at the other documented Skou languages shows the following: Wutung has *qa* 'to hit', *qbaqba* 'to hit', *qaqwa* 'kill.1SG > 3SG.M' (Marmion 2010:374) as well as *lô* 'sharp' and *lálqè* 'sharp' (2010:372). Here it seems that *lô* 'sharp' is a contact-induced second adjective, conveying a meaning already present, that goes back to the Border stem for 'shoot, hit'. I'saka has *-a* 'hit' and *-o* 'shoot' (Donohue and San Roque 2004:95). Barupu has *ti* 'to shoot' (Corris 2005:388). These vocabulary findings support the loan origin of Skou *lú* 'shoot'.

9 The Skou family words for 'good' are as follows: Donohue gives *héfèng* 'good' (2004:525). Smits and Voorhoeve (1994) attest *efe|hèfè|hè:pè* 'good'. For Wutung we find *félàì* 'good, nice' and *muti* 'good' (Marmion 2010:370; 373). I'saka has *èi* 'good' (Donohue and San Roque 2004). In Barupu 'good, be good' comes in the two variants *neman/nevai* (Corris 2005:381).

has the pragmatic value of affirmation of an ongoing process. Note that Wutung has *so* ‘okay’ (Marmion 2010:375), which also fits the meaning of the Kilmeri particle. Most likely both languages borrowed the word from Kilmeri.

6.3 Summary

The number of transferred lexical items is low again: seven words in sum, with four nouns, one verb, one adjective, and a pragmatic particle. In six instances Skou is the recipient language and in one instance Wutung. In one case (‘sago package’) the loan co-exists with inherited terms and adds a specialised concept. In all instances of borrowing, Skou and Wutung need to integrate the loans from the Border languages into their tonal systems. The verb ‘shoot, hit’ is probably taken over in its past form *lu* and then integrated into the morphological structure of Skou.

7 Lexeme Resemblances across the Border, Nimboran, Sentani, and Sko Families

Lexeme resemblances across many languages and several families suggest the phenomenon of *wanderwörter* that spread over a geographical area (cf. Haspelmath 2009:45). They are either the result of direct contact between several languages, or else they spread via extensive use by traders who cross different, rather small language areas, as we find them in Central Northwest New Guinea. Candidates for such *wanderwörter* could be the words discussed in this section: ‘water’, ‘tree’, ‘leaf’, and ‘arrow’; these words can be associated with bird of paradise hunting. Two of the words are basic lexical items that are otherwise not easily borrowed, viz., ‘water’ and ‘leaf’ (Tadmor et. al. 2010:239–241);¹⁰ since the authors include the age score of a word in determining the “basicness” of a vocabulary item (2010:237), the spread of a certain form designating water is remarkable as it counts as a stable item. But bird hunting and plume trading may have facilitated the acquisition of these words that became lasting items of the vocabularies of several families.

10 Foley mentions two examples of language contact on the northeast coast of New Guinea that resulted in quite a number of loanwords that belong to the basic vocabulary, be it among genetically related languages like Watam and Kopar of the Lower Sepik family or be it among an Austronesian and a Papuan language like Mangap-Mbula and Kovai (2010:799). Cf. also van den Heuvel and Fedden (2014:32–33). Gasser (2019) examined Austronesian loans in Papuan languages of the Bird’s Head and the Cenderawasih-Bay.

The distribution across languages of the putative *wanderwörter* is shown in Tables 8.5–8.9:

TABLE 8.5 ‘water, rain, river’

Language	Reference	‘water’	‘rain’	‘river’
Border family				
Waris	Brown 1986	<i>pɔ</i>	<i>pɔ</i>	<i>pɔ</i>
Imonda	Seiler 1985	<i>pɔ</i>	<i>pɔ</i>	<i>pɔ</i>
Kilmeri	Gerstner-Link 2018	<i>pu</i>	<i>pu</i>	<i>pu</i>
Pagi	Gerstner-Link 2000	<i>pɔ</i>	<i>pɔ</i>	<i>pɔ</i>
Taikat	Smits & Voorhoeve 1994	<i>wea</i>	<i>bu, mu</i>	<i>wea</i>
Elseng	Menanti 2005, Burung 2000	<i>vetev</i>	<i>jai</i>	<i>vetev</i>
Nimboran family				
Nimboran (lang.)	Anceaux 1965	<i>bu</i>	<i>sai</i>	<i>bu</i>
Kemtuik	van der Wilden 1987	<i>bu</i>	<i>sa</i>	
Sentani family				
Sentani (lang.)	Cowan 1965	<i>pu, bu</i>	<i>ja</i>	<i>wi</i>
Tabla	Gregerson & Hartzler 1987	<i>bu</i>		<i>wai</i>
Sko family				
Skou	Donohue 2004	<i>pa</i>	<i>fu</i>	<i>pa</i>
Wutung	Marmion 2010	<i>tʃã</i>	<i>fə</i>	
I’saka	Donohue & San Roque 2004	<i>wì</i>		<i>wì</i>
Sumo (Bouni)	Miller 2017	<i>pi:</i>	<i>bɔ:</i>	

The lexical item that spread is *pɔ/pu*. It occurs in 11 languages; gaps in the columns are due to lack of data. In the Nimboran family it only refers to ‘water’ and ‘river’, while in Taikat, Skou, and Wutung it specifically denotes ‘rain’. In Taikat, Nimboran, Kemtuik, Sentani, Tabla,¹¹ Skou, and Sumo it co-exists with other terms of the same lexical field. Skou *pa* and Wutung *tʃã* belong to a

11 Tabla *bu* and Sentani *pu* occur in nominal collocations like *doi bu* ‘sweat’ and *roi pu* ‘sweat’ (Gregerson and Hartzler 1987:10). This shows that *bu/pu* are conventionalised, compositionally productive lexemes in these languages.

TABLE 8.6 'tree, wood'

Language	Reference	'tree', 'wood'
Border family		
Waris	Brown 1986	<i>ti</i>
Imonda	Seiler 1985	<i>ti</i>
Kilmeri	Gerstner-Link 2018	<i>ri</i>
Pagi	Gerstner-Link 2000	<i>ki</i>
Taikat	Smits & Voorhoeve 1994	<i>ti, di</i>
Nimboran family		
Nimboran (lang.)	Anceaux 1965	<i>di, ri</i>
Kemtuik	Smits & Voorhoeve 1994	<i>di</i>
Sentani family		
Sentani (lang.)	Cowan 1965	<i>o</i>
Tabla	Gregerson & Hartzler 1987	<i>o</i>
Sko family		
Skou	Donohue 2004	<i>ri</i>
Dumo	Donohue 2002	<i>ti</i>
Dusur	Donohue 2002	<i>ti</i>
I'saka	Donohue & San Roque 2004	<i>téi</i>
Sumo (Bouni)	Miller 2017	<i>?ái</i>
Barupu	Corris 2005	<i>ai</i>

well-established cognate set (Donohue 2002:187); therefore the spread words are *fu* and *fə*. Since *pɔ/pu* has the widest denotational range in two branches of the Border languages, I assume that it spread from these languages into others in which it takes over one or two meanings.

The lexical item that spread is *ti*, yet it is not found in the Sentani family. In Skou the sound change /t/ > /r/ took place (Donohue 2002:200). In Border we have the following correspondences between the Waris branch and Kilmeri: /t/ corresponds to /r/ syllable-initially; /ⁿd / corresponds to /r/ in other positions (see Appendix). Kilmeri and Pagi show the regular correspondence /r/ < > /k/ (Gerstner-Link 2018:31–35). The Piore branch of the Sko family has another word for 'tree'. I conclude that *ti* spread from the Border family; because of the sound changes in the Border family and Skou it is an old transfer.

TABLE 8.7 'leaf'

Language	Reference	'leaf'
Border family		
Waris	Brown 1986	<i>βeɛ</i>
Imonda	Smits & Voorhoeve 1994	<i>lop</i>
Kilmeri	Gerstner-Link 2018	<i>pɛɛ</i>
Pagi	Gerstner-Link 2000	<i>pɛɛ</i>
Taikat	Smits & Voorhoeve 1994	<i>fælej</i>
Nimboran family		
Nimboran (lang.)	Anceaux 1965, May 1997	<i>pró, plo</i>
Kemtuik	Smits & Voorhoeve 1994	<i>dɔp</i>
Gresi	Smits & Voorhoeve 1994	<i>dɔp</i>
Sentani family		
Sentani (lang.)	Cowan 1965, Gregerson & Hartzler 1987	<i>fe, fæ</i>
Tabla	Gregerson & Hartzler 1987	<i>{kə}pei</i>

The spread form is shaped $C(p, \beta, f)V(e, \varepsilon, \text{æ})lV(\varepsilon, o)$, with /l/ in third position in all Border forms except Imonda; in Nimboran we have vowel elision and /l/ appears in second position. In Sentani only the first syllable is present. Because of the syllable structures I conclude that the word originated in the Border family as a bisyllabic item.¹²

The spread form is shaped $C(p, f, \phi)V(\varepsilon, \text{æ}, a)C(l, r)V(e, a)$. In Waris, Kilmeri, and Pagi the word is monosyllabic; additionally, there is a meaning shift to 'bow' in Waris. Kilmeri and Pagi lack labial fricatives in their inventories; the sound correspondence Kilmeri /p/ < > Waris /β/ and /p/ is regular (see Appendix). Elsens has /ϕ/. In Nimboran proper and Kemtuik vowel elision of the first syllable took place. All Nimboran languages have the onset /p/; Nimboran proper and Kemtuik lack labial fricatives in their consonant inventories (Anceaux 1965:9; van der Wilden 1975:51). The sound change from Tabla /p/ to Sentani proper /f/ is regular (Gregerson and Hartzler 1987:4–5).

12 In Northwest New Guinea, a number of Papuan languages borrowed their words for 'leaf' from Austronesian languages in their vicinities (Gasser 2019:651; 654), yet in her sample 'leaf' belongs to the least borrowed items (2019:635).

TABLE 8.8 'arrow'

Language	Reference	'arrow'
Border family		
Waris	Smits & Voorhoeve 1994	<i>pæ</i> 'bow'
	Brown & Wai 1986	<i>βalɔ</i> -{ <i>ngɔ</i> } 'bow'
Imonda	Seiler 1985	<i>fal</i>
Kilmeri	Gerstner-Link 2018	<i>pɛ</i>
Pagi	Gerstner-Link 2000	<i>pai</i>
Taikat	Smits & Voorhoeve 1994	<i>fale, fara</i>
Elseng	Menanti 2005	<i>φal</i>
Nimboran family		
Nimboran (lang.)	Smits & Voorhoeve 1994	<i>pro</i> { <i>daj</i> }
Kemtuik	van der Wilden 1975	<i>ple</i>
Gresi	Smits & Voorhoeve 1994	<i>para</i> { <i>daj</i> }
Sentani family		
Sentani (lang.)	Cowan 1965, Gregerson & Hartzler 1987	<i>fəla</i>
Tabla	Gregerson & Hartzler 1987	<i>pəra</i>

8 Conclusion and Discussion

8.1 *Types of Borrowed Items*

The lexical transfer between the Border, Nimboran, Sentani, and Skou families presents a manifold scenario. We see *wanderwörter* that are found in languages across several language families and we see words that are found in only two language families, viz., in the Border family and in just one of the other families. Regarding the word classes transferred items belong to, we count 15 non-nouns vs. 20 nouns plus four nouns of the category *wanderwort*. This distribution shows that nouns are indeed more easily borrowed and dispersed than other words. With 24 to 9 items, the ratio of nouns to verbs is close to three-to-one, and is roughly in line with the average ratio found by Tadmor (2009:61–62) in the database representing the languages of the world.

Semantically, the nouns belong to the domains of nature and environment, kinship, body parts, natural kinds, and material culture. The verbs belong to the domains of motion, existence/posture, hunting, eating, and being sick. Field-

TABLE 8.9 Identified transfers in numbers and word class^a

Source language	Target language				
	Kilmeri / Border family	Nimboran proper	Sentani proper	Skou proper and eastern Sko	Wander- wörter
Kilmeri/Border family		6 nouns 1 verb		4 nouns 1 verb 1 adjective 1 particle	3 nouns
Nimboran proper	6 nouns 5 verbs 1 adverb 1 numeral				1 noun
Sentani proper	2 nouns 2 verbs 1 deictic 1 neg. particle				
Skou proper and eastern Sko					
Direction of transfer unknown		2 nouns			

a There may be unidentified items of transfer among the languages under investigation.

related constraints or preferences cannot be detected; instead, the words in question appear to be a selection across the whole lexicon. Quite a few meanings of borrowed or areally dispersed items discussed in the present study occur in “The Leipzig-Jakarta List of Basic Vocabulary” (Tadmor et al. 2010:239–241); the meanings are given with their rank in this list: to go (3), water (4), tongue (6), neck (23), to stand (45), child (51), to burn *intr.* (53), good (56), not (56), leaf (64), wood (80); some meanings obtain the same rank in the list. The ratio of all borrowings to core vocabulary borrowings is 38:11; that is, just under 30% belong to the core vocabulary. This result suggests that items of the core vocabulary are not in principle resistant to borrowing.

The lexical transfer shows a strong tendency to asymmetry. In the case of Nimboran proper versus Kilmeri/Border family, Nimboran is the source language in 13 instances and the target language in 7 instances. Sentani is only a source language with respect to the Border family lexicon. On the other hand, the Border family, and Kilmeri in particular, is the main source for loans into Skou. The number of traceable transfers is low; the Border and Nimboran fam-

ilies possess the highest number of contact-related lexical items.¹³ The relative high amount of transfers between Kilmeri and Nimboran (in detail Section 4 above) is a surprising insight, since today the two languages are farthest away from each other on the east-west axis. Socially, the Kilmeri people don't seem to have any ties that far west; some clan relations across the state border only exist to Manem and maybe Taikat speaking clans.¹⁴

Muysken (2010:272) describes the scenario of borrowing generally as asymmetrical from a dominant superstrate to a socially subordinate language; Winford (2010:177) sees this (a)symmetry relation as a tendency. Because of the very low numbers of borrowing in the present context one should be cautious to draw inferences about social hierarchies between the peoples concerned. Extra-linguistic sources of former social hierarchies between the languages in question that may support possible dominance are not available. Today, however, the Kilmeri are not bilingual in the contiguous vernacular languages I'saka and Pagi; the eastern Pagi villages (Imbio, Imbinis) are looked down upon by them. The recent Kilmeri people are clearly the dominant group in the proximate area. In former times, this may have been the other way round vis-à-vis the Nimboran and Sentani in the west, from whom the Kilmeri borrowed some vocabulary. Presumably, the Kilmeri and Border people were "jungle-dwellers" who got in contact with "river-dwellers" (cf. Aikhenvald 2008:2, 14). The Sentani were clearly lake-dwellers with a fair amount of fish production; they may have traded fish for sago (cf. Cowan 1965:72–74).

Lexical transfer is usually said to be the outcome of bi-/multi-lingualism. Foley (2010:797) describes Papuan multilingualism as extensive in the whole

13 According to their numbers of loanwords, Tadmor (2009:57) classifies languages in "very high borrowers" (> 50%), "high borrowers" (25–50%), "average borrowers" (10–24%), and "low borrowers" (< 10%). The documented Kilmeri lexicon comprises roughly 800 words/stems; in the present study 19 instances of loans into Kilmeri are identified (13 words from Nimboran, 6 from Sentani). So, with 2,3% extra-family loans from vernacular languages, Kilmeri looks like a low borrower regarding those sources. In a similar magnitude, Ross identifies 1,7% loan words from Bargam into Takia, plus 0,6% from Waskia, which makes for 2,3% Papuan loans in sum (2009:758). Gasser also reports very low rates of loan involvement for a number of Papuan languages in her sample (2019:637).

14 My consultants were reluctant to touch this topic because of the tensions between the native Papuan population and the Indonesian military; there were OPM ("Organisasi Papua Merdeka", Organisation for a Free Papua) activities in the area including the Papua New Guinea side of the state border (see also Marmion 2010:31). The Kilmeri people's reservation towards this political subject continued to hold over the years. So I refrained from asking questions.

New Guinea region, but characterises it as a mainly male affair.¹⁵ Yet the marriage of women into another language group is common social behaviour and usually results in some degree of bilingualism, at least in the family and village contexts. Thus it is plausible to assume degrees of bi-/multi-lingualism for the speakers of the languages under investigation, though little can be said about its type. It may have been related to life stages and sex of the speakers. The imperfect second language acquisition is in line with adult bilingualism, insofar new phonological oppositions cannot be acquired any more (cf. Ross 2013:20). Instead, the new words are phonologically adapted (Kilmeri) and tonally integrated (Skou, Wutung).

The overall lexical transfer among the languages is low. This may be caused by the lack of long-lasting and frequent direct contacts. On the other hand, it could be indicative of language loyalty as a means of group identity, which may have played a major role in language attitude, especially since languages are often spoken by (very) small groups of speakers (Winford 2010:178; Foley 2010:796). For the Kilmeri clans and villages, their shared language is a firm pillar of their shared identity; this view was confirmed by all my language consultants.

8.2 *Structural Convergence?*

In the light of the very little lexical transfer among the languages in question one may ask about possible structural convergence regarding the (greater) area, in which these languages are spoken. The retainment of the vocabulary could then be interpreted as a general means of highlighting and preservation of group identity. A reference case for lexical divergence paired with high structural convergence are the languages of the Banks and Torres Islands in north Vanuatu, described by Francois (2011). Grammatically, the 17 languages spoken there (some are moribund) build a linguistic area; they show almost perfect intertranslatability based on identical word order and (almost) identical grammatical categories (2011:178; 214). Clearly, the area of the Border, Sko, Sentani, and Nimboran families cannot be regarded as a structural convergence zone like northern Vanuatu. However, what we do find, is selective convergence regarding some special grammatical features among a few languages from two or more families.

15 A perfect instance of indigenous multilingualism is the following: Considering the Papuan loanwords found in the Oceanic language Takia of the Bel family (Karkar island, Madang province), Ross assumes that the Takia speaking people used to be bilingual in their language and coastal mainland Bargam and maybe more languages (Ross 2009:764).

Structural isomorphism in the lexicon Kilmeri and Skou show surprising similarity in their kind-referring terms, which are usually composite words consisting of a generic term and a specific term. The generic term indicates the class in accord with folk taxonomy, while the specific term adds the necessary distinction. Kilmeri lexically distinguishes twelve faunal classes and seven floral classes (Gerstner-Link 2018:644–659), whose members comprise different kinds numbering between 64 (trees/shrubs), 34 (birds and bats) and three (yams; blood sucking insects; caterpillars). As for Skou, the vocabulary lists in Donohue (2004; 2002) offer the following easily recognisable classes: (i) animals moving in the air, (ii) animals moving in water, (iii) animals with fur, and (iv) snakes. I illustrate this structural lexical isomorphy for the class of animals moving in the air with a few examples: ‘hornbill’ S *tángung* and K *iwán*, ‘pigeon’ S *tángángue* and K *imalo*, ‘heron’ S *tángpa* and K *iwai*, ‘lorikeet, parakeet’ S *tánglè* and K *ipumiya*, ‘small bat’ S *tángkengkeng* and K *imero*.¹⁶

No other Border language shows this pattern for the faunal and floral domain of their lexicons; the Sko languages Wutung (Marmion 2010:283–284), Tsaka (Donohue and San Roque 2004), Dusur (Ross 1980:101–105), and Barupu (Corris 2005) illustrate it to a certain degree. The Kilmeri pattern of kind-referring terms is a structural innovation due to transfer from Skou.

Phonological isomorphy Kilmeri and Skou make the phonological distinction between /l/ vs. /r/, while the other members of both families have only one liquid. In Kilmeri, /t,d/ changed to /r/ (see Appendix); in Skou, *t became /r/. Donohue says that the development of /r/ in Skou must have been due to areal pressure (2002:192; 200). The change from /t,d/ to /r/ is not only observable in Kilmeri, but also in eastern Sentani: *d changed to r word-initially, while *t became r intervocally between central or back vowels (Gregerson and Hartzler 1987:10–11). Foley describes Sentani’s consonant inventory as employing both liquids /l/ and /r/ (2018:439). Farther west, Berik, a member of the Tor family, also distinguishes /r/ and /l/ (Foley 2018:472).

Pronoun system In the Border, Nimboran, and Tor families as well as in the Kaure family to the south we have pronoun systems that distinguish only four categories: first person, second person, third person, and inclusive.

16 Note that the folk taxonomic class membership of specific kinds is semantically not isomorphic, but language-specific for Kilmeri and Skou.

There is no number distinction (Anceaux 1965:167; Foley 2018:470–471; 456). The Sentani and Sko families, on the other hand, employ number distinctions: Sentani distinguishes singular and plural forms (Cowan 1965:16), while Skou distinguishes three numbers and even adds gender distinctions in the dual and third singular (Donohue 2004:186). In contrast to Waris and Imonda, the current pronoun system of Kilmeri has singular, dual, and plural forms and consists of eleven different forms (Gerstner-Link 2018:109; 111). The dual forms are transparently bimorphemic forms that add a locative suffix, which is also used to build pairs of people referred to by proper names (2018:238). The plural forms are more opaque and less easily analysable, but have certainly a bimorphemic history. Kilmeri second plural *inε* may go back to the Border stem *ind* ‘person, man’ plus *dε* ‘you’, resulting in *inε* literally meaning ‘you person’. Note that the plural ‘they’ is often expressed by *jena* ‘people’, which is presumably cognate with *ind*. It is a plausible hypothesis that we observe structural transfer in the current Kilmeri pronoun system, under the older influence of Sentani (plural) and the newer influence of Skou (dual).

Dative verbs Kilmeri possesses 13 verbs with obligatory recipient/dative agreement (Gerstner-Link 2018:386–387) and Nimboran proper possesses 11 such verbs (May 1997:86–88). In this agreement class, they share three common dative verbs (‘tell sb’, ‘show sb’, ‘give sb’), but they also share five verbs with meanings that are not commonly dative verbs: ‘ask sb’, ‘answer sb’, ‘gossip about sb, call sb names’, ‘wait for sb, meet sb’, ‘share food with sb’. In view of the fact that, like the other (documented) Border languages, Kilmeri is predominantly a language with number agreement (2018:323–385), this convergence of role-based person agreement illustrates constructional isomorphy (Francois 2011:212) and may well be due to contact and mutual transfer.

Conclusion: The little lexical transfer among the language families under investigation does not correlate with a high structural convergence via transfer of grammatical properties. However, the transfer of categories in the pronoun system shows that the overall system can be modified under contact influence, while the formal substance of pronouns is indeed quite resistant to borrowing (cf. Tadmor et al. 2010:233). The Border family is the only one which participates in all of the above patterns of argued convergence. This hints at a complex contact scenario over time, i.e., to a series of successive contact events.

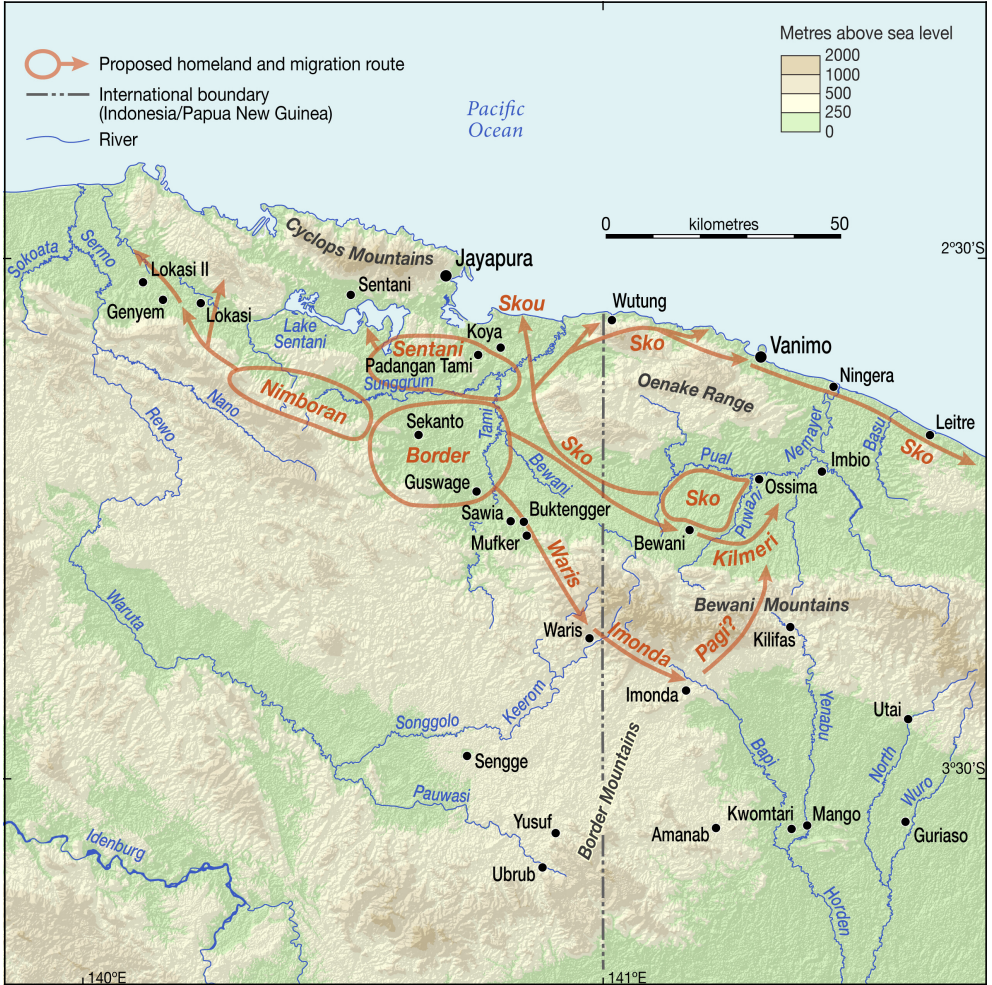


FIGURE 8.2 Putative migration routes of the Kilmeri, Nimboran, Sentani, and Sko people

Comment on the map: The proposed migration routes shown in the map are not exhaustive in the sense that they are not meant to comprise all migrations of the Border people. Quite probably, many more movements away from the proposed homeland took place over time, especially to the south. The migration route of the Pagi is hypothetical. Furthermore, some clans of the Sko people may have gone east or north directly.

8.3 *Traces of Contact and Migration Patterns*

Language contact among vernaculars presupposes vicinity or even contiguity of the languages concerned. Hence, we need to assume that clans speaking

Kilmeri and clans speaking Nimboran and Sentani settled in the same area during a certain time span in the past. The location of their more or less adjacent homelands and hunting grounds may have been in the northern part of the area which is now assigned to Elseng on language maps (see Introduction, Figure 8.1). I hypothesise that the middle Tami river area is the place from where the Border languages spread southeast and east. This hypothesis is supported as follows. For Waris and Imonda there are oral accounts of their origin west of their current sites (Section 2 above). For Kilmeri we have linguistic data that put them in contact with the speakers of Nimboran and Sentani who nowadays live in a region (more than) 100 kilometers further west. In addition, we have the oral source of the clan genealogy over ten generations provided by my Kilmeri consultant Margaret Osi, who was married to the late clan leader Lis Osi and possesses remarkable knowledge of the clans' past. The genealogy dates back the arrival of their ancestor Si in the Puwani-Pual basin to about 200–250 years ago, with Lis Osi's lifetime as reference point (Gerstner-Link 2018:16–20). Assuming this oral account is historically reliable we get roughly 1800AD as the date *ante quem* of contact between speakers of Kilmeri and Nimboran/Sentani. At the same time, about 1800AD is also the date *post quem* at which the Kilmeri got in contact with the Sko speaking people. According to Donohue, the modern Skou trace their ancestors to the mountainous area to the south-east, that is, the western Oenake range. He assumes that Proto Macro-Sko speakers had lived in the Puwani-Pual basin before the intrusion of people speaking (one of) the languages of the Bewani branch of the Border family (Donohue 2004:5–6).

This migration pattern correlates with the relative chronology of external borrowing that can be ascertained based on sound correspondences and sound changes within the Border family. The contact between Border/Kilmeri speaking people and the Nimboran people must have been prior to the regular sound change from Waris / t,d/ > Kilmeri /r/, which is attested by a number of cognate pairs (see Appendix). The Nimboran forms show the same phonological pattern as the languages of the Waris branch, while Kilmeri is different: 'child' is *du* in Nimboran and *tuendis* in Waris, but *ruri* in Kilmeri; 'tongue; mouth' is *méndu* in Nimboran and *minde* in Waris, but *ber* in Kilmeri; the *wanderwort* 'tree' is *di* in Nimboran and *ti* in Waris, but *ri* in Kilmeri. The same sound correspondence applies for the transfer of the distal deictic from Sentani into the Border languages. Sentani and Tabla (*n*)*di*- occurs as *dí* in Waris, but as *ri*- in Kilmeri. Therefore this contact is also old.

Turning to Skou and Kilmeri, we see that they both contrast with Waris: 'empty; hole, hollow' is *bí* in Skou and *bí* in Kilmeri, but *mɛ* in Waris; 'burn' is *rà* *li* in Skou and *rɛ* in Kilmeri, but *ta* in Waris. This means that Skou borrowed the words after the sound changes took place that we observe between the Waris

branch and Kilmeri, viz., Waris /t,d/ > Kilmeri /r/ and Waris /m/ > Kilmeri /b/ in syllable-initial position (see Appendix). Thus the contact is younger.

The above scenario is compatible with the eastward movement of the Kilmeri speakers in the past. From the greater Tami river area they (slowly) migrated to the east through the wide Bewani river valley, which connects to the Puwani-Pual basin. Somewhere during their journey they encountered the Skou people who had been forced or were then forced to leave their sites; when and where exactly this happened cannot be reconstructed. Only the people speaking I'saka retained their sites on the easternmost hills of the Oenake range, where they are the traditional and undisputed landowners (p.c. Simon Tapi of Krisa).¹⁷ The fact that I'saka is a first-order split from Macro-Skou is in line with an old and stable settlement.

However, the Kilmeri also came upon the speakers of Pagi. So far no oral accounts have emerged of the Pagi speakers' clan history, former dwelling sites, or migrations. Linguistically, the sound correspondences between Waris, Kilmeri, and Pagi suggest that Kilmeri and Pagi underwent different phonological developments: for instance, Waris /t,d/ corresponds to Kilmeri /r/, but to Pagi /k/. This regular triple correspondence (see Appendix) can only be understood, if one goes back to Proto Border and tries to reconstruct a proto phoneme that governs all three language-specific developments. A good candidate would be *t. Then the Waris branch of the Border languages would be the conservative branch that retained the alveolar plosive, while Kilmeri and Pagi show independent innovations. Yet in other environments, Pagi still shows an old /t/ that corresponds to Kilmeri /l/ (Gerstner-Link 2018:31–35). The arrival of the Pagi in the Puwani-Pual basin probably predates that of the Kilmeri, since nowadays the Kilmeri live on better land while the Western and Eastern Pagi are found in minor, much more swampy places around Bewani in the west and Imbio/Imbinis in the east. This suggests land grabbing by the Kilmeri. The Pagi may have come from the south, thereby crossing the Bewani mountains, which must have been possible. The official map of the area shows two foot track routes from the Bapi valley to Bewani; there is also a foot track from Ossima to Kilifas (Jeffrey Osi, p.c.).

What exactly caused the Kilmeri to turn east in search for new dwelling and hunting sites can only be guessed. Whenever my consultant Margaret Osi and I talked about game and hunting, she raved about the golden hunting opportunities in earlier times, when her ancestors had arrived in Ossima and its vicin-

17 According to Donohue (2004:6), the modern Skou people are faced with the lack of substantial, undisputed land holdings, and with ongoing disputes about compensations for transmigration lands between the Skou and Elseng (2004:14).

ities. This might suggest that the Kilmeri people had been under economic-ecological pressure at their original places. It is known that, in the Upper Sepik region, over-hunting indeed caused people to move away in order to look for places providing better livelihood.¹⁸

9 Summary and Outlook

The linguistic history of regions for which written sources lack completely can at least be partially reconstructed. Comparison of the lexicon of several language families unveils non-inherited items that came to be shared by contact between their speakers. In addition to often attested types of meanings shifts among transferred words like contiguity and (visual) similarity of the designated items (Blank 1997), more abstract features known from noun categorisation devices (Aikhenvald 2000) could also be taken into account to uncover putative borrowings. The revealing of different types of structural transfer likewise points at some contact among languages. When these families and single languages are not located in proximity today, a history of migration is suggested whose relative chronology can be argued for by historical linguistics, viz., the discovery of sound changes that borrowed items have or have not undergone. In cases where such evidence is supported by oral tradition that tells of peoples' distant origin and land grabbing in their current area, migration is the most plausible scenario. The successive structural transfer into Kilmeri resulted in a grammatical hybridisation of this language acquiring several new properties, while the other Border languages retained the inherited structural properties in question (cf. Section 8.2). So the language transitioned from its original convergence cluster of a minimal system of four pronouns into the more widespread group of number distinctive pronoun systems. By contrast, despite its acquisition of person marking for a special verb class, Kilmeri continues to be a member of the verbal number cluster of the area (Gerstner-Link 2018:383–385; Foley 2018:488–490). In sum, the dynamics of language change by contact is low with regard to the four language families here. While the results are still preliminary, a first step is done in understanding their common history, but much more needs to be investigated.

18 In his introductory article presenting results of the Upper Sepik-Central New Guinea Project Craig (1980:9) writes: "Another tradition, reported in 1968 by informants at Bibiyun on the mid-August River, is that the Yimnai originally lived in the Simaiya valley, east of the Idam valley. They exhausted the supply of game—mainly wallaby—and moved west, near to present-day Bisiaburu on the Idam; part of that group then moved up the August River to present-day Bibiyun, and to Buliap on the Sepik within West Papua."

Acknowledgements

Many thanks to the organisers of the APLL 11 Conference in Leiden, June 2019, for inviting me, and to the participants for helpful comments. In particular I like to thank Timothy Asher, Don Daniels, Bethwyn Evans, Harald Hammarström, Marian Klamer, Godehard Link, and Peter-Arnold Mumm for their questions and discussions. Special thanks again to Marian Klamer and to an anonymous reviewer whose comments on drafts of the paper brought me additional insights that I didn't have before. All errors are my own.—Finally, I feel gratitude to my late father Dr. Wolfgang Erich Gerstner to whom I owe so much of my funding.

Appendix

Border Family: Putative Cognate Sets and Sound Changes for Waris (Waris Branch), Kilmeri (Bewani Branch), and Pagi (Bewani Branch)

The direction of the sound changes is not easy to determine. Some correspondences suggest sound change from Waris to Kilmeri, e.g., Waris /ⁿd,t/ > Kilmeri /r/. For this change and direction we have also areal support. More difficult are the correspondences Waris /ŋ/ < > Kilmeri /k/ and Waris /m,n/ < > Kilmeri /^mb,ⁿd/. Kilmeri lacks voiced velars, while all the other Boder languages possess these phonemes, so it is reasonable to assume that Kilmeri lost these sounds. But the issue of the voiced labials and alveolars is less clear. Universally, lenition is more frequently attested than fortition. Yet Kilmeri seems to show, in word-initial position, the development from nasals to prenasalised plosives, namely /m/ > /^mb / and /n/ > /ⁿd/, which sets it apart from the Waris branch of the Border family and also from Pagi. Actually, in current Kilmeri quite some words with initial /m/ show free variation with [m] and [^mb] like *musi* 'to shut' as [musi] and [^mbusi]. This supports word-initial occlusion.¹⁹ The observable sound changes occurred probably at different times under different phonological conditions and/or pressure. This can be exemplified as follows. Kilmeri /p/ has two Waris correspondences. In several instances Kilmeri /p/ also occurs as /p/ in Waris, e.g., 'water' K *pu* < > W *pɔ*, 'betelnut' K *puel* < > W *pul*, and 'diarrhoea' K *eper* < > W *epɔnda*. Here /p/ appears to be old. Yet in many other

19 The phonological change of occlusion is quite rare, but it systematically occurs in the Kaure [Nawa River] family (Timothy Asher, p.c.; see Introduction, Figure 8.1).

words it occurs as /β/ in Waris: 'do' K *pi* < > W *βε*-{β}, 'leaf' K *pεε* < > W *βεεε*, 'liquid of fruit' K *pul* < > W {*mo*}-β*ɔ*l, 'wind' K *pupi* < > W *βuβi*, 'wooden signal horn' K *pup* < > W *βuβ*, 'house' K *jip* < > W *deuβ*, 'earthquake' K *ninɔp* < > W *nεnβ*. Kilmeri has no labial fricatives; it probably lost them like the velar fricative, and so /p/ in these words is a newer development. The sound changes below also provide evidence that Kilmeri and Pagi underwent different sound changes with respect to Waris. Hypothesis: Waris represents an older stage of the Border family's sounds, while Kilmeri and Pagi show independent, different innovations.

The sound changes that the above discussion of lexical transfer refer to are summarised in the following table. In the column *Sound change*, the first row gives the sound change from Waris to Kilmeri and the second row the change from Waris to Pagi. In many cases, the sound changes are positionally constrained. Curly brackets indicate (morphological) elements that are not part of the compared pair.

Sound change	Meaning	Waris	Kilmeri	Pagi
t > r syl-initial	'tree'	<i>ti</i>	<i>ri</i>	<i>ki</i>
t > k syl-initial	'feather'	<i>tai</i>	<i>rε</i>	<i>kai</i>
	'wet'	<i>puti</i> -{l}	<i>puri</i>	
d > r syl-final, intervoc	'child'	<i>tuεnd</i> -{is}	<i>ruri</i>	<i>kɔkεi</i>
d > k syl-final, intervoc	'foot(print)'	<i>dand</i>	<i>dɔr</i>	<i>nɔk</i>
	'tongue'	<i>mindε</i>	<i>ber</i>	<i>mεki</i>
	'dog'	<i>windε</i>	<i>wɔr</i>	<i>wɔk</i>
	'netbag'	<i>wɔnda</i>	<i>urɔ</i>	<i>uk</i>
	'diarrhoea'	<i>εpɔnda</i>	<i>εper</i>	
	'penis gourd'	<i>pɛnda</i>	<i>ber</i>	
	'marsupial, possum'	<i>pind</i>	{ <i>bi</i> }_per	
	'flat'	<i>pund</i>	<i>pur</i>	
t > l syl-final	'sugarcane'	<i>atxa</i>	<i>εlɔ</i>	<i>æt^h</i>
t > t ^(h) syl-final	'leech'	<i>at</i>	<i>al</i>	<i>wat</i>
	'fish'		<i>wal</i>	<i>vit</i>
	'snake'		<i>pial</i>	
m > b syl-initial	'stone axe'	<i>mand</i>	<i>buar</i>	<i>mɔk</i>
m > m syl-initial	'tongue'	<i>mindε</i>	<i>ber</i>	<i>mεki</i>

(cont.)

Sound change	Meaning	Waris	Kilmeri	Pagi
	'leg, back limbs'	<i>mɔ</i> -{ɣala}	<i>bɔu</i>	<i>mɔu</i> -{l}
	'saliva'	<i>mius</i>	<i>bɪs</i> -{ɛp}	
	'pig'	<i>miɛ</i>	<i>bi</i>	
	'hole, hollow'	<i>mɛ</i> -{l}	<i>bɪ</i>	
	'sound, speech'	<i>mɔa</i>	<i>bɔ</i>	
	'in-house fireplace'	{ <i>jua</i> }- <i>masa</i>	<i>bæs</i>	
	'dead body'	<i>mind</i> -{il}	<i>bir</i>	
n > d syl-initial	'sago (swamp)'	<i>nə</i>	<i>duɛ</i>	<i>na</i>
n > n syl-initial	'bush'	<i>na</i>	<i>du</i>	<i>nɔ</i>
	'meat'	<i>nix</i>	<i>dɔ</i>	<i>ni</i> -{l}
	'grass skirt'	<i>nai</i>	<i>diɛ</i>	
	'eye'	<i>nɔβ</i>	<i>dɔb</i>	<i>nɔp</i> -{ɔl}
	'axe'		<i>dawa</i>	<i>nawa</i>
	'night'		<i>dupuni</i>	<i>nɔpuni</i>
ŋ > k syl-final, intervoc	'mouth'	<i>mɛŋ</i>	<i>mɛk</i>	
Waris > Pagi ???	'sago grubs'	<i>mɛŋɛ^{mb}</i>	<i>bɛkup</i>	
	'wife'	<i>ɔŋa</i> -{l}	<i>akɔ</i>	
	'to think'	<i>nɛŋ</i> -{β}	<i>nɛki</i>	
	'to lie flat; place of sth'	<i>liŋi</i> -{l-β}	<i>liki</i>	
	'underneath'	{ <i>demus</i> }- <i>siŋi</i>	<i>sikil</i> -{jɔ}	
k > k all positions	'bone'	<i>kəl</i>	<i>kili</i>	<i>ɛli</i>
k > Ø syl-initial	'mosquito'	<i>kles</i>	<i>kles</i>	<i>ɛɛs</i>
	pronoun 1SG	<i>ka</i>	<i>kɔ</i>	<i>a</i>
	'small frog'	<i>keu</i>	<i>k^wɛr</i>	<i>ɛtu</i>
	'chin'	<i>kisi-l</i>	<i>kæau</i>	
	'fish scales, fish bones'	<i>ku</i>	<i>kisi</i>	
	'headlouse'		<i>kɔ</i>	
x > Ø all positions	'thunder'	<i>xul</i>	<i>ul</i>	<i>aunɔi</i>
x > Ø all positions	'stomach'	<i>ɛxna</i> -{l}	<i>ɛni</i>	<i>ni</i> -{l}
	'meat'	<i>nix</i>	<i>dɔ</i>	

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers. A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2008. "Language Contact along the Sepik River, Papua New Guinea." *Anthropological Linguistics* 50:1, 1–66.
- Anceaux, Johannes C. 1965. *The Nimboran language. Phonology and Morphology*. 's Gravenhage: Martinus Nijhoff.
- Blank, Andreas 1997. *Prinzipien des lexikalischen Bedeutungswandels*. Tübingen: Niemeyer.
- Brown, Bob 1990. *Waris grammar sketch*. Ms. Ukarumpa: Summer Institute of Linguistics.
- Brown, Bob and Honoratus Wai 1986. *Diksenari. A short dictionary of the Walsa (Waris) language, Tok Pisin and English*. Ms. Ukarumpa: Summer Institute of Linguistics.
- Burung, Wiem 2000. A brief note on Elsen. Unpublished manuscript. SIL International Electronic Survey Reports 2000–001: Summer Institute of Linguistics International.
- Collier, Kenneth and Kenneth J. Gregerson 1985: "Tabla verb morphology." *Papers in New Guinea Linguistics* 22, 155–172.
- Corris, Miriam 2005. *A grammar of Barupu: a language of Papua New Guinea*. PhD Thesis, University of Sydney.
- Cowan, H.K.J. 1965. *Grammar of the Sentani language*. 's Gravenhage: Martinus Nijhoff.
- Craig, Barry 1980. "Introduction to the Legends of the Abau of Idam Valley and of the Amt of Simaiya Valley, West Sepik Province, Papua New Guinea." First published in *Oral History* 8, Nrs 4 and 5.
- Donohue, Mark 2002. "Which Sounds Change: Descent and Borrowing in the Skou Family." *Oceanic Linguistics* 41:1, 171–221.
- Donohue, Mark 2004. *A grammar of the Skou language of New Guinea*. Ms. published online via eSICDOC MPG. Accessed September 2020 and October 2021.
- Donohue, Mark and Melissa Crowther 2005. "Meeting in the middle: Interaction in North-Central New Guinea." In Andrew Pawley, Robert Attenborough, Jack Golson, and Robin Hide (eds.), *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Australian National University, Research School of Pacific and Asian Studies. 15–65.
- Donohue, Mark and Lila San Roque 2004. *Isaka. A sketch grammar of a language of North-Central New Guinea*. Canberra: Pacific Linguistics 554.
- Foley, William 2010. "Language Contact in the New Guinea Region." In Raymond Hickey (ed.), *The Handbook of Language Contact*, Oxford: Wiley-Blackwell. 795–813.
- Foley, William A. 2018. "The languages of Northwest New Guinea." In Bill Palmer (ed.), *The languages and linguistics of the New Guinea area*. Berlin/Boston: De Gruyter Mouton. 433–567.

- Gasser, Emily 2019: "Borrowed Color and Flora/Fauna Terminology in Northwest New Guinea." *Journal of Language Contact* 12, 609–659.
- Gell, Alfred 1992. "Barter and gift exchange in old Melanesia." In Caroline Humphrey and Stephen Hugh-Jones (eds.), *Barter, Exchange and Value. An Anthropological Approach*. Cambridge: Cambridge University Press.
- Gerstner-Link, Claudia 2019. "Loans into and from Kilmeri as indicators of the people's migration route." Paper presented at the 11th International Austronesian and Papuan Languages and Linguistics Conference Leiden.
- Gerstner-Link, Claudia 2020. "Elseng as a member of the Border family." Ms. University of Munich.
- Gerstner-Link, Claudia 2018. *A grammar of Kilmeri*. Pacific Linguistics 654. Berlin/Boston: De Gruyter Mouton.
- Gerstner-Link, Claudia. *A dictionary of Kilmeri*. Work in Progress. To appear in Martin Haspelmath and Barbara Stiebens (eds.), *Dictionaria*. Leipzig.
- Gerstner-Link, Claudia 2000. *Pagi fieldnotes*.
- Gregerson, Kenneth J. and Margaret Hartzler 1987. "Towards a reconstruction of Proto Tabla-Sentani phonology." *Oceanic Linguistics* 26:1/2, 1–29.
- Hartzler, Margaret 1976. "Central Sentani phonology." *Irian: Bulletin of Irian Jaya Development* 5:1, 66–81.
- Haspelmath, Martin 2009. "Lexical borrowing: Concepts and issues." In Martin Haspelmath and Uri Tadmor (eds.), *Loanwords in the World's Languages—A Comparative Handbook*. Berlin/New York: De Gruyter Mouton. 35–54.
- Haspelmath, Martin and Uri Tadmor 2009. "The Loanword Typology project and the World Loanword Database." In Martin Haspelmath and Uri Tadmor, *Loanwords in the World's Languages—A Comparative Handbook*. Berlin/New York: De Gruyter Mouton. 1–34.
- van der Heuvel, Wilco and Sebastian Fedden 2014. "Greater Awyu and Greater Ok: Inheritance or Contact?" *Oceanic Linguistics* 53:1, 1–36.
- Juillerat, Bernard (ed.) 1992. *Shooting the Sun. Ritual and Meaning in the West Sepik*. Washington / London: Smithsonian Institution Press.
- Juillerat, Bernard 1996. *Children of the Blood: Society, Reproduction and Cosmology in New Guinea*. Oxford: Berg Publishers.
- Kouwenhoven, Willem J.H. 1956: *Nimboran. A study of social change and social-economic development in a New Guinea society*. Doctoral Thesis. Leiden.
- Marmion, Douglas E. 2010. *Topics in the Phonology and Morphology of Wutung*. PhD dissertation. Canberra: Australian National University.
- Matras, Yaron 2009. *Language contact*. Cambridge: Cambridge University Press.
- May, Kevin R. 1997: *A Study of the Nimboran Language*. MA thesis. Melbourne: La Trobe University.
- Menanti, Jackie 2005. *Sociolinguistic report on the Elseng language in Sia-Sia village, Keerom county, Papua, Indonesia*. Unpublished ms. Ukarumpa: Summer Institute of Linguistics.

- Miller, S.A. 2017. "Skou Languages Near Sissano Lagoon, Papua New Guinea." *Language and Linguistics in Melanesia* 35, 1–24.
- Minch, Andrew 1992. "Amanab grammar essentials." *Data Papers on Papua New Guinea Languages* 39, 100–173. Ukarumpa: Summer Institute of Linguistics.
- Muysken, Pieter 2010. "Scenarios for Language Contact." In Raymond Hickey (ed.), *The Handbook of Language Contact*. Oxford: Wiley-Blackwell. 266–281.
- Ross, Malcolm 1980. "Some Elements of Vanimo, a New Guinea Tone Language." *Papers in New Guinea Linguistics No. 20*. Pacific Linguistics A—56, 77–109.
- Ross, Malcolm 2009. "Loanwords in Takia, an Oceanic language of Papua New Guinea." In Martin Haspelmath and Uri Tadmor (eds.), *Loanwords in the World's Languages—A Comparative Handbook*. Berlin/New York: De Gruyter Mouton. 747–770.
- 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* 4:1, 84–132.
- Seiler, Walter 1985: *Imonda, a Papuan Language*. Pacific Linguistics B-93. Canberra: Australian National University.
- Smits, Leo and Clemens L. Voorhoeve (eds.) 1994. *The J.C. Anceaux Collection of Wordlists of Irian Jaya Languages B: Non-Austronesian (Papuan) Languages*. Part II. Irian Jaya Source Material 10 Series B 4. Leiden-Jakarta: DSALCUL/IRIS.
- Tadmor, Uri 2009: "Loanwords in the world's languages: Findings and results." In Martin Haspelmath and Uri Tadmor (eds.), *Loanwords in the World's Languages—A Comparative Handbook*. Berlin/New York: De Gruyter Mouton. 55–75.
- Tadmor, Uri, Martin Haspelmath and Bradley Taylor 2010: "Borrowability and the notion of basic vocabulary." *Diachronica* 72:2, 226–246.
- van der Wilden, Jaap and Jelly van der Wilden 1975: "Kemtuk phonology." *Irian: Bulletin of Irian Jaya Development* 4:3, 31–60.
- van der Wilden Jaap 1976: "Simplicity and detail in Kemtuk predication." *Irian: Bulletin of Irian Jaya Development* 5:2, 59–84.
- Voorhoeve, Clemens L. 1971: "Miscellaneous notes on languages in West Irian." *Papers in New Guinea Linguistics* 14, 47–114.
- Voorhoeve, Clemens L. 1975: *Languages of Irian Jaya Checklist: Preliminary Classification, Language Maps, Wordlists*. Pacific Linguistics. Canberra: Australian National University.
- Winford, Donald 2003. "Contact and Borrowing." In Raymond Hickey (ed.), *The Handbook of Language Contact*. Oxford: Wiley-Blackwell. 170–187.
- Wirz, Paul 1934 [1928]. "Beitrag zur Ethnologie der Sentanier (Holländisch Neuguinea)." *Nova Guinea* 16. Brill.
- Wohlgemuth, Jan 2009. *A Typology of Verbal Borrowings*. Berlin/New York: De Gruyter Mouton.

Spanish Suffixes in Tagalog: The Case of Common Nouns

Ekaterina Baklanova and Kate Bellamy

1 Introduction

The intense contact that took place between Spanish and Tagalog during Spanish colonial rule in the Philippine archipelago from the mid-16th until the turn of the 20th century was not characterized by widespread bilingualism (e.g. Lipski et al., 1996: 272–275; Thompson, 2003: 17). However, it did lead to heavy lexical borrowing,¹ which has resulted in significant changes to Tagalog derivation (see notably López, 1965; Goulet, 1971; Rau, 1992; Alcántara y Antonio, 1999; Steinkrüger, 2008; Potet, 2016). Less attention has been paid to morphological borrowing from Spanish, such as the adoption of several Spanish nominative and adjectival affixes, which constitute mostly suffixes (Wolff, 1973, 2001; Baklanova, 2004, 2017; Quilis and Casado-Fresnillo, 2008). This chapter will address the characteristics and impact of Spanish noun-forming suffixes in Tagalog, using the framework of Seifart (2015) to identify whether these constitute direct or indirect borrowings.

1.1 *On the Traces of Spanish in the Tagalog Lexicon*

Of the several dialects of Spanish present in the Iberian Peninsula in the 16th century, Castilian Spanish dominated in most administrative centers of the American colonies of Spain, including Mexico, “since most officials of the Crown came from this area, in particular from Toledo and Madrid” (Gómez Rendón, 2008, 1: 126). As the Philippine colony was under the jurisdiction of the Vice-royalty of New Spain established in Acapulco in 1535, Mexican Spanish and thus also Castilian might well have been the main variants of Spanish that

1 Following Thomason and Kaufman (1988: 37), we shall use “borrowing” as the traditional cover term for both lexical and structural linguistic items transferred into the recipient language, as well as the process of this transfer. The term “loanword” will be used, as in Haspelmath (2009: 36), only for “a word that at some point in the history of a language entered its lexicon as a result of borrowing (or transfer, or copying)”.

influenced Tagalog.² Philippine contacts with Spain were initially mostly limited to galleon trade via Mexico, since only from the 19th century and the independence of Mexico onwards were the Philippines, as other Spanish Pacific territories, administered directly from Spain (Sippola, 2020: 455; see also Quilis and Casado-Fresnillo, 2008).

Lipski et al. (1996: 272–275) show that there was no significant group of Spanish mestizos in the Philippines at this time, nor a large Tagalog-Spanish bilingual community. By the end of the Spanish rule “the census indicated that less than three percent of the population spoke Spanish” (Thompson, 2003: 16). Sippola (2020: 455) elaborates:

Local laws and customs were largely maintained, although the legal code was codified in Spanish. For most of the Spanish period, the policy was for priests to interact with Filipinos in the local vernaculars rather than teach Spanish, and Spanish education was limited mostly to a small elite.

With the advent of transoceanic steam navigation in the second half of the 19th century, increased trade with the Philippines “created a new wealthy class of Chinese mestizos who controlled commerce throughout the islands. They eagerly learned Spanish and spread it throughout the Philippines along with their business interests” (Thompson, 2003: 16). These bilinguals might, then, have become the main agents of the spread of Spanish language influence to Tagalog speakers from lower social strata.

Overall, the language situation in Manila and other Tagalog-speaking regions appeared to roughly correspond to diglossia (Fishman, 1967), where the High language (in this case Spanish) operated as the written/formal-spoken code and the Low language (Tagalog) as the vernacular, with no interaction between the two. The cases of Spanish-Quechua and Spanish-Otomí contact also indicate that in a diglossic situation where speakers of the Low language are socio-politically subdominant to speakers of the High language, the latter typically becomes a source of active borrowing into the former (Bakker and Hekking, 2012; Gómez Rendón, 2008). This is similar to the Philippine case: Spanish was a marker of high social status (Wolff, 2001: 234; Quilis and Casado-Fresnillo, 2008: 62–66). Hence, more than three centuries of influence by Spanish as a high prestige language of the colonial administration and local elite, even without a significant degree of bilingualism, has resulted in heavy lexical borrowing into

² Loanwords of both Indo-American and Spanish origin adopted by Tagalog via Spanish are considered hispanisms and marked as Mexican Spanish (Mex Spanish).

Tagalog. According to various estimates, Tagalog vocabulary consists of around 20% Spanish borrowings (Baklanova, 2017: 333–334), or even up to 32% (Rau, 1992: 101), with loanwords appearing in all domains (Wolff, 2001). Spanish influence on Tagalog rates at least as the third stage (“more intense contact”) on Thomason and Kaufman’s (1988) scale, where notably basic vocabulary is borrowed, including function words and discourse markers; new phonemes are added to the Tagalog inventory; and also derivational morphemes from Spanish are borrowed (see also Wolff, 1973, 2001; Baklanova, 2004, 2017; Quilis and Casado-Fresnillo, 2008; Steinkrüger, 2008).

The case of Tagalog is particularly interesting as the Spanish influence overlaps with English influence. Even after the replacement of Spanish rule by that of the USA in 1898, Spanish remained the second official language of the Philippines alongside English, and dominated in the courts and high society until the early 1930s (Lipski et al., 1996: 272; Thompson, 2003: 63). Already around 1920, society had seen an increase in the number of educated Filipinos who could speak English, often, however, with a Spanish-like accent (Fernández, 2013: 369).

We assume that a certain Spanish adstrate influence still persists in Tagalog through the following processes: 1) mildly productive nominal and adjectival derivation with Spanish affixes; 2) the development of a marginal gender system, as discussed in Stolz (2012) and Baklanova (2016); and 3) the “hispanization” of English borrowed lexical items.³ The third phenomenon needs some elaboration because examples of it are sometimes regarded simply as “mistakes” in the everyday speech of Filipinos (see, e.g., Alcántara y Antonio, 1999; Ortograpiyang Pambansa, 2013). It is highly probable that very few, if any, English words were borrowed into Tagalog via Spanish during the Spanish rule. Except for some culturally-specific borrowings, numerous English words began to enter the Spanish lexicon only from the 1950s onwards (Dworkin, 2012: 217–218). Examples of some early borrowings from English that had entered Spanish by the end of the 19th century, whence they were then borrowed into Tagalog are: Spanish *bistec* > Tagalog *bístik* ‘beef steak’, Spanish *cheque* > Tagalog *tséke* ‘check’, Spanish *turista* > Tagalog *turísta* ‘tourist’ (Dworkin, 2012:215). In the present study the immediate donor language of a loanword is taken as the source of the borrowing, thus the above examples are also considered hispanisms in Tagalog.

3 With thanks to Dr. Anthony Grant (p.c. Oct. 2020) for sharing a similar view on the adstrate character of Spanish influence on Tagalog.

Following Haugen (1969), Aikhenvald (2012: 178) observes, that “grammatical and lexical morphemes may not be borrowed directly, and yet come to share their form and meaning with a morpheme in the contact language”. In the case of Tagalog, the tendency to create neologisms through analogy with Spanish loanwords has been attested since the 20th century, along with the reshaping of English loanwords into Spanish-like forms as a means of accommodation (Goulet, 1971; Wolff, 1973, 2001). This pattern is similar to the way in which English loanwords are adopted into Indonesian based on an earlier way of borrowing Dutch words (Tadmor, 2009).⁴ General hispanization patterns of English borrowings in Tagalog are presented in Baklanova (2017: 336–337), two of which are reproduced in (1a–b).

- (1) a. English *-er* > Tagalog *-ero*: English *abus-er* > Tagalog *abus-éro* (cf. Spanish *abusador*)
 b. English *-ist* > Tagalog *-ista*: English *cartoon-ist* > Tagalog *kartun-ísta* (cf. Spanish *caricaturista*)

Perhaps surprisingly, this tendency in Tagalog developed independently of a similar mode of adopting Anglicisms and the creation of English-Spanish hybrid neologisms in Spanish, which has only been attested since the second half of the 20th century, such as English *adherence* > Spanish *adherencia* (Dworkin, 2012: 220–224). This process increases the frequency of Spanish and Spanish-like grammatical items in Tagalog discourse, which may foster the use of Spanish borrowed suffixes in Tagalog word formation.

1.2 *Aims and Methodology of the Present Study*

The present study investigates the borrowing of the Spanish agentive suffixes *-ero/a*, and *-ista*, the diminutives *-illo/a*, *-ito/a*, and *-ete*, and the adjectival *-eño* into Tagalog nominal derivation. The focus will be their impact on the contemporary derivation of common nouns.

Winford (2003b: 134) observes that “certain structural innovations in an RL [recipient language] appear to be mediated by lexical borrowing”, i.e. adopted through *indirect* borrowing. Cases of *direct* borrowing of structural elements typically involve free morphemes, while bound morphemes “appear to

4 Tadmor (2009: 702) describes the integration pattern of English loanwords as “based on an earlier pattern of borrowing similar Dutch words ending in *-atie* [asi] and *-isatie* [isasi]”: Dutch *proclamatie* ‘proclamation’ > Indonesian *proklamasi*. Hence English *-(iz)ation* is reshaped into *-(is)asi*: English *stagflation* > Indonesian *stagflasi*.

be borrowed only in cases where they substitute for RL morphemes that are semantically and structurally congruent. Moreover, such borrowing requires a high degree of bilingualism among individual speakers” (ibid.). Seifart (2015: 511) defines *indirect affix borrowing* as follows:

This scenario involves two subprocesses. First, a language borrows a number of complex loanwords containing an affix, and second—possibly much later—these complex loanwords come to be analyzed within the recipient language, and eventually the affix becomes productively used on native stems.

The scenario of *direct affix borrowing* (Seifart, 2015: 512) occurs when:

An affix is recognized by speakers of the recipient language in their knowledge of the donor language and used on native stems as soon as it is borrowed, with no intermediate phase of occurring only in complex loanwords.

Thus, Seifart’s (2015) definitions corroborate those of Winford (2003b), including the observation that direct borrowing requires a significant degree of bilingualism among speakers of the RL. However, such borrowing does not necessarily imply “full familiarity with the donor language”, or source language (SL; Seifart, 2015: 512). Moreover, the distribution of borrowed affixes and the ratio of corresponding *complex* (with the borrowed affix) and *simplex* (without the borrowed affix) loanwords in a corpus can be used to assess whether borrowing has been direct or indirect (ibid.). This also supports the observation that complex loanwords of low token frequency relative to corresponding simplex forms tend to be decomposed and analyzed by RL speakers more easily (Hay, 2001; Baayen, 2008). The analogically deduced affix may then be used to produce *hybrid formations* with the RL stems. According to Seifart (2017: 394):

[an affix] is considered effectively borrowed only if it is used with at least some native stems, i.e. it is not considered borrowed if it only combines with equally borrowed stems to form complex loanwords.

However, Tagalog hybrid formations with Spanish affixes may also be derived from borrowed stems, adopted from Spanish or another donor language, such as English (see Appendix, Table 9.13). If a stem has been borrowed into Tagalog from a source language other than Spanish, we consider its hybridization

with an affix of Spanish origin as evidence of the productiveness of this affix in Tagalog.

The crucial condition for primarily *indirect* affix borrowing is the presence of complex loanwords with this affix in the RL, while certain proficiency in the SL is necessary for *direct* affix borrowing (Seifart, 2015: 513–515). Based on this methodology and classification, we will identify the primary character of borrowing of the above-mentioned Spanish suffixes into Tagalog. This will also entail an assessment of the distribution and ratio of these suffixes in the research data described in Section 1.3.

Our second goal is to investigate the semantics of the borrowed suffixes. As observed by many scholars, such as Aikhenvald (2007: 23), “a borrowed bound morpheme, reanalysed and reinterpreted, may acquire a quite different meaning in the target language”. Wolff (2001: 248) suggests that Tagalog semantic deviations from the Spanish original be analyzed, for they “reveal the extent to which Spanish concepts were not taken over but reinterpreted into a Filipino understanding of the world”.

Thus, the present study focuses on three major groups of research questions:

1. Are all of the above-mentioned Spanish suffixes attested in derivations of Tagalog native stems, thus producing hybrid formations? What are the characteristics of the Tagalog stems receiving these suffixes?
2. What are the characteristics of the borrowing process for each of these suffixes? First, is it predominantly direct or indirect borrowing, following Seifart (2015)? Second, is the adoption of each of these Spanish suffixes older, pertaining to the colonial period (i.e., when Spanish still persisted in the Philippines); or is it more recent, being dateable to the 20th century (thus without the influence of Spanish)?
3. What new meanings do the borrowed Spanish suffixes introduce into Tagalog nominal derivation, if any? What is the overall impact of the Spanish suffixes on the Tagalog derivation of common nouns?

1.3 *Research Data*

To address these questions, and also in view of the present-day English influence on Tagalog, two datasets have been employed for the analysis: (a) historical data from the 19th–early 20th century (i.e., before the spread of English-Tagalog bilingualism); and (b) contemporary data of the 20th–early 21st century (when English-Tagalog bilingualism is widespread).

The early data are difficult to obtain, so dataset (a) is rather limited, consisting of the available Spanish-Tagalog dictionaries (Laktaw, 1889; Calderón, 1915), 34 sample Tagalog texts of 20,500 tokens (Bloomfield, 1917: ch. 1), and six liter-

ary texts from 1906–1922 by six Tagalog writers (Project Gutenberg), comprising around 60,000 tokens in total. This dataset is only used to check whether derivatives with each of the above-mentioned Spanish suffixes may already be found in the pre-English Tagalog lexicon. If hybrid formations with one of these Spanish suffixes are found in the early sources, this indicates that the suffix was borrowed into Tagalog during the period of direct Spanish influence on Tagalog. However, the lack of Tagalog hybrids with a Spanish suffix in the early dataset is not sufficient evidence that the suffix was not borrowed in the Spanish period. Since dataset (a) comprises mostly written texts and is rather small, it may not reflect colloquial Tagalog use from that period, and in fact, innovations might already have emerged.

The main source is the contemporary dataset, which comprises two large Tagalog dictionaries (English, 1987; Rachkov, 2012), and the recent Tagalog Leipzig Corpus (Goldhahn et al., 2012), hereafter LC, which consists of around 20 million *tokens* (total number of words), and about 472,000 *types* (each word form counted once). It was compiled in 2012–2016 from more than 500 sources, predominantly from the leading Filipino e-dailies (Abante, Abante Tonite, PhilStar, Journal.com.ph) and Tagalog Wikipedia, but also from some Tagalog blogs, thus it partly reflects colloquial, contemporary Tagalog.

Both datasets were first searched for complex nominal formations containing the suffixes *-ero/a*, *-ista*, *-ito/a*, *-ilyo/a* (*-illo/a*), *-eryo* (*-eño*) and *-ete/a*. The lists of derivatives from datasets (a) and (b) with each suffix were then analyzed in terms of provenance (namely, a Spanish complex loanword or a Tagalog hybrid formation), type of stem, semantics, and distribution in the datasets.

The rest of this chapter is organized as follows. Section 2 gives a description of some relevant aspects of Tagalog nominal derivation. Section 3 presents an overview of the characteristics of the agentive suffixes *-ero/a*, *-ista* and the suffix *-eño* in Spanish, an analysis of their distribution in the Tagalog datasets (a) and (b), as well as their impact on Tagalog nominal derivation. In Section 4 the same analysis is carried out for the Spanish diminutive suffixes *-ito/a*, *-illo/a*, and *-ete/a* in Tagalog lexical derivation. A discussion of the mechanism of borrowing of each suffix, based on the methodology of Seifart (2015) is presented in Section 5, followed by some concluding remarks in Section 6. The Appendix presents the characteristics of all Tagalog hybrid formations with *-ero/a* found in the datasets.

2 Notes on Tagalog Lexical Derivation

Tagalog is a language of the Central Philippine group of the Austronesian family, whose standardized variant—Filipino—is the national language of the Republic of the Philippines. Tagalog is characterized typologically as agglutinative-synthetic, with a relative abundance of affixes and clear morpheme boundaries (Blust, 2013: 41; 355–356). As such, Tagalog possesses a large inventory of derivations. A stem may be derived into different lexical categories (Shkarban, 1995: 38–42; De Guzman, 1996: 312–315). Shkarban (2004: 319–320) claims that the major rules regulating the functioning of Tagalog affixes operate “at the level of semantic relations between root-morphemes and affixes” (see also Wolff, 1993).

Nominal derivation may involve verbal, nominal or adjectival stems, and may include affixation, reduplication, compounding, conversion with prosodic changes, or a combination of the above. Nouns are stem lexemes or derivatives that do not take the verbal inflections of voice and aspect, nor the adjectival affixes of degree. They also do not inflect for case or number.

The class of nouns includes as its most productive:

- i. Names of persons and objects
- ii. Abstract names of quality or situation
- iii. Places

In class (i) common nouns are distinguished from personal names by the particles with which they co-occur: *ang* for common nouns, and *si* for personal names, which become *ng/ni* and *sa/kay* in genitive/ergative and oblique constructions (Schachter and Otones, 1972: 93–96).⁵ The present paper focuses on common nouns in this first class, that is, names of persons and objects. For this class, the main native derivation strategies are presented in Table 9.1 (following Blake, 1925; Schachter and Otones, 1972; Rachkov, 1981; Shkarban, 1995; De Guzman, 1996).

With regard to the strategies presented in Table 9.1, a number of observations can be made. Firstly, prefixation clearly prevails over suffixation, as illustrated in examples (3a–3e).

5 As stress is phonemic in Tagalog, in all the Tagalog examples stressed vowels are marked with an accent /'/, and the voiceless glottal stop is represented orthographically as /ʔ/ in word-final position.

TABLE 9.1 Tagalog native derivation of the class 'names of persons and objects'

Derivation strategy*	Derivation type	Meaning
maŋ-r-V (w/prosodic change)	Prefix	'a regular/professional doer of V' 'a person prone to do V'
mag-r-V (w/prosodic change)	Prefix	'a regular/professional doer of V' 'a person prone to do V'
mag- + N	Prefix	'a pair of persons (rarely, objects) bearing the relation designated by the stem'
ka- + N/V/Adj	Prefix	'a person/object reciprocally associated with another'
R+N+ -(h)an	Two-syllable reduplication + suffix	'a person/object imitating what the stem designates' 'diminutive of an object' (n/prod?)
N + -(h)in	Suffix (n/prod.)	'a similarity subject'
N + <in>	Infixation (n/prod.)	'a similarity subject'
pala- + V	Prefix	'(a person) prone to do V'
taga- + V	Prefix	'a person charged to V', 'a regular doer of V'
taga- + N	Prefix	'a person born/living/working at the place designated by the stem'
N(-ŋ)+N	Compound	'a person/object designated by the compounded stems'

* Adj – adjectival stem, LNK – linker (ligature), N – nominal stem, n/prod – not productive, r – one-syllable reduplication, R – two-syllable reduplication of the stem, V – verbal stem

- (3) a. *mam-(b)ángká?* 'to sail by boat' > *mámamangká?* 'boatman'; *mag-la síng* 'to get drunk' > *maglalásing* 'drunkard'
 b. *mag-lólo* 'grandfather with a grandchild'
 c. *ka-palít* 'a substitute'
 d. *palá-káin* 'frequent eater (of)'
 e. *taga-báyan* 'city resident'; *taga-showbiz* 'person from showbusiness' (< English)

There is only one productive suffixal strategy, namely *R+N+ -(h)an*, see examples (4a–b).

- (4) a. *báhay-bahay-an* '1. toy house; 2. small house'
 b. *bulág-bulág-an* 'person pretending to be blind'

The presently unproductive suffixal strategy $N + -(h)in$ seems connected to the infixal $N + \langle in \rangle$ of roughly the same meaning, namely 'a similarity subject', as in (5a–d).

- (5) a. *wika-ín* 'dialect, i.e. like-a-language'
 b. *k-in-arayom* 'long thin rice, i.e. like-a-needle'
 c. *ama-ín* 'uncle, i.e. as close as father'
 d. *k-in-ákapatíd* 'person as close as brother'

Tagalog lacks its own affixal inventory to derive agent nouns of the semantic group 'a doer of N/A' from non-verbal stems. For the diachronically polysemantic strategy $R+N+-(h)an$ (i.e. two-syllable reduplication of a nominal stem plus suffix $-(h)an$), the contemporary corpus data attest only word types meaning 'a person/object imitating what the stem designates', with no new diminutive types found. There is no evidence for the present-day productivity of this pattern. Diachronically, Tagalog derived a number of diminutive names of objects, some meaning both 'a small object' and 'an imitation of the object'. These occurred with both native (6a) and borrowed (6b) stems.

- (6) a. *ílog* 'river'—*ílug-ilúgan* 'rivulet, small river'
 b. *báso* (< Spanish *vaso*) 'glass'—*básu-basúhan* 'small glass; toy glass'

In the next section we shall discuss further how the suffixes borrowed from Spanish have contributed to Tagalog nominal derivation.

3 Spanish Suffixes in Tagalog Derivation of Agentive Nouns

Spanish is a fusional language, that is, its morphemes can simultaneously encode several meanings (Payne, 1997: 28). Most words contain more than one morpheme, and morpheme boundaries can be difficult to identify (Gómez-Rendón, 2008, 1: 156; Rainer, 2011). Spanish also has grammatical gender, so many of its nominal and adjectival suffixes are marked with the masculine or feminine exponents $-o/-a$, including $-ero/a$, $-illo/a$, $-ito/a$, $-eño/a$ (Gramática: § 2).

Due to heavy borrowing from Spanish, a wide variety of simplex-complex pairs and groups of Spanish loanwords have been adopted into the Tagalog lex-

icon, such *káha* ‘box’, *kah-íta* ‘small box’, *kah-éro* ‘cashier’ (< *caja* ‘box’); and *espírito* ‘spirit’ (identical in Spanish), *espirit-ísta* ‘spiritualist’, *espirítu-ál* ‘spiritual’. Evidently this process has enabled Tagalog speakers to contrastively analyze the semantic and structural differences between simplex and complex loanwords formed with the same stem. As a result, the hypothetical semantics of the suffixes *-ero/a*, *-ista*, *-eño/a*, *-ito/a*, *-illo/a*, *-ete* might have been acquired and eventually transferred to native noun formations.

3.1 Spanish Suffix *-ero/a*

As Muysken (2012: 485) observes, Spanish agentive suffixes such as *-ero/a*, *-dor/a* “almost operate in paradigmatic opposition [with] a series of related meanings <...>: profession, typical behavior, personal propensity, remarkable physical characteristic, resemblance, affective negative, pejorative, affective positive, endearment, diminutive”. The suffix *-ero/a* combines mostly with nominal and adjectival stems, and derives both nouns and adjectives (Gramática: 5.1.b). In nominal derivation it forms mostly agentive nouns with the meanings ‘a person of a profession/occupation related to N’, where N is mostly ‘an object of action’ (7a) or ‘a place of action’ (7b).

- (7) a. *reloj* ‘watch/clock’—*relojero* ‘watch/clock-maker’
 b. *taquilla* ‘box-office’—*taquillero* ‘box-office clerk’

It can also refer to ‘a person of a certain propensity related to N’, in diachrony often with a negative (deprecatory) connotation, as in (8a–b).

- (8) a. *aventura* ‘adventure’—*aventurero* ‘adventurer, prone to adventures’
 b. *política* ‘politics’—*politiquero* ‘political manoeuvrer (neg.)’

Moreover, it can also refer to nouns of objects meaning ‘place’, ‘container’, ‘instrument/utensil’, ‘group/set’, ‘tree/plant’ (ibid.: § 6.8i–6.8m, 6.8s).

In the historical dataset used in the present study, 25 types of Spanish complex loanwords (CL) with *-ero/a* and eight hybrid formations (HF), i.e. Tagalog neologisms with *-ero*, are attested (see Table 9.2).⁶

6 As most of the Tagalog stem words cannot be attributed to a concrete class outside of their context, for the purposes of the present analysis we shall take nominal stems as roughly referring to a person/object/place, adjectival stems as referring to a quality/trait, and verbal stems as referring to an action.

TABLE 9.2 Characteristics of nouns with *-ero/a* in the Tagalog dataset of 1900s

Semantic group of derivates	Type of stem	CLS	Simplex	HFS		Simplex
			related to CLS	Tagalog/ non-Spanish stem	Spanish stem	related to HFS
Object/place	N	7	2	-	-	-
Person of certain profession/occupation	N	18	11	2	1	3
	V	-	-	1	1	2
Person of certain	N	-	-	1 neg.	-	1
	V	-	-	2 neg.	-	2
TOTAL # of types		25	13	6	2	8

Examples (9a–b) contain two of the complex loanwords attested in the historical dataset.

- (9) a. Object/place: Spanish *candeléro* > Tagalog *kandeléro* ‘candelabrum’
 b. Person of certain profession/occupation: Spanish *fogonero* > Tagalog *pugonéro* ‘stoker’

Among the examples of the hybrid formations found in the dataset are those presented in (10a–b).

- (10) a. Person of profession/occupation: *sípa?* ‘kick with the boot; game with rattan ball’ > *sipéro* ‘sipa player’; *salamángka* ‘conjuring; magic; sleight of hand’ (< Spanish *salamanca* ‘cave for sorcery’) > *salamangkéro* ‘magician; juggler’
 b. Person of certain propensity/trait: *baság-úlo* ‘altercation; scuffle’ > *baság-uléro* ‘squabblers’ (neg.)

Moreover, among the entries of Calderón’s (1915) dictionary, there are around 50 more Spanish complex words along with some simplex-complex pairs, which do not appear in the dictionaries from the 1890–1900s. Yet these forms eventually entered the Tagalog lexicon, presumably not later than the early 20th century, while Spanish still had influence on Tagalog through its bilinguals (recall Thompson, 2003: 17, 63). The vast majority of these later Spanish complex loanwords also pertain to agentive nouns meaning ‘a person of profession/occupation’, but there are also a few meaning ‘a person of certain propensity’, mostly negative (11a), or referring to an object (11b).

- (11) a. Spanish *calle* ‘street’, *callejero* ‘loiterer, gadabout’ > Tagalog *kálye*, *kalye-héro*
 b. Spanish *grano* ‘grain’, *granero* ‘granary’ > Tagalog *gráno*, *granéro*

Based on the above data, several observations may be made regarding derivation with *-ero/a* in Tagalog at the beginning of the 20th century. First, the presence of hybrid formations with a Tagalog stem and the Spanish suffix *-ero* (including the masculine exponent *-o*) indicates that the suffix had been borrowed in that form into Tagalog not later than the turn of the 20th century. The form *-era* with the feminine exponent *-a* is not attested in the same dataset. Second, although in the early Spanish complex loanwords only the meaning ‘person of certain profession/occupation’ occurs, two of the three original meanings of Spanish *-ero* are registered in Tagalog hybrids, see (12a–b).

- (12) a. Person of certain profession/occupation: *bangká?* ‘boat’—*bangkéro* ‘boatman’
 b. Person of certain propensity/trait, with a negative connotation: *satsát* ‘babble, chatter’—*satsatéro* ‘chatterbox’

Third, the agentive *-ero* in Tagalog, unlike its original in Spanish, combines not only with nominal stems, but also with stems referring to an action, as in (12b). Fourth and final, following Seifart’s (2015) methodology, we can observe that the ratio of Spanish complex loanwords (25) to Tagalog hybrid formations (8) with *-ero*, and that to related simplex (stem) words, indicates a primarily indirect character of suffix borrowing from Spanish into Tagalog. This will be discussed further in Section 5.

Let us now turn to the recent Tagalog dataset, from the late 20th–early 21st century, in order to assess the contemporary usage and semantics of the borrowed suffix *-ero/a*. This dataset rendered far more Spanish complex loanwords and Tagalog hybrids: a dictionary search, cross-checked with the corpus data (see Section 1.3), gave 158 Spanish complex loanwords, including 150 items as actor nouns. These 150 nouns pertain to the same two semantic groups as above, namely ‘person of certain profession/occupation’ (n = 143) and ‘person of certain propensity/trait’ (n = 7), mostly with a negative connotation.

There is also a considerable number of hispanized English loanwords in the contemporary dataset, which are not included in the count. These are English lexemes which have been reshaped in Tagalog by analogy with a Spanish pattern, as in (13a–b), see also Section 1.1.⁷

⁷ A similar pattern of hispanization of English loanwords is observed in Chamorro, e.g.: English *upholsterer* > Chamorro *apostero/a* ‘upholsterer, m/f’ (Rodríguez-Ponga, 2009: 241–248).

TABLE 9.3 Characteristics of hybrid formations with *-ero/a* in contemporary Tagalog

Semantic group of HF	Type of stem	Derivates of Tagalog stems	Derivates of Spanish stems	Derivates of English stems
Person of certain profession/occupation	N	7	9	2
	V	1	2	1
Person of certain propensity/trait	N	4, neg.	-	1, neg.
	A	5, neg.	2, neg.	1, neg.
	V	5, neg.	4, neg.	1, neg.
TOTAL # of types		22 types	17 types	6 types

- (13) a. *blogéro* ‘blogger’ (coll.) < English *blogger*⁸
 b. *debatéro/a* ‘one who often disputes, m/f’ < English *debater* (cf. Spanish *polemista*)

Tagalog hybrid formations with *-ero/a* are also significant in number in this dataset, occurring 45 times. The full list of Tagalog hybrids with *-ero/a*, their stems and source forms, and the information on their token quantities in the corpus is presented in the Appendix (Table 9.13). Table 9.3 summarizes their main characteristics.

There are ten more Tagalog hybrids with *-ero/a* attested only in the contemporary corpus. These types have the lowest frequency (from 2 to 38 tokens in total), indicating their recent creation. All of them also carry the meaning ‘person of certain propensity/trait’ with a negative connotation, as in (14a–b).

- (14) a. *ingleséro/a* ‘Filipino who prefers English to his mother tongue, m/f’ < Tagalog *Inglés* < Spanish *Inglés* ‘English’
 b. *emotéro/a* ‘too emotional person, m/f’ < English *emotion/(to) emote*

Sixteen of the 45 hybrid forms listed in the Appendix (Table 9.13) are attested as both *-ero*, for masculine or generic and *-era* for feminine, which corroborates Stolz’s (2012) observations on the emergence of “marginal gender” in Tagalog (see also Bowen, 1971; Baklanova, 2016).

The suffix *-ero/a* demonstrates a growth of productivity over time in Tagalog. Although the size of the historical dataset is much smaller than the con-

8 The same tendency has evolved independently in contemporary Spanish (Gramática: 6.8p).

TABLE 9.4 Semantic distribution of CLS and HFS with *-ero/a* in contemporary Tagalog

Semantic group of derivatives	% of Spanish CLS	% of Tagalog hybrids
Person of certain profession/ occupation	c. 95%	c. 40%
Person of certain propensity/trait	c. 5%, mostly negative	c. 60%, all negative
Total # of types	150	55

temporary one, and thus cannot be directly compared, the scarcity of Tagalog hybrids (8 items) in the early dataset, and their much larger number in the contemporary data (55), including some recent creations, certainly implies a certain growth in productivity. Unlike in Spanish, in Tagalog *-ero/a* can combine with all type of stems: with nominal and verbal stems for the semantic group ‘person of certain profession/occupation’, and with nominal, adjectival and verbal stems for ‘person of certain propensity/trait’. Note also that the ratio between the two semantic groups for Spanish complex loanwords and Tagalog hybrid forms reveals a significant shift in Tagalog towards ‘person of certain propensity/trait’ with a distinct negative connotation, as illustrated in Table 9.4.

3.2 Spanish Suffix *-ista*

The Spanish suffix *-ista* is mostly added to nominal stems, both common and proper (Gramática: 6.9b), with rare cases of verbal and adjectival derivation, (see Rainer, 2011: 490). Its productivity reportedly correlates with that of derivatives with the abstract nominal suffix *-ismo* (Gramática: 6.9c). Diachronically *-ista* appears to be mostly productive in forming agentive nouns with the following semantics: ‘a person of a certain profession/occupation’ (15a), often also used as a corresponding relational adjective (Gramática: 7.7h); ‘a person of certain propensity/trait’ (15b), with weak productivity; and ‘a follower/participant of a tendency/movement/party’ (15c) (see, e.g., Gramática: 6.9b).

- (15) a. *técnico electricista* ‘electric technician’—*electricista* ‘electrician’
 b. *anécdota* ‘anecdote’—*anécdotista* ‘one who is prone to anecdotes; one who composes anecdotes’
 c. *absolutismo* ‘absolutism’—*absolutista* ‘supporter of absolutism’

In the historical dataset, 14 types of Spanish complex loanwords with *-ista* and only one Tagalog hybrid formation are attested. Their characteristics are summarized in Table 9.5.

TABLE 9.5 Characteristics of nouns with *-ista* in the Tagalog historical dataset

Semantic group of derivatives	Type of stem	CLS	Simplex	HFS		Simplex
			related to CLS	Tag. stem	Sp. stem	related to HFS
Person of certain profession/occupation	N	11	8	-	1?	1
Person of certain propensity/trait	V	1	1	-	-	-
Follower of a trend/party/movement	N	2	2*	-	-	-
TOTAL # of types		14	11	0	1	1

* The related simplex forms for the attested *anarkista*, *sosyalista* are *anarkíya* ‘anarchy’ and *sosyál* ‘social’ respectively. They are found in the later dictionaries (i.e. English, 1987; Rachkov, 2012), but are absent in the early dataset, presumably due to its small size. Nonetheless, it is possible that they might have been borrowed into Tagalog in the early 20th century, but were infrequent.

From this earlier data it may be noted that Spanish agentive complex loanwords of all the three original meanings are attested in Tagalog, with the items of group (a) prevailing (as in (15a)). Note further that all the stems of the complex loanwords except one are nominal, as in (16a–b).

- (16) a. *sálmo* ‘psalm’ (< Spanish *salmo*)—*salmísta* ‘psalmist’ (< Spanish *salmista*)
 b. Mex. Spanish *jaranista* ‘prone to revelry; player of a *jarana* (small four-string guitar)’ > Tagalog *haranísta* ‘person prone to revelry (archaic); with the simplex *harana* ‘revelry’ also attested

There is only one hypothetical hybrid form with *-ista* (marked with ‘?’ in Table 9.5) presumably derived from a Spanish stem (17).

- (17) *dibúho* ‘drawing’ (< Spanish *dibujo*) > *dibuhísta* ‘draftsman’ (cf. Spanish *dibujador/dibujante*)

However, it is also possible that *dibuhísta* is a Mexican Spanish complex loanword, as lexical items display geographical variation in agentive suffixes, such as Peninsular Spanish *jaranero* versus Mexican Spanish *jaranista* ‘prone to revelry’ (see DRAE 2014; Rainer, 2011). Thus the historical data is insufficient to confirm whether *-ista* had been borrowed into Tagalog by the early 20th century.

The lack of hybrids indicates either very weak productivity, or the complete absence of *-ista* in Tagalog lexical derivation in the (early) 1900s. However, the presence of a number of simplex-complex pairs of Spanish loanwords with *-ista* may have provided the basis for a possible reanalysis and subsequent decomposition of stem and *-ista* suffix in complex loanwords by Tagalog speakers.

The recent data displays a considerable increase in the number of complex loanwords in *-ista*, with around 140 counted in the dictionaries of English (1987) and Rachkov (2012). The vast majority of these forms have a corresponding simplex lexeme, thus fostering their reanalysis in Tagalog. There are also 19 hybrid formations, of which 13 are formed with Spanish stems, for example (18a), four with Tagalog stems (18b), and two more with recently borrowed English stems (18c).

- (18) a. *independísta* ‘person of independent character’ < *independénte* ‘independent’ (< Spanish)
 b. *balagtasísta* ‘follower of poet Balagtas’ < *Balagtas*
 c. *raliyísta* ‘demonstration participant’ < *ráli* ‘mass demonstration’ (< English ‘rally’)

The characteristics of these nouns in *-ista* are outlined in Table 9.6.

Both the complex loanwords and the hybrid forms belong to the three original Spanish semantic groups (18a–c). A further 15 Tagalog hybrids with *-ista* are attested in the LC, but with the lowest frequencies (2 to 13 tokens in total), which may indicate their very recent creation. There are items for each of the three meanings presented above among them, mostly derived from Spanish or English nominal and adjectival stems, see (19a–f).

- (19) a. *aghamísta* ‘scientist’ < Tagalog *agham* ‘science’ (< Skt *āgama* ‘religion; sacred science’)⁹
 b. *iligalísta* ‘one who is involved in an illegal business’ < Spanish *ilegal* ‘illegal’
 c. *parlorísta* ‘one who works in a beauty parlor/salon’ < English [*beauty*] *parlor*
 d. *mujerísta*¹⁰ ‘crossdresser or effeminate gay’ (slang) < Spanish *mujer* ‘woman’

9 See Casparis (1997).

10 There is a recent tendency in Tagalog to retain the original orthography of both Spanish and English donor words. Baklanova (2017: 353, Tab. 3) rates such cases as 0.2% of the total number of Spanish and English borrowings in her data.

TABLE 9.6 Characteristics of nouns with *-ista* in the contemporary Tagalog dataset

Semantic group of derivatives	Stem type	Avg CL	Simplex related to CL	Hybrid formations			Simplex related to HF
				Tag. stem	Spanish stem	Eng. stem	
Person of certain profession/ occupation	N	c. 85	c. 70	1	6	1	8
	V	5	5	2	1	1	4
Person of certain propensity/ trait	N	2	2	-	2	-	2
	A	2	2	-	2	-	2
Follower of a trend/ party/ movement	N	c. 45	c. 40	1	2	-	3
TOTAL types		c. 140	c. 120	4	13	2	19

- e. *wangwangísta* 'one who uses special car signal to demonstrate authority' (neg. coll.) < Tagalog *wangwáng* '1. completely exposed; 2. special car signal to give a priority pass'
- f. *punkísta* 'punk' < English *punk*

Compared with *-ero/a*, *-ista* appears to be a more recently borrowed suffix in Tagalog, with an observable growth in productivity attested in the contemporary sources. It derives agentive nouns of the same semantic groups as *-ero/a*, with the semantics 'person of a certain profession/occupation' prevailing (see Table 9.6). However, *-ista* tends to convey the meaning 'person of certain propensity' in a neutral manner, whereas *-ero/a* conveys a negative connotation for this semantic group (see Table 9.4). The suffix *-ista* also derives nouns meaning a 'follower of a tendency/movement/party', which *-ero* lacks.

In Tagalog *-ista* combines with the same types of stems as the Spanish complex loanwords, with nominal stems most common for all three semantic groups. The contemporary data also comprise many complex loanwords with *-ista* that are not Spanish loanwords, but rather English cognates or false cognates formed with the suffix *-ist*, which have been reshaped in Tagalog by analogy with Spanish (20a–b).

- (20) a. *kolon-ísta* < English *colon-ist* (cf. Spanish *colono*)
 b. *loyal-ísta* < English *loyal-ist* (cf. Spanish *partidario del régimen*)

This reshaping of *-ist* > *-ista* in the context of large-scale assimilation of English lexical items makes the Spanish suffix *-ista* more frequent in Tagalog speech which, in turn, may lead to an increase in its productivity with native stems.

3.3 Spanish Suffix *-eño/a*

The Spanish suffix *-eño/a* is one of the suffixes that can form relational adjectives from proper nouns (place and personal names) and common nouns, usually with the following meanings (Gramática: 7.6ñ–7.6o; Rainer (2011: 475)):

- Born/living in N, e.g. *Madrid*—*madrileño/a* ‘born/living in Madrid’
- Pertaining to N, e.g. *Velazquez*—*velazqueño/a* ‘pertaining to Velazquez (or his painting), m/f’; *águila* ‘eagle’—*aguileño/a* ‘pertaining to an eagle, aquiline, m/f’

It is claimed that *-eño* was borrowed into Tagalog in the form *-enyo* with the meaning ‘person born/residing in some place’ (Rachkov, 1981: 59; Alcántara y Antonio, 1999). However, no clear evidence of Tagalog hybridization with the Spanish suffix *-eño* was found in this study.

No such derivatives can be attested with certainty in the historical dataset; only a small number of personal names were found. Indeed, the search found no evidence of *-enyo* hybridization in Tagalog until the end of the 19th century. During the 20th century there was a growth in number of *-enyo* derivatives in the texts. The dictionaries queried give 10 *enyo*-formations meaning ‘person born/residing in some place’, mostly with names of big cities, provinces and countries as stems. Four overt Spanish loanwords, with names of countries (21a), a city and the word ‘island’ (21b) as stems were attested.

- (21) a. *Brasilényo/a* ‘Brazilian (resident) m/f’ < Spanish *Brasileño*
 b. *islényo* ‘resident of an island’ < Spanish *isleño* ‘pertaining to an island’

Five derivatives with names of Philippine provinces as stems were also attested, as in (22).

- (22) *Batángas*—*Batang(g)ényo/a* ‘resident of Batangas province, m/f’

Finally, we also found one derivative with the name of a capital as stem (23).

- (23) *Manila*—*Manilényo/a* ‘resident of Manila, m/f’

The most recent data show that derivatives with *-enyo* are in use in contemporary Tagalog, although with low frequencies (from 2 to 40 total tokens). A number of

formations has been attested, predominantly with names of the old Philippine cities and provinces as stems, in both Spanish and Tagalog orthography, as in (24a–b).

- (24) a. *Davaoéño* ‘1. born/resident of Davao city/province; 2. native dialect of Davao’
 b. *Palawéño / Palawényo* ‘born/resident of Paláwan island’

It is still unclear whether the nouns/adjectives related to such important geographical areas are indeed derived in Tagalog, or whether they were simply diffused in the early 20th century as loanwords from Spanish language newspapers, legal documents and other sources. The double orthography of the suffix *-enyo/-éño* at present may reflect present-day Filipinos’ awareness of its Spanish provenance, and their positive attitude to the foreign spelling. This interpretation is supported by the official introduction of some Spanish letters into the Filipino alphabet (Ortograpiyang Pambansa 2013).

All such items co-vary with the derivatives of the native nominative strategy *taga+place*, see (25a–b), where token frequencies from the LC are given in brackets.

- (25) a. *taga-Ma(y)níla(?)* (55)—*Manilényo, Maniléño* (41) ‘Manila-born/resident’
 b. *taga-Táguig* (6)—*Taguig(u)éño* (5) ‘Taguig-born/resident’

Since further research is needed to identify cases of *-enyo* derivation with recent stems and thus to verify the productivity or lack of productivity of this Spanish suffix in Tagalog, it is not included in the analysis in the next section.

3.4 *Impact of the Spanish Agentive Suffixes ero/a and -ista on Tagalog Derivation*

Table 9.7 presents the impact of *-ero/a* and *-ista* on the Tagalog agentive derivation inventory outlined in the preceding sections.

For the semantic groups ‘person of certain profession/occupation’ and ‘person of certain propensity/trait’ Tagalog lacks native affixal inventory to derive an agent noun from a nominal or adjectival stem. The introduction of the Spanish suffixes *-ero/a* and *-ista* into Tagalog morphology partly fills this gap. That said, with the addition of the Spanish strategies to the two existing Tagalog ones (*taga-* and *may/mag+r*), native derivation with verbal stems has become redundant, and a functional differentiation of these four strategies may be expected in the future.

TABLE 9.7 Comparison of native and Spanish strategies of agentive derivation in Tagalog

Semantic group of agentive noun	Stem type	Prefix <i>taga-</i>	Prefixes <i>mag/mag + r</i>	Prefix <i>pala-</i>	Suffix <i>-ero/a</i>	Suffix <i>-ista</i>
Person of certain profession/occupation	N	-	-	-	+	+
	A	-	-	-	-	-
	V	+	+	-	+	+
Person born, living, or working at place	N	+	-	-	-	-
	A	-	-	-	-	-
	V	-	-	-	-	-
Person of certain propensity/trait	N	-	-	-	+ neg.	+
	A	-	-	-	+ neg.	-
	V	-	+	+	+ neg.	-
Follower of tendency/movement/party	N	-	-	-	-	+
	A	-	-	-	-	-
	V	-	-	-	-	-

The *-ero* derivation adds a negative connotation to the nouns referring to ‘person(s) of a certain propensity/trait’, while *pala-* and *-ista* are neutral. Thus, it appears to be the first item of affective morphological inventory in Tagalog. Moreover, *-ista* has introduced the new meaning ‘follower of tendency/movement/party, etc.’ to the Tagalog derivational inventory. Finally, it should be noted again that *-ero* and *-ista* are similar to the corresponding English suffixal forms *-er* and *-ist* and thus enable the phonetic assimilation of English borrowings which, in turn, appears to foster the adoption of English lexical items into Tagalog.

4 Spanish Diminutive Suffixes in Tagalog Lexical Derivation

Spanish possesses many suffixes that produce diminutives of nominal, adjectival and adverbial stems (Gramática: 9.1b). They help to express “a wide range of affective notions (size, affection, disapproval, irony, etc.)”, thus a noun + *ito/ita* “spring[s] more readily to the tongue of a Spanish-speaker than a noun+ *pequeño* [‘small’], especially in Mexican Spanish (Batchelor and San José, 2010: 450). Jurafsky (1996: 543) shows that the basic meaning of diminutives refers to the concepts of being ‘small’ or ‘a child’, with a metaphorical development into a meaning conveying an attitude of the speaker. It has been claimed

that the Spanish diminutive suffixes *-ito/a*, *-illo/a* and *-ete* have been adopted into Tagalog nominal derivation (Wolff 1973, 2001; Baklanova, 2004; Quilis and Casado-Fresnillo, 2008).

4.1 Suffixes *-ito/a* and *-illo/a*

-Ito/a is considered to be currently the most productive diminutive suffix in Spanish, whereas historically *-illo/a* predominated (Gramática: 9.1j). This preference is manifested in, for example, the prevalence of Spanish diminutive toponyms with *-illo/a* in Spain (Gramática, 9.1m, j). Although *-illo/a* and *-ito/a* alternate with some stems (26a), in Latin America *-illo/a* is regarded as having mostly negative connotations (26b) (Batchelor and San José, 2010: 452).

- (26) a. *cuchara* ‘spoon’—*cucharita/cucharilla* ‘small spoon, teaspoon’
 b. *guerra* ‘war’—*guerrilla* ‘1. insignificant war, skirmish; 2. guerilla’

The search of the historical dataset produced the ratio of Spanish complex loanwords with *-ito/a*, *-ilyo/a* (*-illo/a*) to their related simplex loanwords, and to possible hybrid formations (Table 9.8).

Only four Spanish complex loanwords with the suffix *-ito/a* are attested in the early dataset, all of which are related to ‘an object smaller than that designated by the stem’. Two of them have their simplex pairs, such as in (27).

- (27) *palíto* ‘toothpick; matchstick; small stick’ (< Spanish *palito* ‘small stick’)—
pálo ‘stick’ (< Spanish *palo* ‘idem.’)

At least two types of hybrid formations with *-ito/a* are: one with a Spanish stem (28a), and one with a non-Spanish stem that was borrowed earlier into Tagalog (28b).

- (28) a. *naran(g)híta* ‘tangerine; small orange’ (cf. Spanish *naranjillo* ‘small green citrus’)—*narán(g)ha* ‘orange’ (< Spanish *naranja*); cf. the later loan-blended form *dalanghíta* (< Tagalog *dalandán* ‘orange’)
 b. *sampag(u)íta*—*sampága* ‘Jasminium sambac, Arabian jasmine’ < Skt *campaka* ‘Michelia Champaka’ (M-W, 1899: 388.3),¹¹ probably via Malay *cempaka* ‘Michelia Champaka tree’ (cf. Casparis, 1997: 15)

11 Skt *campaka* ‘Michelia Champaka’ as the etymon for Tagalog *sampága* with a close meaning casts doubts on the supposition of Blust and Trussel (2010) that the base of Ilokano *sampága* ‘may be native to some Philippine languages, the longer word with diminutive suffix appears to be a Spanish loan in both the Philippines and the Marianas’.

TABLE 9.8 Characteristics of nouns with *-ito/a* and *-ilyo/a* (*-illo/a*) in the historical dataset

Semantic group of derivatives	Type of stem	-ITO/-ITA		-ILYO/-ILYA (-illo/illa)	
		Ratio of CLS: simplex	Ratio of HFS: simplex	Ratio of CLS: simplex	Ratio of HFS: simplex
Smaller object	N	4:2	2:2 1 Spanish, 1 non-Spanish	8:4	-
	A	-	-	-	-
	V	-	-	-	-
TOTAL of types		4:2	2:2	8:4	0

Though it is uncertain whether the two hybrid forms were created by Tagalog bilinguals, there is also no evidence for the Spanish provenance of these hybrids in the early dictionaries (Serrano Laktaw, 1889; Lopes and Bensley, 1895; Calderón, 1915). It is therefore possible that they may be early Tagalog hybrid forms.

Slightly more complex loanwords with *-ilyo/a* (*-illo*) are attested in the historical dataset. They are also related to ‘a smaller object’, and three occur with their simplex pairs, as in (29).

- (29) *ganchílyo/gantsílyo* ‘crochet hook’ (< Spanish *ganchillo*)—*gáncho/gántso* ‘hook; staple’ (< Spanish *gancho*)

No hybrid formations with *-ilyo/a* were found, although there are two instances of diminutive Tagalog hybrid formations that may pertain to the lexicon of the early 20th century, despite being unattested in this limited dataset. Both forms are derived from Tagalog stems and have the basic meaning of ‘younger, child’. As will be seen below, both are attested in the contemporary data, where they have a much higher token frequency (c. 500 tokens each) than other hybrid forms from the same period, which may indicate their older provenance. Two examples can be observed in (30a–b).

- (30) a. *binatílyo* ‘preadolescent boy’ < Tagalog *bináta?* ‘young man, bachelor’
 b. *dalagíta* ‘preadolescent girl’ < Tagalog *dalága* ‘maiden’

There are 39 complex loanwords with diminutive *-ito/a* attested in the contemporary data, the majority of which have nominal stems. All of them pertain to

one of the two semantic groups ‘object smaller than the stem’ (31a) and ‘child (of human or animal)’ (31b–c). See also Table 9.9.

- (31) a. *labahíta* ‘small razor; small penknife’ (< Spanish *navajita* ‘small clasp-knife’ (archaic))—*labáha* ‘razor; knife’ (< Spanish *navaja* ‘clasp-knife; razor’)
 b. *guwapíto/a* ‘pretty boy/girl’ (< Spanish *guapito/a*)—*guwápo* ‘nice, pretty’ (< Spanish *guapo*)
 c. *kabríto/a* ‘goatling, m/f’ (< Spanish *cabrito*)—*kábra* ‘goat’ (< Spanish *cabra*)

Spanish complex loanwords in *-ilyo/a* outnumber those in *-ito/a*, with a total of 53. They have nominal stems and pertain to ‘an object smaller than the stem’, as in (32).

- (32) *granílyo* ‘small grain’ (< Spanish *granillo*)—*gráno* ‘grain; pimple’ (< Spanish *grano*)

No complex loanwords in *-ito/a* or *-ilyo/a* were found with negative connotations, although three with the suffixal form *-silyo* (< *-cillo*) have a slightly negative or pejorative meaning, referring to ‘someone less significant than the stem’, see (33).

- (33) *gobrnador-silyo* (< Spanish *governadorcillo*) ‘city authority lower than governor’—*gobrnador* (< Spanish *governador*) ‘governor’

Regardless of the significant number of simplex-complex pairs of diminutive complex loanwords pertaining to the basic meaning ‘small object’, and some meaning ‘child’, contemporary Tagalog hybrid formations with *-ito/a*, *-ilyo/a* show a shift toward human nouns with affective connotation. More specifically, *-ito/a* appears to have recently developed an ironical connotation to a person denoted by the stem, close to the meaning ‘one who looks like/imitates the stem’, such as in (34a–c), where token frequency in the LC is provided in brackets.

- (34) a. *bagíto* ‘newbie; someone unskilled’ (359)—Tagalog *bágo* ‘new’
 b. *baklíta* ‘effeminate male’ (coll.) (59)—Tagalog *baklá?* ‘gay’
 c. *puríta* ‘one who looks like a poor person’ (ironic) (5) < English *poor* (as an unassimilated borrowing)

The hybrid formation types with *-ito/a* meaning ‘child’ (35a) or ‘small object’ (35b) are scarce and formed with Spanish stems. Again, LC token frequency is provided in brackets.

- (35) a. *Tsinító/a* ‘(one who looks like) a Chinese boy/girl’ (8)—*Tsino* ‘Chinese’ (< Spanish *Chino*)
 b. *platító* ‘small portion; small dish’ (cf. Spanish *platillo* ‘small dish’ (archaic)), (18)—*plátó* ‘dish; portion’ (< Spanish *plato*)

The suffix *-ilyo/a* is attested in only three hybrid forms whose attribution is problematic. They are all derived from Spanish borrowed stems, however, these stems do not appear in the dictionaries consulted (Serrano Laktaw, 1889; Lopes, Bensley, 1895; DRAE). All the formations are agentive human nouns with an ironic/negative connotation, as in (36).

- (36) *maestrílyo* ‘one who likes to sermonize’—*maéstro* ‘teacher’ (< Spanish *maestro*)

The characteristics of the attested complex nouns and hybrid forms with *-ito/a* and *-ilyo/a* in Tagalog and their associated ratios are presented in Table 9.9.

As discussed in Section 2, Tagalog lacks a clear native diminutive suffixal strategy, relying instead on the suffix-stem duplication (*R+(h)an*) construction. Rather than conveying the canonical meaning of ‘small object’ for inanimate stems and ‘child’ for animate stems, the *R+(h)an* strategy conveys the mixed meaning ‘small or imitated object’ for inanimate stems, with rare cases of nouns with human-related stems conveying a mildly negative connotation, namely ‘one who imitates/pretends’. Thus, in this case the trigger for transfer cannot have been functional and structural congruency of the affixes between the two languages (Winford, 2003: 92–93; Matras, 2007: 34; Chamoreau, 2012: 85–86). That said, the morphotactic transparency of the Spanish suffix might have facilitated its borrowing into the Tagalog system (see Gardani, 2008). Moreover, as Tagalog lacks native affixal inventory for the semantic group ‘younger, child’, the borrowing of *-ito/a* shows potential, albeit weakly, to fill this gap. The recent hybrids with *-ito/a* are formed purely as agentive nouns, with nominal and adjectival stems of Tagalog, Spanish and English provenance.

The derivation with *-ito/a* thus provides Tagalog with a clear diminutive strategy. Its interaction with the native *R+(h)an* strategy may account for the development of a similar meaning for human noun derivations with *-ito/a*, such as *sántu-santú-han* ‘one who pretends to be holy, a prude’ and *santo-santító* with the same meaning. Thus the new pattern with *-ito/a* seems to undergo

TABLE 9.9 Characteristics of nouns with *-ito/a* and *-ilyo/a* in contemporary Tagalog

Semantic group of derivates	Type of stem	-ITO/-ITA		-ILYO/-ILYA	
		CLS: Simplex	HFS: Simplex	CLS: Simplex	HFS: Simplex
Smaller object/animal	N	32:17	2:2 1 Spanish, 1 non-Spanish	52:21	-
	A	-	-	1:1	-
Child	N	6:6	3:3 2 Spanish, 1 non-Spanish	-	1:1 1 Spanish
One who is similar to/imitates the stem	N	-	6:6 2 Spanish, 4 non-Spanish	-	2:2 (neg.) 2 Spanish
	A	-	3:2 1 Spanish, 2 non-Spanish	-	-
	V	-	1:1 1 non-Spanish	-	-
TOTAL ratio of types		38:23	15:14	53:22	3:3

a functional differentiation towards an affective connotation, mostly of the meaning 'one who looks like/imitates the stem'. The current emergence of personal names (nicknames) with *-ito/a* attested in the LC corroborates this view, since they also bear affectiveness. Take, for example, *Milk-ita* as a brand name of milk products, *Dracul-ita* as a movie character, and nicknames such as *Daldal-ita* (< Tagalog *daldál* 'talkative').

It appears that the borrowing of *-ito*, *-illo/-ilyo* into Tagalog might have begun in the early 20th century, or perhaps even earlier, but has not yet reached its completion. There is clear evidence of only a small number hybrids adopted by the masses, such as *sampagita* as a Philippine national symbol; *dalagita* and *binatilyo* as the terms filling the lexical gap 'teenager' with relatively high frequencies (c. 500 tokens each in the LC). The suffix *-ito/a* still shows weak productivity, mostly with a mildly negative or ironical meaning. Low token frequency and the absence of some of the hybrid forms with *-ito/a* in the dictionaries consulted indicate their most recent creation. Such items still appear to be

cases of individual usage by (educated) Tagalog speakers. The scarcity of hybrid forms with *-ilyo/a* in the dictionaries and their absence in from the corpora indicates that this derivation strategy is unproductive in contemporary Tagalog.

4.2 *Spanish Suffix -ete*

The suffix *-ete/a* is among the less frequent diminutive suffixes in Spanish, being used both neutrally and affectionally or pejoratively (Batchelor and San José, 2010: 452). It is also productive to a certain extent as a nominal suffix denoting an instrument or utensil, such as *color* ‘color’—*colorete* ‘blusher, rouge’ (Rainer, 2011: 217–218).

Ten complex loanwords with *-ete* are attested in the historical Tagalog dataset, all of which refer to an instrument or utensil, such as *bilyéte* ‘bill; ticket’ (< Spanish *billete*). No simplex pairs or diminutives were registered. Only one hybrid form with *-ete* occurs (37), which might have been created by Tagalog-Spanish bilinguals rather than by analogy, since there are no simplex-complex pairs attested in the data. This form is still in use at present (70 tokens in the LC).

(37) *kaliwéte* ‘left-handed; leftist’—*kaliwá?* ‘left’

The contemporary Tagalog dataset includes 39 complex loanwords with *-ete*, which relate to the semantic groups of ‘instrument/utensil’ (38a), ‘smaller object/animal’ (38b) and ‘person of certain occupation’ (38c); note that almost half of these forms also have a related simplex loanword.

- (38) a. *asuléte* ‘bluing (for linen)’ (< Spanish *azulete*)—*asúl* ‘blue’ (< Spanish *azul*)
 b. *toréte* ‘a small bull’ (< Spanish *torete* ‘small bull; difficult point’)¹²—*tóro* ‘bull’ (< Spanish *toro*)
 c. *gruméte* ‘yunker, ship’s boy’ (< Spanish *grumete*)—(no simplex)

There are only two more hybrid formations in the recent data, one with a Spanish borrowed stem (39a) registered only in Rachkov (2012), the other with a Tagalog stem (39b) that is an analogical creation based on (37).

- (39) a. *negosyéte* ‘huckster, haggler’ (neg.)—*negósyo* ‘commerce, business’ (< Spanish *negocio*)
 b. *kananéte* ‘right-handed’—*kánan* ‘right (side)’

12 See Lopes and Bensley (1895: 599).

TABLE 9.10 Characteristics of nouns with *-ete* in contemporary Tagalog

Semantic group of derivates	Type of stem	Ratio	
		CL:Simplex	HF:Simplex
Instrument/utensil	N	25:7	-
	A	1:1	-
Smaller object/animal	N	10:7	-
Person of certain occupation	N	3:0	1:1 (neg.) Spanish stem
Person of certain trait	N	-	2:2 Tagalog stems
TOTAL ratio of types	39:15	3:3	

Table 9.10 outlines the characteristics of nouns with *-ete* in contemporary Tagalog.

Thus, although the complex loanwords with *-ete* in Tagalog mostly refer to an 'instrument/utensil' or a 'smaller object', no hybrid forms exist with such meanings. The three attested hybrids do not show consistency in semantics, with one Spanish-derived item referring to 'a person of certain occupation', and another denoting a 'person of certain trait'. Indeed, except for (39) as a clear analogical creation, there is no other evidence for the productivity of *-ete* in the recent corpus.

5 Discussion of the Results

Contact-induced change requires a certain degree of bilingualism in the recipient community for linguistic innovations to spread (Winford, 2003a). However, until the 19th century there had been only a very small stratum of bilingual Spanish-Tagalog mestizos in the Philippines (Lipski et al., 1999). Only in the late 19th century did the bilingual community grow significantly due to a "new wealthy class of Chinese mestizos" who readily learned and used Spanish for their commercial interests (Thompson 2003: 16). Additionally, "individuals who have large numbers of weak ties outside the community tend to be innovators, and to serve as instigators of language change" (Bright, 1998: 90–91; see also Milroy and Milroy, 1992). In the case of the Philippines, individuals with higher socioeconomic status and stronger inter-community ties, namely

TABLE 9.11 Scale of directness of affix borrowing

Directness of borrowing	Direct borrowing				
	Direct borrowing			Indirect borrowing	
Complex loanwords:	None	Few	Few	Many	Many
Frequent simplex loanwords:	None	None	Many	Many	Many
Knowledge of donor language:	Yes	Yes	Yes	Yes	No

SEIFART, 2015: 527, FIG. 3

Spanish-Tagalog mestizos (in local administration) and active Chinese-Tagalog mestizos (as leaders in trade) might have been the only agents of Spanish borrowing and innovations in Tagalog up to the early 20th century.

The Tagalog-Spanish contact situation corroborates Winford's (2003b: 134) observation that *direct borrowing* of bound morphemes "requires a high degree of bilingualism among individual speakers", while "certain structural innovations in an RL appear to be mediated by lexical borrowing", i.e. adopted through *indirect borrowing*. As shown in Sections 3 and 4, the majority of hybrid creations with Spanish suffixes have a number of simplex-complex pairs of Spanish loanwords as the foreground for the indirect borrowing process. However, there are some Tagalog-Spanish hybrids which do not have such corresponding pairs of simplex-complex loanwords.

This situation correlates with Seifart's (2015) assumption, that both *direct* and *indirect* scenarios of affix borrowing may apply in the majority of cases, making it possible to define only the *primary* character of the borrowing in a given RL. As such, Seifart (2015: 527) proposes a scale of directness of affix borrowing, which is reproduced in Table 9.11.

Three major criteria indicate that *indirect* borrowing (i.e. the borrowing of an affix from the loanwords adopted in the RL) was "the only or primary process involved" in the transfer of an affix to the RL (Seifart, 2015: 514):

- 1) The number of complex loanword types is larger than the number of hybrid formations;
- 2) The existence of pairs of loanwords with and without a certain affix; and
- 3) Low token frequencies of complex loanwords, in comparison to the frequencies of their corresponding simplex forms.

These three conditions provide a strong basis for reanalyzing the structure of a complex loanword in the RL, and for extracting its affix for subsequent use in analogical creation. As observed by Bybee (1995: 434), "the more forms that bear an affix, the stronger the representation of that affix, the greater likelihood

TABLE 9.12 Summary of distribution of agentive *-ero/a* in Tagalog historical data

CRITERION	VALUE	RATIO
Ratio of CL to HF	25:8	3:1
Ratio of total CL to the simplex-complex pairs	25:13	2:1
Ratio of total simplex-complex pairs to infrequent CL*	13:3	4:1

* The limited dataset appears insufficient to check criterion 3.

that that affix will be productive". Consequently, "if no complex loanwords that would include the borrowed affix are attested, this is a strong indicator of direct borrowing" (Seifart, 2015: 528). In this case, there is no lexical basis in the RL for extracting the affix, so a speaker may only receive it directly from their knowledge of the donor language.

Regarding the simplex-complex pairs of Spanish loanwords attested in Tagalog, *indirect* borrowing appears to be the primary mode of adopting most of the suffixes discussed in the previous sections. To verify this assumption, Seifart's methodology is applied to analyze the ratio of complex loanwords to hybrid formations with each Spanish agentive and diminutive suffix discussed. Table 9.12 illustrates this analysis using the case of *-ero*.

Table 9.12 indicates that Seifart's criteria 1 and 2 are well met in our data: the number of compound loanword types with *-ero/a* is three times larger than that of hybrid formations; and half of the compound loanwords have their simplex pairs in Tagalog. Criterion 3 is only partially met, partly due to the rather limited early text dataset, which extends to only about 82,500 tokens, making it difficult to correctly assess token frequencies. Thus, the above distribution ratio should be regarded as a preliminary estimate, which requires a follow-up study using a larger corpus, preferably including texts from early newspapers as a vehicle for lexical innovations. Nonetheless, on the basis of criteria 1 and 2, it seems fair to propose that the primary process involved in the transfer of the Spanish suffix *-ero/a* to Tagalog was *indirect borrowing* from a number of complex loanwords.

The second adopted Spanish agentive suffix *-ista* is less productive and appears to have been borrowed into Tagalog more recently than *-ero*, since the early dataset does not include any *-ista* hybrids with a Tagalog stem (see Table 9.5). The ratio of complex loanwords to hybrid formations (with a Spanish stem) is 14:1, while the ratio of total complex loanwords to their simplex-complex pairs is 1:0.8. These distributions served as a sound basis for the decomposition of the suffix from the complex loanwords by speakers. A sig-

nificant growth in the number of complex loanwords with *-ista* and the corresponding recent simplex loanwords (c. 140:120) also correlates with a growth in hybrid formations (19, including items with Tagalog stems; recall Table 9.6). This again indicates the *indirect* character of borrowing of the suffix *-ista* into Tagalog.

As shown in Tables 9.8 and 9.9, several compound loanwords and hybrid formations with *-ito/a* are already attested in the early dataset, and both increase in frequency in the contemporary data, with the same ratio of 2:1 during the two periods. Many of the compound loanwords are less frequent, however, than their simplex pairs. Thus *-ito/a* also meets Seifart's criteria for the primarily *indirect* character of borrowing.

As only one possible hybrid with *-ilyo/a* and one with *-ete* are attested in the early dataset, both with Spanish stems, we assume that these Spanish suffixes might not have been adopted into Tagalog until the 1900s. The simplex-complex pairs with and without *-ilyo/a* in Tagalog, the lower frequency of many compound loanwords compared with their related simplex types, as well as the lack of hybrid formations with Tagalog stems strongly suggest that this possible *indirect* suffixal borrowing is not yet complete, and that the suffix *-ilyo* is not productive in Tagalog.

As for *-ete*, Seifart's criteria 2 and 3 are not met, due to the lack or absence of the simplex corresponding forms for the complex loanwords. Thus it is possible that the only hybrid formation attested in the early data (37) could be an individual creation by Spanish-Tagalog bilinguals who might have directly transferred the Spanish suffix onto the Tagalog stem. In other words, they may have extracted the suffix using knowledge of Spanish (the source language) "with its subsequent use on native stems" (Seifart, 2015: 529). Except for (39) as a clear analogical creation, there is no other evidence for the productivity of *-ete* in the recent corpus, thus it appears to not yet have become a part of Tagalog lexical derivation. However, a more detailed investigation with a larger dataset would be instructive for clarifying the status of *-ilyo/a* and *-ete* in Tagalog.

6 Concluding Remarks

The Tagalog data presented in this study corroborate the observation that "in adstrate situations, borrowing affects the lexicon first, before it extends to other domains of language structure" (Haspelmath, 2009: 50). The majority of the Spanish suffixes discussed here appear to have been adopted through a primarily *indirect* borrowing process, that is, from Spanish complex loanwords (Seifart, 2015).

It has been demonstrated that structural items from a source language are borrowed more easily if the function they express already exists in the recipient language, but in a less analytic form (Gómez Rendón, 2008: 102). This is also true for the Tagalog case: the Spanish suffix *-ero/a* displays clear morpheme boundaries, and has thus provided a comprehensive strategy for deriving agentive nouns from any type of stem.

We found that *-ero/a* is the most productive Spanish nominal suffix in Tagalog. As it may combine with any type of stem in Tagalog, including unassimilated borrowings, this may also foster its hybridization with English borrowings by Tagalog-English bilinguals, such as *stiréro* ‘teaser; cheater; prankster’ < English (*to*) *stir*. Moreover, the suffixes *-ero* and *-ista*, which correspond to English *-er* and *-ist*, promote the phonetic assimilation of English borrowings, thus increasing the adoption of more English lexical items into Tagalog. Indeed, the growing adaptation of English lexemes through such a hispanization process may increase the amount of *-ero* and *-ista*-derivatives in Tagalog which, in turn, may lead to the hybridization of the suffixes with a wider range of stems (see Wolff, 2001).

Spanish suffixes in Tagalog provide a good example of the widely attested tendency for polysemantic morphemes from a source language to be borrowed into a recipient language with their most concrete meanings and functions (Winford, 2003a: 91–92). However, “the erstwhile patterns come to coexist with new ones, and new rules develop governing the functional differentiation of new and old patterns” (Aikhenvald 2007: 46). Indeed the derivation with *-ito/a* in Tagalog seems to interact with the native diminutive strategy *R+(h)an* possessing the mixed semantics of ‘smallness’ and ‘imitation’. This interaction may account for the development of a similar meaning for the human noun derivation with *-ito/a*, namely ‘one who looks like/imitates the stem’.

To conclude, it should be noted that the use of these Spanish suffixes as nominalizers only enlarges the purely nominal morphological base of Tagalog, and in the future may lead to a more distinct functional distribution of the Tagalog derivational inventory, with clearer boundaries between the lexical classes.

Acknowledgements

The authors would like to thank John Wolff, Marian Klamer, Francesca R. Moro, Maria Kristina Gallego and the anonymous reviewers for their valuable comments and suggestions on earlier versions of this chapter. Any remaining shortcomings are the authors’ responsibility.

Appendix

TABLE 9.13 Hybrids with *-ero/a* in contemporary Tagalog

No	Tagalog hybrid formation	# of tokens in LC	Tagalog simplex word	# of tokens in LC	SL
1.	ansikutero 'loiterer, truant'	0	ansikót 'loitering; truancy'	0	Tag
2.	babaero 'philanderer'	143	babae 'woman'	12027	Tag
3.	balitero 'reporter'	0	balitaq 'news'	4573	Tag
4.	bangkero 'boatman'	19	bangkáq 'boat'	537	Tag
5.	baság-ulero/-a 'trouble-maker, m/f'	7	baság-ulo 'altercation; scuffle'	3	Tag
6.	boksingero 'boxer'	543	bóksing 'box, boxing'	348	Eng
7.	bulsero 'pickpocket' (cf. Spanish, Mex Spanish bolseador, carterista)	0	bólsá [< Mex Spanish bolsa 'pocket; pouch']	387	MexSp
8.	bombéra 'porno actress' (cf. Spanish bombero 'fireman; worker on petrol pump')	0	bómbsa [< Spanish bomba 'pump, fire engine; bomb']	435 as 'bomb; pump' 2 as 'porno scene'	Sp
9.	boratséro/-a 'drunkard, m/f' (cf. Spanish borrachera 'drunkenness')	0	borátso [< Spanish borracho 'drunk; drunkard']	0	Sp
10.	bosero 'peeper, voyeur'	6	boso [<? Mex buzo 'Look out! Watch it!']	1	MexSp?
11.	bulakbulero/-a 'truant; vagabond m/f'	1	bulakból [< English black ball]	12	Eng
12.	bungangéro/-a 'chatterbox, m/f'	5	bunganga 'gullet of animals/fish; mouth'	151	Tag
13.	butangéro 'bandit, gangster'	5	butáng 'beating up; thrashing'	0	Tag
14.	kaing(in)éro 'one who clears land for farming'	1	kaingín 'burning off in field for cultivation; cleared land in a forest'	12	Tag
15.	kartomanséro 'fortune-teller by cartomancy' (cf. Spanish cartomante)	0	kartomans(i)ya [< Spanish cartomancia]	0	Sp
16.	kaskaséro/a 'speed maniac, m/f'	27	kaskás 'sudden effort; spurt; rush'	1	Tag

TABLE 9.13 Hybrids with *-ero/a* in contemporary Tagalog (*cont.*)

No	Tagalog hybrid formation	# of tokens in LC	Tagalog simplex word	# of tokens in LC	SL		
17.	Katipunéro/a	'revolutionary of Katipunan society'	36	Katipúnan	'revolutionary society'	234	Tag
18.	komikéro	'comic, clown' (cf Spanish payaso 'clown', cómico 'comic')	8	komiko [<small><</small> Spanish cómico]	'clown, comedian; comic (adj.)'	4	Sp
19.	daldalero/-a	'gabbler; gossiper; chatterbox, m/f'	37	daldál	'gossiping; jabber; talkative'	1	Tag
20.	dupléro	'participant of duplo poetry competition'	1	dúplo [<small><</small> Spanish duplo 'double; a group of two']	'poetic duel as competition'	4	Sp
21.	hambugéro	'boaster, braggart'	0	hambóg	'boastful, arrogant'	25	Tag
22.	isnabéro/-a	'snob, m/f'	18	isnáb [<small><</small> English snob]	'snob'	1	Eng
23.	lakwatséro/-a	'truant; loiterer'	6	lakwátsa [<small>? <</small> Mex (el)acuache 'buddy, mate']	'truancy; staying away from school or work'	12	Mex Sp?
24.	langiséro/-a	'smoothie, flatterer'	0	langís	'oil'	1383	Tag
25.	lasing(g)éro	'drunkard'	7	lasíng	'drunk; inebriated'	520	Tag
26.	madyongéro	'player of mah-jong'	0	madyóng / majóng [<small><</small> ?Ch/Mal]	'game of mah-jong'	4	?Ch/Mal
27.	musikéro/-a	'musician, m/f' (cf. Spanish músico)	182	músika [<small><</small> Spanish música]	'music'	1099	Sp
28.	osyoséro/-a usyoséro/-a	'unduly curious person, m/f'	8	osyóso/usyóso * [<small><</small> Spanish ocioso 'idle']	'curious; idle'	2	Sp
29.	pakialaméro/-a	'meddler; busybody'	31	pakialám	'interfering, meddling'	561	Tag
30.	palikéro	'man who is too free and insincere with women, philanderer'	10	?palíki?, mamalíki? 'to philander'	*'philandering'	0	Tag
31.	pangging(g)éro/-a	'player of panggingge, m/f'	0	panggingge / panguingue	'card game of unknown origin, resembling rummy' (popular in the Philippines at least in late 19th-early 20th century)	0	?

TABLE 9.13 Hybrids with *-ero/a* in contemporary Tagalog (*cont.*)

No	Tagalog hybrid formation	# of tokens in LC	Tagalog simplex word	# of tokens in LC	SL
32.	parakaidéro 'paratrooper' (cf. Tagalog parakaidista < Spanish paracaidista)	0	parakáida/ parakáyda [< Spanish paracaídas]	'parachute'	2 Sp
33.	pasyaléro 'gadabout, wanderer, flaneur'	0	pasyál [< Spanish pasear 'to take a walk; to go for a ride']	'stroll; taking a walk; a walk for pleasure'	5 Sp
34.	panitikéro 'bookman; member of Panitikan society'	0	pánitik(án)	'literature; Panitikan literary society'	440 Tag
35.	sabungéro 'fan/frequent participant of cockfight'	47	sábong	'cockfight'	91 Tag
36.	salamangkéro 'conjurer; wizard'	42	salamángka [< Spanish salamanca 'cave for sorcery']	'conjuring; magic; sleight of hand'	46 Sp
37.	satsatéro/a 'chatterbox; scandalmonger, m/f'	0	satsát	'idle talk; gossip'	15 Tag
38.	sorbetéro 'ice cream vendor' (cf. Spanish vendedor de hielo)	3	sorbétes [< Spanish sorbete 'sherbet; iced drink']	'ice cream'	22 Sp
39.	stiréro 'teaser; cheater; prankster' (slang)	0	N/A	English (to) stir	0 Eng
40.	tinahéro 'producer/seller of tinaha jars'	0	tináha	'earthen jar for water; 12,5 gallon liquid measure'	0 Tag
41.	tsineléro '1. producer/seller of slippers; 2. homebody'	0	tsinélas [< Spanish chinelas, pl]	'slipper(s)'	142 Sp
42.	tubéro 'plumber, pipe fitter' (cf. Tagalog plomero < Spanish)	13	túbo [< Spanish tubo]	'tube, pipe'	ab. 56** Sp
43.	umbagéro 'pugnacious; prone to beat up'; 'brave man' (Rachkov 2012)	8	umbág	'a punch'	0 Tag
44.	usiséro/-a 'very inquisitive person, m/f'	41	usísa? [< Spanish ocioso 'idle; pointless'] *	'inquiry; examination'	23 Sp

TABLE 9.13 Hybrids with *-ero/a* in contemporary Tagalog (*cont.*)

No	Tagalog hybrid formation	# of tokens in LC	Tagalog simplex word	# of tokens in LC	SL
45.	utangéro/-a 'one who often makes debts, m/f' (neg.); 'debtor, m/f?' (Rachkov 2012)	2	útang	'debt'	1441 Tag

PATTERN ADOPTED FROM BAKKER AND HEKING (2012: TABLE 7)

Abbreviations: Ch – Chinese (incl. dialect), Eng – English, f – feminine, LC – Leipzig Corpus, Mal – Malay, m – masculine, Sp – Spanish, Mex Sp – Mexican Spanish, SL – source language, ? – origin uncertain.

* The simplex forms *usísa?* and *osyóso* are both from Spanish *ocioso* 'idle'. The difference in meaning, phonetics, number of derivatives, and frequency (22 vs. 2 tokens in LC) allow us to assume that the Spanish lexeme has been borrowed twice, with *usísa?* adopted at an earlier stage of Spanish colonization than *osyóso*.

** Due to the ambiguity of the type *tubo* in the LC, comprising the homonyms 'pipe, tube', 'born', 'profit, income' and 'sugarcane', the quantity of tokens for 'pipe, tube' in the first 250 entries has been counted manually (38 tokens), and an average of such tokens for the total 376 entries with *tubo* has been estimated (56.4).

Bibliography

- Aikhenvald, Alexandra Y. (2007). Grammars in contact: a cross-linguistic perspective. In: Aikhenvald, Alexandra Y., and Dixon, R.M.W. (eds.), *Grammars in contact: A cross-linguistic typology*. Explorations in Linguistic Typology, 4. Oxford: Oxford University Press, pp. 1–66.
- Aikhenvald, Alexandra Y. (2012). 'Invisible' loans: how to borrow a bound form. In: Johanson, Lars, and Robbeets, Martine (eds.), *Copies versus cognates in bound morphology*. Brill's Studies in Language, Cognition and Culture, 2. Leiden: Brill, pp. 167–185.
- Alcántara y Antonio T. (1999). *Mga Hispanismo sa Filipino (Batay sa Komunikasyong Pangmadla ng Filipinas: Pagaaral Lingguwistiko)*. Quezon City: UP Diliman.
- Baayen, R. Harald. (2008). Corpus linguistics in morphology: morphological productivity. In: Lüdeling, Anke, and Kytö, Merja (eds.), *Corpus Linguistics. An international handbook*. Berlin: Mouton De Gruyter, pp. 900–919.
- Bakker, Dik, and Hekking, Ewald. (2012). Constraints on morphological borrowing: Evidence from Latin America. In: Johanson, Lars, and Martine Robbeets (eds.), *Copies versus cognates in bound morphology*. Brill's Studies in Language, Cognition and Culture, 2. Leiden: Brill, pp. 187–220.
- Baklanova, Ekaterina. (2004). Interference in Tagalog as a Result of Borrowing. *Pilipinas. A Journal of Philippine Studies*, 42: 95–116.

- Baklanova, Ekaterina. (2016). On marginal gender in Tagalog: A case study. *Proceedings on XII LESEWA*: 25–33. Moscow: Yazyki Narodov Mira.
- Baklanova Ekaterina. (2017). On contact-induced changes in modern Tagalog: A case study. In: Ogloblin Alexander, Yanson Rudolf, Dmitrenko Sergey (eds.), *Intercivilizational Contacts In South-East Asia: Historical Perspectives and Globalization*. Saint Petersburg: Nestor-Istoriya, pp. 329–359.
- Batchelor, R.E., and San José, Miguel Ángel. (2010) *A Reference Grammar of Spanish*. Cambridge: Cambridge University Press.
- Blake, Frank R. (1925). *A grammar of the Tagalog language, the chief native idiom of the Philippine Islands*. New Haven, Conn.: American oriental society. URL: <https://archive.org/details/agy9509.0001.001.umich.edu/page/80/mode/2up>. Accessed on December 12, 2020.
- Bloomfield, Leonard. (1917). *Tagalog Texts with Grammatical Analysis*. Urbana: University of Illinois Press.
- Blust, Robert and Trussel, Stephen. (2010). *Austronesian Comparative Dictionary*, web edition. 2010 (rev. 2016). URL: www.trussel2.com/ACD. Accessed on July 2, 2020.
- Blust, Robert. (2013). *The Austronesian Languages*, revised edition/SEAMLES Series. Canberra: Australian National University. URL: <https://openresearch-repository.anu.edu.au/bitstream/1885/10191/6/Blust-2013-AustronesianLanguages.pdf> Accessed on 14 March 2021.
- Bowen, J. Donald (1971). Hispanic languages and influence on Oceania. In: Sebeok, Thomas A. (ed.), *Current Trends in Linguistics 8, pt. 2*. The Hague: Mouton, pp. 938–952.
- Bright, William. (1998). Social Factors in Language Change. In: Coulmas, Florian (ed.), *The Handbook of Sociolinguistics* (Blackwell Handbooks in Linguistics 4). Oxford & Cambridge, MA: Blackwell, pp. 81–91.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5): 425–455.
- Calderón, Sofronio G. (1915). *Diccionario Ingles-Español-Tagalog (Con partes de la oración y pronunciación figurada)*. Primera edición. Manila: Librería y papelería de J. Martínez.
- Casparis, J.G. de (1997). *Sanskrit loan-words in Indonesian: An annotated check-list of words from Sanskrit in Indonesian and Traditional Malay*. Jakarta: Universitas Katolik Indonesia Atma Jaya.
- Chamoreau, Claudine. (2012). Spanish diminutive markers -ito/-ita in Mesoamerican languages. A challenge for acceptance of gender distinction. In: Vanhove, Martine, Stolz, Thomas, and Urdze, Aina (eds.), *Morphologies in Contact* (Studia Typologica 10). Berlin & Boston: Akademie Verlag, pp. 71–92.
- De Guzman, V.P. (1996). Lexical categories in Tagalog. *Pan-Asiatic Linguistics: Proceed-*

- ings of the Fourth International Symposium on Languages and Linguistics, vol. 1.* Salaya: Mahidol University, pp. 307–396.
- DRAE. (2014). *Diccionario de la Lengua Española*. 23rd ed. Madrid: Real Academia Española. URL: <https://dle.RAE.es/>. Accessed Jan-Feb 2020.
- Dworkin, Steven N. (2012). *A History of the Spanish Lexicon: a Linguistic Perspective*. Oxford: Oxford University Press on Demand.
- English, Leo James. (1987). *Tagalog-English Dictionary*. Manila: National Book Store.
- Fernández, Mauro (2013). The Representation of Spanish in the Philippine Islands. In: Del Valle, José (ed.), *A Political History of Spanish: The Making of a Language*. Cambridge: Cambridge University Press, pp. 364–379.
- Fishman, J.A. (1967). Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues*, 23(2): 29–38.
- Gardani, Francesco. (2008). *Borrowing of inflectional morphemes in language contact*. Frankfurt/M.: Peter Lang.
- Gómez Rendón Jorge A. (2008). *Typological and social constraints on language contact: Amerindian languages in contact with Spanish, vols. 1 & 2*. Doctoral dissertation, University of Amsterdam.
- Goulet, Rosalina Morales. (1971). English, Spanish and Tagalog: A study of grammatical, lexical and cultural interference. *Philippine Journal of Linguistics*, Special Monograph Issue, no. 1.
- Gramática. (2011). *Nueva gramática de la lengua española. Morfología. Sintaxis. Fonética y fonología*, ed. by Ignacio Bosque. Madrid: Real Academia Española/Asociación de Academias de la Lengua Española.
- Haspelmath, Martin. (2009). Lexical borrowing: Concepts and issues. In: Haspelmath, Martin, and Tadmor, Uri, *Loanwords in the world's languages*. Berlin & New York: Mouton De Gruyter, pp. 35–54.
- Hay, Jennifer. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, 39: 1041–1070.
- Haugen, Einar. (1969). *The Norwegian language in America. A study in bilingual behavior*. Bloomington: Indiana University Press.
- Himmelman, Nikolaus P. (2009). Notes on Tagalog nominalism. *Theoretical Linguistics*, 35(1): 115–123.
- Jurafsky, Daniel. (1996). Universal tendencies in the semantics of the diminutive. *Language*, 72(3): 533–578.
- Lipski, John M., with Mühlhäusler, P., and Duthin, F. (1996). Spanish in the Pacific. In: Wurm, Stephen A., Mühlhäusler, Peter, and Tryon, Darrell T. (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*. Berlin & New York: Mouton de Gruyter, pp. 271–298.
- Lopes, José M., and Bensley, Edward R. (1895). *Nuevo diccionario inglés-español y español-inglés*. Paris: Garnier. URL: <https://archive.org/details/nuevodiccionario00lopeoft>. Accessed on 2–10 November 2020

- López, Cecilio. (1965). The Spanish overlay in Tagalog. *Lingua*, 14: 467–504.
- Matras, Yaron. (2007). The borrowability of structural categories. In: Matras, Yaron, and Sakel, Jeanette (eds.), *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter, pp. 31–73.
- Milroy, Lesley, and Milroy, James. (1992). Social network and social class: Toward an integrated sociolinguistic model. *Language in Society*, 21(1): 1–26.
- M-W – Monier-Williams, Monier. (1899). *A Sanskrit-English Dictionary, new edition*. Delhi, Varanasi: Motilal Banarsidass.
- Muysken, Pieter. (2012). Spanish affixes in the Quechua languages: A multidimensional perspective. *Lingua*, 122(5): 481–493.
- Ortograpiyang Pambansa. (2013). *Unang Edisiyon*. Manila: Komisyon sa Wikang Pambansa.
- Payne, Thomas E. (1997). *Describing morphosyntax: A guide for field linguists*. Cambridge & New York: Cambridge University Press.
- Potet, Jean-Paul G. (2016). *Tagalog Borrowings and Cognates*. Raleigh: Lulu Press.
- Quilis, Antonio, and Casado-Fresnillo, Celia. (2008). *La lengua Española en Filipinas: Historia, situación actual. El Chabacano. Antología de textos*. Madrid: Consejo Superior de Investigaciones Científicas.
- Rachkov, Gennadij E. (1981). *Введение в Морфологию Современного Тагальского Языка [Introduction to the morphology of Modern Tagalog]*. Leningrad: Leningrad State University.
- Rachkov, Gennadij E. (2012). *Новый тагальско-русский словарь [New Tagalog-Russian Dictionary]*, vols. 1 & 2. Saint Petersburg: Saint Petersburg State University.
- Rainer, Franz. (2011). *Spanische Wortbildungslehre*. Tübingen: Max Niemeyer Verlag.
- Rau, Der-Hwa V. (1992). Language contact in the Philippines. *Philippine Journal of Linguistics*, 23(1–2): 91–109.
- Rodríguez-Ponga, Rafael. (2009). *Del español al Chamorro. Lenguas en contacto en el Pacífico*. Madrid: Ediciones Gondo.
- Schachter, Paul, and Otones, Fe T. (1972). *Tagalog reference grammar*. Berkeley: University of California Press.
- Seifart, Frank. (2015). Direct and indirect affix borrowing. *Language*, 91(3): 511–532.
- Seifart, Frank. (2017). Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics*, 7(3): 389–343.
- Seifart, Frank. (2020). *AfBo: A world-wide survey of affix borrowing*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Serrano Laktaw, P. (1889). *Diccionario Hispano-Tagalog*. Manila: Estab. Tipografico La Opinion.
- Shkarban, Lina I. (1995). *Грамматический Строй Тагальского Языка [Tagalog grammatical structure]*. Moscow: Vostochnaya literatura.
- Shkarban, Lina. (2004). On Tagalog morphology in the context of parts-of-speech typo-

- logy. In: Alieva, Natalia (ed.), *Malay and Indonesian Studies*, Issue XVI, pp. 314–324.
- Sippola, Eeva. (2020). Contact and Spanish in the Pacific. In: Hickey, Raymond (ed.), *The handbook of language contact*. Hoboken, NJ: Wiley-Blackwell, pp. 453–469.
- Steinkrüger, Patrick O. (2008). Hispanisation processes in the Philippines. In: Stolz, Thomas, Bakker, Dik, and Salas Palomo, Rosa (eds.), *Aspects of language contact. New theoretical, methodological and empirical findings with special focus on Romanisation processes*. Berlin: Mouton de Gruyter, pp. 203–236.
- Stolz, Thomas, Bakker, Dik, and Salas Palomo, Rosa (eds.). (2008). *Aspects of language contact. New theoretical, methodological and empirical findings with special focus on Romancisation processes*. Berlin: Mouton de Gruyter.
- Stolz, Thomas. (2012). Survival in a niche. On gender-copy in Chamorro (and sundry languages), in: Vanhove, M., Stolz, T., Urdze, A. (eds.), *Morphologies in Contact. Studia Typologica Series (Book 10)*. Berlin, Boston: Akademie Verlag, pp. 93–140.
- Tadmor, Uri. (2009). Loanwords in Indonesian. In: Haspelmath, Martin, and Tadmor, Uri (eds.), *Loanwords in the world's languages: A comparative handbook*. Berlin: De Gruyter Mouton, pp. 686–716.
- Thomason, Sarah Grey, and Kaufman, Terrence. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley & LA: University of California Press.
- Thompson, Roger M. (2003). *Filipino English and Taglish. Language switching from multiple perspectives*. Amsterdam & Philadelphia: John Benjamins.
- Winford, Donald. (2003a). *An introduction to contact linguistics*. Oxford: Blackwell.
- Winford, Donald. (2003b). Contact-induced changes: Classification and processes. *OSUWPL* 57: 129–150.
- Wolff, John U. (1973). The character of borrowings from Spanish and English in the languages of the Philippines. *Philippine Journal of Linguistics*, 4(1): 72–82.
- Wolff, John U. (1993). Why roots add the affixes with which they occur: A study of Tagalog and Indonesian adjective formations. In: Reesink, Ger P. (ed.), *Topics in descriptive Austronesian linguistics*. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden, pp. 217–244.
- Wolff, John U. (2001). The influence of Spanish on Tagalog. In: Zimmermann, Klaus, and Stolz, Thomas (eds.), *Lo propio y lo ajeno en las lenguas austronésicas y amerindias*. Frankfurt a.M./Madrid: Vervuert/Iberoamericana, pp. 233–253.

Data Sources

- Almarío R. 1910. *Ang Mánanayaw* [short story], in: *Project Gutenberg E-books*, URL: <https://www.gutenberg.org/ebooks/>. Accessed in Jan-Feb 2020.
- Calderón S.G. 1907. Dating Pilipinas, in: *Project Gutenberg E-books*, URL: <https://www.gutenberg.org/ebooks/>. Accessed in Jan-Feb 2020.

- Cruz H. 1922. Kartilyang Makabayan ... [essay], in: *Project Gutenberg E-books*, URL: <https://www.gutenberg.org/ebooks/>. Accessed in Jan-Feb 2020.
- Francia J.R. 1919. *Hindi biro!* ... [short story], in: *Project Gutenberg E-books*, URL: <https://www.gutenberg.org/ebooks/>. Accessed in Jan-Feb 2020.
- LC—Goldhahn D., Eckart T. & Quasthoff U. (2012). *Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages*. In: Proceedings of the 8th International Language Resources and Evaluation (LREC'12), 2012. Available at: https://corpora.uni-leipzig.de/en?corpusId=ttl_community_2017. Accessed in Feb 2020.
- Mariano P. 1906. *Juan Masili* ... [short story], in: *Project Gutenberg E-books*, URL: <https://www.gutenberg.org/ebooks/>. Accessed in Jan-Feb 2020.
- Rivera J.M. 1910. *Sa Tabi ng Bangin* ... [play], in: *Project Gutenberg E-books*, URL: <https://www.gutenberg.org/ebooks/>. Accessed in Jan-Feb 2020.

The Structural Consequences of Lexical Transfer in Ibatan

Maria Kristina S. Gallego

Introduction*

In accounting for contact-induced language change, it is argued that different linguistic materials have varying degrees of transferability, where some tend to be transferred more easily than others. The general consensus in the field is that lexicon is highly transferable in contact situations, whereas structural transfer (i.e. morphology and phonology) is less likely to occur.¹

This paper investigates a particular contact-induced change in the morphology of Ibatan, an Austronesian language spoken in the far north of the Philippines. In particular, it focuses on the paradigm of the durative verbal prefix *pag-*, which is traced to Ilokano, the main language in contact with Ibatan. What are the mechanisms and scenarios that led to the development of a non-native² set of verbal prefixes which exists parallel to the native paradigm in the language? The main argument taken here is that this current morphological structure reflecting both native and non-native verbal morphology is an outcome of layers of contact-induced language change driven by different agents with varying degrees of (psycholinguistic) dominance in Ibatan.

Explaining contact-induced outcomes requires us to determine the processes that have shaped the language, and this means linking contact outcomes to the sociolinguistic contexts of the multilingual individuals and community. In this paper, explanations for the development of a non-native morphological paradigm in Ibatan, a phenomenon that has been argued to be dispreferred in situations of language contact, are grounded in past and present patterns of language dominance, both at the levels of the individual and the com-

* This paper is adapted from the thesis Gallego (2022b).

1 Such scales or hierarchies can be seen as early as Whitney (1881), to Haugen (1950), Weinreich (1953), and Thomason and Kaufman (1988).

2 The term *non-native* is used in this paper to describe contact-induced features in Ibatan. It is used in its neutral sense, and unless otherwise specified, refers to features from any source language in contact with Ibatan.

munity. Such context-driven frameworks, such as van Coetsem (2000) focusing on the individual, Thomason and Kaufman (1988) on the community, and Muysken (2010) on different scenarios of language contact, allow for a nuanced treatment of contact-induced outcomes, which can ultimately provide a more satisfactory account of the phenomenon than what have been proposed in the early literature (that is, context-free, language-internal approaches to contact).

This paper begins with a detailed description of the dynamic sociolinguistic landscape of the Ibatan community (Section 1), as well as an overview of the verbal morphology of both Ilokano and Ibatan (Section 2). Data on the distribution and current usage of the parallel durative verbal paradigms in Ibatan (Section 3) are based on the Ibatan dictionary by Maree, et al. (2012), supplemented by recordings of naturalistic speech and interviews with speakers gathered during the author's 2018 fieldwork. Explanations behind the development of the parallel paradigms in the language are grounded on the socio-historical changes that happened in the community, following context-based frameworks for studying language contact (Section 4).

1 The Ibatans of Babuyan Claro

Babuyan Claro (or Babuyan) is an island community in the far north of the Philippines with a dynamic sociolinguistic landscape that has been shaped by its history. At present, the majority of people on Babuyan Claro are multilingual in at least three languages: Ibatan (IVB), Ilokano (ILO), and Filipino (FIL).³ Ibatan, the local language of Babuyan Claro and the smallest of the three, belongs to the Batanic subgroup of Philippine languages along with Itbayaten, Ivatan (with dialects Ivasay and Isamorong), and Yami (also known as Tao) (Figure 10.1). Ilokano, the main language in contact with Ibatan, is a Northern Luzon⁴ language, and it is the trade language of the Babuyan group of islands (to which the community of Babuyan Claro administratively belongs) and the regional lingua franca of northern Luzon. Lastly, Filipino is the national language of the Philippines, and is the main language used in print and broadcast media in the country.

3 In this paper, Filipino is used to refer to the language as it is the term mandated in the Philippine constitution, but at the same time, acknowledging that this language is primarily based on Tagalog, a Greater Central Philippine language.

4 Also known as Cordilleran.



FIGURE 10.1 The location of Ibatan

In terms of linguistic features, the three languages share significant similarities in lexicon and structure because of their common ancestry within the Malayo-Polynesian branch of the Austronesian language family.⁵ At the same time, however, a great number of features across phonology and morpho-syntax make the languages distinct from each other, and they are not at all mutually intelligible. These are part of the evidence to argue for the separation of the three languages into three different subgroups of Philippine languages.

As for the people of Babuyan Claro, the first of the founding families of the community were of Batanic ancestry who were shipwrecked on the island around 1869 in their attempt to return to Batanes after having been relocated to the Babuyan islands. Soon after, two more groups, but this time of Ilokano ancestry, arrived on Babuyan Claro. For the next 50 years or so, the population on Babuyan Claro grew with the arrival of small groups of people from both Batanic- and Ilokano-speaking backgrounds (Maree 1982, Maree, 2005).

While ethnographic evidence suggests that these first families generally kept the two ethnolinguistic lines separate (Maree 1982), the harsh conditions on the island also required the families to rely on each other, particularly in terms of economic and livelihood activities. There were also some cases of marriage across linguistic groups, especially because the population on the island at that time was very small. This setting must have fostered the maintenance of bilingualism in the community in these initial years.

The general tendency to maintain ethnolinguistic boundaries in the community has led to the geographical distinction between Ibatan and Ilokano-speaking networks. While residential settlements are scattered across the island, the greatest density can be found along the southern coast of Babuyan Claro, and this is divided into *daya* 'east' and *laod* 'west'. This geographic distinction has come to coincide with social networks that reflect different patterns of language choices and uses. Families who reside in *daya* have acquired both Ibatan and Ilokano in their childhood, and they show greater affinity towards Ibatan. They are referred as **Ibatan-dominant early bilinguals** in this paper. In contrast, a small but significant network of families situated in *laod*, who like-

5 It has long been debated whether there is a single Philippine subgroup of languages within Malayo-Polynesian. The languages spoken in the Philippines share significant similarities but scholars such as Ross (2005), and Smith (2017) question the integrity of the subgroup. See Blust (2019, 2020), Liao (2020), Reid (2020), Ross (2020), and Zorc (2020) for the most recent discussion of this debate.

wise have acquired both languages in their childhood, tend to prefer the use of Ilokano as their everyday language. They constitute the **Ilokano-dominant early bilinguals** referred in this paper.

Around the 1970s, Ilokano, being the language for wider communication in northern Luzon, became more prominent on Babuyan Claro as the community became more integrated within the administrative region of Calayan. During this time, Ilokano was the main language for administration, religion, and education on the island. This had dramatic effects on the patterns of multilingualism of the community, where the domains in which Ibatan was used became more limited, thus severely threatening the vitality of the language.

Starting in the 1980s, Babuyan Claro witnessed further changes in its socio-political landscape, the most pivotal of which is the granting of the Certificate of Ancestral Domain Title⁶ to the Ibatans in 2007. This and other significant changes reversed the expansion of Ilokano, and this is clearly reflected in the more vigorous use of Ibatan even in the domains outside the home. Currently, there is also an increasing number of immigrants on the island, typically from Ilokano-speaking backgrounds, who are learning the Ibatan language as adults. They tend to have varying degrees of proficiency in Ibatan depending on the networks of speakers with whom they frequently interact. This final group of Ibatan speakers are characterized as **Ilokano-dominant late bilinguals** in this paper.

Finally, in recent years, Babuyan Claro has come to be more integrated within the larger nation state. This means that the influence of Filipino has become more pronounced in the community as well. In addition to Filipino being taught formally in basic to higher education, the Ibatan people are able to travel to and from the mainland more frequently, which means greater use of and exposure to Filipino. This has contributed to further changes in the patterns of language use for some speakers, where Filipino, rather than Ilokano, has now become their preferred second language.

The patterns of multilingualism on the island are evidently shaped by the changing socio-political landscape of Babuyan Claro and the larger region to which it belongs. These changes comprise different phases in the history of Babuyan Claro, summarized in (1).⁷

6 This gives the Ibatan people collective rights to natural resources on Babuyan Claro, and this was granted by the National Commission on Indigenous Peoples of the Philippines through the Indigenous Peoples Rights Act of 1997 (Ebarhard, Simons & Fennig 2022).

7 For a detailed account of the linguistic landscape of Babuyan Claro, see Gallego (2020).

(1)	1870s	Phase 1	The arrival of the first Ibatans
	1900s	Phase 2	The emergence of the <i>daya-laod</i> networks
	1970s	Phase 3	The rise of Ilokano
	1980s	Phase 4	The renewed vitality of Ibatan
	<i>ongoing</i>	Phase 5	The influx of Ilokano immigrants
	<i>ongoing</i>	Phase 6	The increasing influence of Filipino

While these phases appear to constitute distinct periods in the history of Babuyan Claro, they are in no way discrete and tend to overlap. The socio-political and linguistic contexts of the community remain dynamic to this day. Thus, the ongoing dynamics of language use and the social value attached to the three languages are in tension with each other. The changing nature of the socio-political and linguistic landscape of Babuyan Claro therefore means that an individual's patterns of language choices and uses may change within their lifetime. At the same time, some patterns of language use can become widespread across the community, and this is how language change (here we put particular focus on contact-induced change) proceeds.

Given the vast difference between Ibatan and Ilokano in terms of social dominance, the relationship between the two languages can best be described as a one-way street. Ilokano is the bigger language, used in a larger area of mainland northern Luzon, and it currently has about 6,482,100 users. In contrast, Ibatan is only mainly used on Babuyan Claro by about 1,240 to 3,000 users (according to Ebarhard, et al. (2022) and the author's fieldwork). Thus, in terms of contact-induced outcomes, Ibatan has shown little to no impact on the overall system of Ilokano.⁸

In contrast, Ibatan is characterized by Ilokano-influenced linguistic features which set it apart from the rest of the Batanic languages, not only in terms of the lexicon, but also in more structured aspects of the language, such as morphology. To illustrate, Ibatan has a significantly high proportion of loanwords in its lexicon. A preliminary investigation following the Loanword Typology Project by Haspelmath and Tadmor (2009) shows a 44% proportion of loanwords⁹

8 It is a different matter, however, when talking about how the Ibatans use Ilokano as their second language, where it is expected that they would show Ibatan features in their use of Ilokano. This is particularly evident in phonology, where Ilokano-dominant speakers would describe the Ilokano spoken by the Ibatans as having a clearly "Ibatan accent". While this is an interesting study in its own right, it is well outside the scope of this study.

9 From various source languages such as Spanish, English, and Filipino, but with a huge proportion of loanwords coming from Ilokano.

TABLE 10.1 Native and non-native affixes and stems

Affix function	Native affixes + native stems		Non-native affixes + non-native stems	
Durative	<i>may-tenek</i>	'stand'	<i>mag-bayad</i>	'pay'
Nominalization	<i>pay-tolas</i>	'write'	<i>pag-sorat</i>	'write'
Pretense	<i>may-sin-CV-asnek</i>	'shame'	<i>magin-CV-singpet</i>	'virtue'

MAREE 2007:173

in Ibatan, which places it as a high borrower in their scale. Beyond vocabulary, Ibatan has also been heavily influenced by Ilokano in terms of structural features. Maree (2007) identifies competing native and non-native affixes in the language, some of which are presented in Table 10.1.

In many cases, non-native affixes occur with non-native stems, constituting complex loanwords, for instance, *mag-bayad* consisting of the non-native prefix *mag-* 'DUR'¹⁰ and Ilokano stem *bayad* 'pay'. However, there are also a few cases of hybrid formations, or non-native prefixes occurring with native stems. In terms of accounting for this durative paradigm in Ibatan, its general distribution as part of complex loanwords appears to be a straightforward outcome of lexical transfer, but the presence of hybrid formations demands a detailed investigation of the various processes governing language contact, which can be linked to the known history of the Babuyan Claro community. That is, the changing patterns of multilingualism in the community, which began when the first families came to Babuyan Claro in 1869, are argued to drive the layers of contact-induced change we see in Ibatan.

2 The Verbal Morphology of Ilokano and Ibatan

In understanding the consequences of language contact, it is necessary to distinguish which features are non-native in a language, and consequently trace the source of such features. In the case of Ibatan and Ilokano, the two languages share a number of similar features because of shared ancestry, which makes teasing apart native from non-native features more challenging.

¹⁰ See Appendix for the list of glossing abbreviations.

In terms of morphosyntax, both languages have a Philippine-type system that is typically described in terms of focus (cf. Reid and Liao 2004 and Liao 2004), or more recently, voice (cf. Wouk and Ross 2002, Riesberg 2014, etc.). This voice system is realized as the affixes on the verb in relation to the role of voice-selected argument in a sentence, which can either be the actor or the undergoer, the latter further categorized into patient, locative, and circumstantial.¹¹ For actor voice, there are further sets of affixes that encode additional semantic features on the predicate, namely inchoative (or punctual), distributive (which implies multiple activities), and durative (which is also associated with reflexive and reciprocal senses). In addition to voice, verbal affixes encode mood and aspect. Mood can either be irrealis (events that are yet to happen, as in future events) or realis (events that are non-future, as in present, past, and habitual activities). Aspect can be perfective (completed events) or imperfective (events that are not yet completed, as in progressive or habitual events) (cf. Reid and Liao 2004).

This section gives a brief description of the verbal morphology of Ilokano and Ibatan, and sets out how the parallel durative paradigm seen in Ibatan can be traced back to Ilokano.

2.1 *Ilokano*

Verbs in Ilokano are marked with voice, aspectual, and mood distinctions by means of different sets of affixes (Table 10.2). For actor voice, the affixes may either be <um> 'INC', *mang-* 'DIST', or *ag-* 'DUR'. Undergoer voices are marked with the suffix *-en*¹² for patient, *-an* for locative, and *i-* for circumstantial. As for aspect, perfective is marked by the infix <in>, and imperfective is typically marked by reduplicating the first CVC¹³ sequence of the stem. For the irrealis mood, Ilokano shows the optional use of the enclitic *=(n)to*, which is a variant of the adverb *into* that indicates future time.

These grammatical specifications on the verb are marked by combining the verbal affixes. To illustrate, the verb stem *gatang* 'buy' marked with <um> for actor voice (inchoative), in combination with the CVC reduplication for realis imperfective, yields the form *g<um>at-gatang* '<AV.INC>IPFV~buy'. As for marking realis perfective, the aspectual infix <in> comes first before the voice infix <um>, and this ordering of the verbal affixes in Ilokano has led

11 Some grammars specify another category, that is, benefactive, typically derived with the circumfix *i...-an* (cf. Reid and Liao 2004).

12 Where <e> is pronounced as a high, central vowel (Rubino 2000: xiii).

13 Sometimes CV, depending on the stem.

to the syncope of the vowel *u* in $\langle um \rangle$, and the subsequent assimilation of *n* in $\langle in \rangle$, leading to the form $\langle im \rangle \langle m \rangle$. Thus, marking the same verb *gatang* 'buy' with actor voice, realis perfective yields the form $g\langle im \rangle \langle m \rangle atang$ ' $\langle PFV \rangle \langle AV.INC \rangle$ buy'

For distributive and durative verbs, marking aspectual distinctions does not reflect the same level of agglutination as inchoative verbs. In particular, the affixes used to mark realis perfective are portmanteau forms that combine the infix $\langle n \rangle$ (a reduction of $\langle in \rangle$) and the voice prefixes *mang-* for distributive and *ag-* for durative. This leads to the perfective forms *nang-* and *nag-* respectively. Realis imperfective and irrealis forms are more transparent, reflecting the expected combination of the voice and aspectual affixes. To illustrate these derivations, *takaw* 'steal' is derived in the actor voice distributive form as *mang-takaw* 'AV.DIST.NTRL-steal', *nang-takaw* 'AV.DIST.PFV-steal', *mang-tak-takaw* 'AV.DIST-IPFV~steal', and *mang-takaw=to* 'AV.DIST-steal=IRR'. *Surat* 'write' is derived in the actor voice durative form as *ag-surat* 'AV.DUR.NTRL-write', *nag-surat* 'AV.DUR.PFV-write', *ag-sur-surat* 'AV.DUR-IPFV~write', and *ag-surat=to* 'AV.DUR-write=IRR' (Rubino 2000: lxvii).

The forms *mang-* and *ag-* that mark actor voice distributive and durative are historically derived from a combination of the actor voice affix $\langle m \rangle$ (a reduction of $\langle um \rangle$) with the prefixes *pang-* and *pag-*. These latter prefixes carry the basic distributive and durative senses, and at present are also used to nominalize verb forms in Ilokano. These prefixes, moreover, are reflexes of Proto Malayo-Polynesian (PMP) $*paN-$ and $*paR-$ respectively, and the resulting portmanteau forms $*maN-$ and $*maR-$ are also reconstructed for PMP (Wolff 1973:72–74). The realis neutral form *ag-* in Ilokano, shows a further reduction of PMP $*maR-$ to its current form *ag-*. The Ilokano verbal morphology is summarized in Table 10.2, with sample verbs to illustrate the various derivations discussed above.

TABLE 10.2 Ilokano verbal morphology

Ilokano	ACTOR				UNDERGOER			
	INCHOATIVE	DISTRIBUTIVE	DURATIVE	PATIENT	LOCATIVE	CIRCUMSTANCIAL		
	<i>gatang</i> 'to buy'	<i>takaw</i> 'to steal'	<i>surat</i> 'to write'	<i>surat</i> 'to write'	<i>punas</i> 'to wipe'	<i>kabil</i> 'to put'		
REALIS	NEUTRAL							
	$\langle um \rangle X$ $g \langle um \rangle atang$	$mang-X$ $mang-takaw$	$ag-X$ $ag-surat$	$X-en$ $surat-en$	$X-an$ $punas-an$	$i-X$ $i-kabil$		
	PERFECTIVE							
	$\langle im \rangle \langle m \rangle X$ $g \langle im \rangle \langle m \rangle atang$	$nang-X$ $nang-takaw$	$nag-X$ $nag-surat$	$\langle in \rangle X$ $s \langle in \rangle urat$	$\langle in \rangle X-an$ $p \langle in \rangle unas-an$	$in-X$ $in-kabil$		
	IMPERFECTIVE							
	$\langle um \rangle CVC-X$ $g \langle um \rangle at-gatang$	$mang-CVC-X$ $mang-tak-takaw$	$ag-CVC-X$ $ag-sur-surat$	$CVC-X-en$ $sur-surat-en$	$CVC-X-an$ $pun-puna-san$	$i-CVC-X$ $i-kab-kabil$		
IRREALIS								
	$\langle um \rangle X = ((n)to)$ $g \langle um \rangle atang = to$	$mang-X = ((n)to)$ $mang-takaw = to$	$ag-X = ((n)to)$ $ag-surat = to$	$X-en = ((n)to)$ $surat-en = to$	$X-an = ((n)to)$ $punas-an = to$	$i-X = ((n)to)$ $i-kabil = to$		
NOMINAL								
		$pang-X$ $pang-takaw$	$pag-X$ $pag-surat$					

2.2 *Ibatan*

Verbs in Ibatan are marked with the same distinctions as those discussed for Ilokano, but by different sets of affixes (Tables 10.4 and 10.5). Given the genetic relationship between the two languages, a number of affixes are identical in the two languages, namely the undergoer voice affixes *-en*¹⁴ 'PV', *-an* 'LV', and *i-* 'CV', as well as the actor voice distributive prefix *maN*.¹⁵ The actor voice infix $\langle om \rangle$ in Ibatan is also phonologically similar to Ilokano $\langle um \rangle$, where *o* is pronounced as a high, back, rounded vowel, but it is represented orthographically with the vowel *o*. Ibatan also shows the use of the future adverb *anchi* as the enclitic $=(a)nchi$ to optionally mark irrealis, parallel to the development of Ilokano *into*.

Ibatan differs from Ilokano in terms of the ordering of the aspectual and voice affixes. Where Ilokano reflects the sequence $\langle im \rangle$ 'PFV' + $\langle m \rangle$ 'AV', Ibatan shows the reverse order, that is, $\langle om \rangle$ 'AV' + $\langle in \rangle$ 'PFV'. This sequence is actually a retention of the ancestral system reconstructed for PMP (Ross 2002), and the current ordering observed in Ilokano constitutes an innovation shared among many Northern Luzon languages (Reid 1992).

What makes Ibatan unique, not only in comparison to Ilokano but also to its sister Batanic languages, is its two distinct but parallel paradigms of verbal affixes, where the use of a particular set typically depends on the etymology of the stem. This is observed in the paradigms for actor voice durative and realis imperfective. For marking durative verbs, Ibatan reflects two sets of prefixes, namely *pay-* (along with *may-* 'AV.DUR.NTRL' and *nay-* 'AV.DUR.PFV') and *pag-* (along with *mag-* 'AV.DUR.NTRL' and *nag-* 'AV.DUR.PFV'). For marking realis imperfective, Ibatan shows different reduplication patterns, namely CV(y)/CVCV and CVC. Native Ibatan stems are marked with the paradigms *pay-* for 'DUR' and CV(y) or CVCV for 'IPFV' (Table 10.4). As an example, the native Ibatan verb *disna* 'sit' occurs as *may-disna* for 'AV.DUR.NTRL-sit' and *may-di~disna* for 'AV.DUR-IPFV~sit'. In contrast, loanwords, typically of Ilokano origin (but also stems from other source languages (SL), such as Filipino, English, and Spanish), are generally marked with *pag-* for 'DUR' and CVC for 'IPFV' (Table 10.5). To illustrate, the Ilokano loanword *kalap* 'fish' is derived as *mag-kalap* for 'AV.DUR.NTRL-fish', and *mag-kal~kalap* for 'AV.DUR-IPFV~fish'. The co-existence of these parallel paradigms in Ibatan is clearly an outcome of contact-induced change, where non-native stems are marked with non-native morphology. To further illustrate these parallel paradigms, (2a) and (2b) show the prefixes *nay-* and *nag-* marking native *abang* '(ride on a) rowboat' and non-native *lampitaw* '(ride on a) motorized boat' respectively.

14 Where $\langle e \rangle$ is pronounced as a high, central vowel, but slightly fronted compared to Ilokano (Maree 2005: 19).

15 The final nasal *N*- can be bilabial *m*, alveolar *n*, or velar *ng*, as it assimilates to the place of articulation of the following segment.

- (2) a. Native actor voice durative prefix *nay-* (Maree 2007:174)
Nayabang si adi a nangay do Calayan.
Nay-abang si adi a nangay do Calayan
 DUR-rowboat.IVB DET younger.sibling LK went DET Calayan
 ‘Younger sibling rode on a rowboat going to Calayan.’
- b. Non-native actor voice durative prefix *nag-* (Maree 2007:174)
Naglampitaw si adi a nangay do Calayan.
Nag-lampitaw si adi a nangay do Calayan
 DUR-motor.boat.ILO DET younger.sibling LK went DET Calayan
 ‘Younger sibling rode on a motorized boat going to Calayan.’

The two sets of durative prefixes in Ibatan can be traced from two sources, both descended from PMP **paR-*. The paradigm consisting of the forms *pay-* ‘DUR’, *may-* ‘AV.DUR.NTRL’, and *nay-* ‘AV.DUR.PFV’ are directly inherited, as evidenced by the final consonant *y*, which is the regular reflex of PMP **R* in the Batanic languages. The non-native paradigm consisting of the counterpart forms *pag-*, *mag-*, and *nag-* respectively is argued to be transferred from Ilokano, albeit with subsequent adaptation into the Ibatan system. Not only do the forms reflect *g* as the reflex of PMP **R*, a feature of Ilokano,¹⁶ but the distribution of the prefixes with mostly Ilokano stems clearly points to Ilokano as the source of this paradigm (see Sections 3 and 4).

This non-native durative paradigm has become regularized in Ibatan, and has come to apply generally to loanwords, including those from English, Filipino, and Spanish (Table 10.3). Its usage and distribution are discussed in detail in Section 3.

As mentioned, these parallel durative paradigms are a unique feature in Ibatan, which is not observed in other Batanic languages such as Ivatan, a closely related language spoken on Batan Island, Batanes. Both native *vidi* ‘return’ and Spanish *eroplano* ‘(ride an) airplane’ take the native verbal prefix *nay-* (3).

16 Ilokano in fact has two reflexes for PAN/PMP **R*, namely *r* and *g*. Blust (1991) characterizes this *g* in the language as the “stereotyped Philippine *g*,” where Ilokano, along with other Philippine languages, exhibit an irregular *g* reflex of **R* alongside the regular reflex of the consonant. Blust (1991) proposes that this is an outcome of the historical expansion of the Greater Central Philippine languages, which are languages that show *g* as the regular reflex of **R*. As an alternative explanation, Reid (personal communication) analyzes this irregular *g* reflex in Ilokano as an outcome of contact with Ibanag and other Cagayan Valley languages of the Northern Luzon subgroup which show *g* as the regular reflex of PMP **R*.

TABLE 10.3 Loanwords from different source languages occurring with *mag-*

Source	Derivation	Definition
English	<i>mag-pichor</i>	take a picture
Filipino	<i>mag-bak~bakla</i>	a man behaves like a woman
Ilokano	<i>mag-dayaw</i>	honour, praise
Spanish	<i>mag-tokar</i>	play music

(3) Ivatan: native *may-* with non-native stem

Nay-eroplano si Maria ta nayvidi du Basco. (elicited)

Nay-eroplano si Maria ta nay-vidi du Basco

DUR-airplane.SPA DET Maria because DUR-return.IVV DET Basco
 'Maria took the airplane because she returned to Basco.'

3 The Parallel Durative Paradigms of Ibatan

In their dictionary, Maree et al. (2012) indicate 1436 stems that can occur with the two sets of durative prefixes in Ibatan (Table 10.6). The vast majority of these stems follow the expected distribution, that is, either as native formations, where native stems occur with native morphology (513 stems or 35.72%), or as complex loanwords, where non-native stems, regardless of their source, occur with non-native morphology (755 stems or 52.58%).¹⁷ Among complex loanwords, the majority are traced back to Ilokano (485 of 755 stems, or 64.24%), followed by Spanish (248 stems, or 32.85%). Other SLs include English, Filipino, Chinese, and Ibanag.¹⁸

Such general distribution not only shows the relative influence of the different SLs in Ibatan in terms of the number of loanwords the different lan-

¹⁷ The remaining 168 stems reflect unexpected formations, discussed in Section 3.1.

¹⁸ The type of contact between Ibatan and the different SLs varies in terms of directness. Given the intense social contact between Ilokano and Ibatan, Ilokano has had more direct influence on Ibatan compared to other foreign SLs such as Spanish, English, and Chinese. That is, while one can expect that the Ibatan speakers are also proficient in Ilokano, they may not have such comparable proficiency in these other SLs. Their influence in Ibatan is thus minimal and is typically restricted within the lexicon, where, in fact, many of the loanwords have been transferred indirectly through another intermediate SL, typically Ilokano, and more recently, Filipino. This process also explains how the non-native durative paradigm has come to be extended to loanwords from these other foreign SLs.

TABLE 10.4 Ibatan verbal morphology for native stems

Ibatan: Native stems		ACTOR				UNDERGOER			
		INCHOATIVE	DISTRIBUTIVE	DURATIVE	PATIENT	LOCATIVE	CIRCUMSTANCIAL		
		<i>sayap</i> 'to fly'	<i>paso</i> 'to roast'	<i>disna</i> 'to sit'	<i>inom</i> 'to drink'	<i>tolas</i> 'to write'	<i>pangay</i> 'to put'		
REALIS	NEUTRAL	<om>X s<om>ayap	maN-X <i>mam-aso</i>	may-X <i>may-disna</i>	X-en <i>inom-en</i>	X-an <i>tolas-an</i>	i-X <i>i-pangay</i>		
	PERFECTIVE	<om><(i)n>X s<om><n>ayap	naN-X <i>nām-aso</i>	nay-X <i>nay-disna</i>	<in>X <in>inom	<in>X-an <i>t<in>olas-an</i>	in-X <i>in-pangay</i>		
	IMPERFECTIVE	<om>CVCV-X s<om>aya~sayap	maN-CVCV-X <i>mam-aso-paso</i>	may-CV(y)-X <i>may-di-disna</i>	CVCV-X-en <i>ino-inom-en</i>	CVCV-X-an <i>tola-tolas-an</i>	i-CVCV-X <i>i-panga-pangay</i>		
IRREALIS		<om>X=((a)nchi) s<om>ayap= <i>anchi</i>	maN-X=((a)nchi) <i>mam-aso=nchi</i>	may-X=((a)nchi) <i>may-disna=nchi</i>	X-en=((a)nchi) <i>inom-en=anchi</i>	X-an=((a)nchi) <i>tolas-an=anchi</i>	i-X=((a)nchi) <i>i-pangay=anchi</i>		
NOMINAL			paN-X <i>pam-aso</i>	pay-X <i>pay-disna</i>					

TABLE 10.5 Ibatan verbal morphology for non-native stems

Ibatan: Non-native stems		ACTOR			UNDERGOER		
	INCHOATIVE	DISTRIBUTIVE	DURATIVE	PATIENT	LOCATIVE	CIRCUMSTANCIAL	
REALIS	<i>gatang</i> 'to buy'	<i>kalod</i> 'to get coconuts'	<i>kalap</i> 'to fish'	<i>osar</i> 'to use'	<i>dalos</i> 'to clean'	<i>lako</i> 'to sell'	
NEUTRAL	<om>X <i>g<om>atang</i>	naN-X <i>mang-alod</i>	mag-X <i>mag-kalap</i>	X-en <i>osar-en</i>	X-an <i>dalos-an</i>	i-X <i>i-lako</i>	
PERFECTIVE	<om>(i)n>X <i>g<om>n>atang</i>	naN-X <i>nang-alod</i>	mag-X <i>mag-kalap</i>	<in>X <in> <i>osar</i>	<in>X-an <i>d<in>alos-an</i>	in-X <i>in-lako</i>	
IMPERFECTIVE	<om>CVC-X <i>g<om>at-gatang</i>	maN-CVC-X <i>mang-al-kalod</i>	mag-CVC-X <i>mag-kal-kalap</i>	CVC-X-en <i>os-osar-en</i>	CVC-X-an <i>dal-dalos-an</i>	i-CVC-X <i>i-lak-lako</i>	
IRREALIS	<om>X=((a)nchi) <i>g<om>atang=anchi</i>	maN-X=((a)nchi) <i>mang-alod=anchi</i>	mag-X=((a)nchi) <i>mag-kalap=anchi</i>	X-en=((a)nchi) <i>osar-en=anchi</i>	X-an=((a)nchi) <i>dalos-an=anchi</i>	i-X=((a)nchi) <i>i-lako=nchi</i>	
NOMINAL		paN-X <i>pang-alod</i>	pag-X <i>pag-kalap</i>				

guages have contributed (Gallego, 2022a), but also the central role of Ilokano in driving contact-induced structural change in Ibatan. Several lines of evidence point to Ilokano as the most likely source of the durative paradigm. First, while the forms of the non-native durative prefixes are actually shared among a number of Philippine languages, most notably Filipino, making any of these languages the possible source of the paradigm, this is highly unlikely because of the limited history of contact between the Ibatans and speakers of these languages. Second, the overall number of loanwords, including complex ones, across the different source languages, shows an overwhelming bias towards Ilokano as the SL. Finally, supported by known patterns of multilingualism, both past and ongoing, Ibatan speakers across generations generally use Ilokano as their second language, as compared to Filipino, which is only starting to be used as a second language among the younger generations of Ibatans.

In terms of form, while Ilokano reflects *ag-* for realis neutral whereas Ibatan reflects *mag-*, this can be analyzed as an outcome of analogy, where the adapted Ibatan form *mag-* has been analogized with the native counterpart *may-*, thus matching the rest of the prefixes, that is, the non-native paradigm *mag-*, *nag-*, *pag-*, with the native *may-*, *nay-*, *pay-* (see Section 4.1. for further explanation).

As for distribution, while the non-native paradigm is by and large restricted to non-native stems, this is not always the case. That is, there is also a small number of hybrid formations observable in the language, which are of two types: non-native prefixes occurring with native stems (Type 1), such as *bwang* ‘go bald’ in (4a), and native prefixes occurring with non-native stems (Type 2), such as *bilag*¹⁹ ‘dry in the sun’ in (4b).

- (4) a. Non-native *mag-* with native stem (Type 1 hybrid formation)

Magbwang si maraan. (elicited)

Mag-bwang si maraan

DUR-bald.IVB DET uncle

‘Uncle is going bald.’

19 Clearly a loanword as evidenced by the final consonant *g*, which is the reflex of *j in Ilokano and a number of Northern Luzon languages, as in PMP *bilaj ‘spread out in the sun to dry’ > Ilokano *bilag*, Isneg *bilag*, Bontok *bilag*, and Proto Austronesian (PAN) *apejux ‘gall, gallbladder, bile’ > Ibanag *aggu*, Ifugaw *apgo*, Pangasinan *apgo* (Blust and Trussel 2020). In the Batanic languages, the consonant is typically reflected as *d*, as in PAN *apejux > Itbayaten *apdo* (Blust and Trussel 2020) and Ibatan *apdo* (Maree et al. 2012).

3.1 *Unexpected Formations*

The first category among the small set of unexpected formations involves hybrid forms, or combinations of native and non-native material. Type 1 involves non-native morphology used with native stems (14 of 1436 stems, or merely 0.97%) and Type 2 involves native morphology used with non-native stems (62 of 1436 stems, or 4.32%). Table 10.7 gives some examples.

Evidently, the development of the non-native durative paradigm in Ibatan has arisen mainly through indirect transfer, that is, via the transfer of complex loanwords (Seifart 2015), as evidenced by the significant number of Ilokano stems that occur with the paradigm. The existence of hybrid formations, however, suggests other mechanisms that must have operated in driving this particular contact-induced change.

Type 1 hybrid formations constitute only a very small fraction of the overall distribution (only 0.97% of all instances of durative formations indicated in Maree et al. 2012). Such kinds of formations raise an important question about how non-native morphology comes to be extended to native stems. As for Type 2 hybrid formations, while they occur more frequently than Type 1 forms, they still constitute a very small portion of the overall distribution (4.32%). These two types of hybrid formations, along with other unexpected distribution, although very few in number, point to further complexity in Ibatan in terms of diversity of structures, as discussed below.

In deriving the basic durative meaning, loanwords occur with the non-native paradigm as expected, but in more complex formations that also involve other affixes, the native morphology is used. Table 10.8 gives some examples, where *bosel* '(develop) buds', *kamoras* '(become sick with) measles', *darop* 'attack', and *tiro* 'shoot' are all loanwords that are marked with the non-native *mag-* for the basic durative form but take the native paradigm *may-* when combined with other native affixes such as the distributive *cha-* and the reciprocal *sin-* along with reduplication to mark additional meanings of the verb.

Such cases suggest how morphology, even in agglutinative languages that have relatively transparent compositionality, such as Ibatan, encodes meanings on the basis of patterns of combination, irrespective of the discrete functions of the component elements (cf. Word and Paradigm approach by Hay and Baayen 2005, Ackerman et al. 2009, etc.). That is, more complex derivations in Ibatan appear as combinations involving native morphology, and these apply even for loanwords that are known to take non-native morphology in basic derivations. The sentences below illustrate this further. The Ilokano verb *labang* 'dappled' in (5a) and Spanish *tiro* 'shoot' in (6a) occur with *mag-/nag-* in the basic durative form, but (5b) shows *may-cha-laba-labang* 'have irregular patches' involving the native prefix *may-* in combination with *cha-*

TABLE 10.7 Hybrid formations in Ibatan

Category	Source	Prefix	Stem	Definition
Type 1	Native	<i>mag-</i>	<i>inen</i>	thrifty; something is gradually consumed, especially food; use sparingly
Type 1	Native	<i>mag-</i>	<i>ippet</i>	an intestinal roundworm
Type 1	Native	<i>mag-</i>	<i>payaw</i>	hoarse (voice)
Type 1	Native	<i>mag-</i>	<i>rongsoh</i>	hammer
Type 1	Native	<i>mag-</i>	<i>sangpah</i>	hold in mouth
Type 2	Ilokano	<i>may-</i>	<i>abagis</i>	a term expressing a close relationship between cousin and sibling
Type 2	Ilokano	<i>may-</i>	<i>bilag</i>	sun dry clothes, grains, etc.
Type 2	Ilokano	<i>may-</i>	<i>ikit</i>	aunt, aunty
Type 2	Spanish	<i>may-</i>	<i>dasal</i>	prayer, prayer time
Type 2	Spanish	<i>may-</i>	<i>tarabako</i>	labor, work

TABLE 10.8 Restricted distribution of the non-native durative paradigm vis-à-vis the native paradigm

Source	Stem	Prefix	Function	Derived form	Definition
ILO	<i>bosel</i>	<i>mag-</i> <i>may-cha-RDP-</i>	durative durative, distributive	<i>mag-bosel</i> <i>may-cha-bos-bosel</i>	develop buds of a fruit or vegetable develop buds together
ILO	<i>kamoras</i>	<i>mag-</i> <i>may-cha-RDP-</i>	durative durative, distributive	<i>mag-kamoras</i> <i>may-cha-kamo-kamoras</i>	become sick with measles have measles at the same time
ILO	<i>darop</i>	<i>mag-</i> <i>may-sin-</i>	durative durative, reciprocal	<i>mag-darop</i> <i>may-sin-darop</i>	attack two or more people or groups from different areas attack each other
SPA	<i>tiro</i>	<i>mag-</i> <i>may-sin-RDP-</i>	durative durative, reciprocal	<i>mag-tiro</i> <i>may-sin-ti-tiro</i>	hit, shoot, throw hit, shoot, throw something at each other

and CVCV reduplication to further derive the distributive meaning, and (6b) shows *may-sin-ti-tiro* 'throw at each other', again involving the native *may-* with the affix *sin-* and CV reduplication to derive the reciprocal meaning to the verb.

- (5) a. Non-native *mag-* with non-native stem (Complex loanword)
Maglabang kodit kwaya, ta nadoplagan. (Maree et al. 2012: labang)
Mag-labang kodit kw=aya ta nadoplagan
 DUR-dappled.ILO skin 1SG.GEN=REF because scalded.
 'My skin becomes dappled because it was scalded.'
- b. Native *may-cha-RDP-* with non-native stem (Type 2 hybrid formation)
Maychalabalabangayaw basket kwaya. (Maree et al. 2012: labang)
May-cha-laba-labang=aya=w basket kw=aya
 DUR-DIST-RDP~dappled.ILO=REF=NOM basket 1SG.GEN=REF
 'My basket has irregular patches of color.'
- (6) a. Non-native *mag-* with non-native stem (Complex loanword)
Nagtiro so amang so pirpiroka. (Maree et al. 2012: tiro)
Nag-tiro si amang so pirpiroka
 DUR-shoot.SPA DET father DET pirpiroka.bird
 'Father shot the pirpiroka bird.'
- b. Native *may-sin-RDP-* with non-native stem (Type 2 hybrid formation)
Maysintitiro saw mangalkem so bwa. (Maree et al. 2012: tiro)
May-sin-ti-tiro sa=aw mangalkem so bwa
 DUR-REC-.SPA RDP~throw 3PL.NOM=REF old.men DET betel.nut
 'The old men threw betel nuts at each other.'

To illustrate further, Table 10.9 presents various derivations involving the durative paradigms found in Maree et al. (2012). The diversity of structures that can co-occur with the non-native durative prefixes is evidently limited compared to those that combine with the native *may-*. Such restricted distribution of the non-native paradigm indicates that it is not yet fully parallel with its native counterpart, especially with structures involving more complex morphological combinations that encode further semantic specifications on the verb.

TABLE 10.9 Further morphological derivations involving the durative paradigms

Form	Function	Example	Meaning
Derivations involving the non-native durative paradigm			
<i>machi-pag-</i>	Associative	<i>machi-pag-ragsak</i>	someone rejoices with someone
<i>pag-X-en</i>	Causative	<i>pag-bolos-en</i>	allow water to flow freely
<i>ma-pag-</i>	Causative	<i>ma-pag-bwenas</i>	someone or something causes someone luck

TABLE 10.9 Further morphological derivations involving the durative paradigms (*cont.*)

Form	Function	Example	Meaning
<i>mag-pa-</i>	Causative	<i>mag-pa-borek</i>	someone boils something in a pot
<i>maka-pag-</i>	Conditional ability	<i>maka-pag-pikar</i>	someone is able to make an engine, machine, or motor go faster
<i>pag-X-an</i>	Locative	<i>pag-mangamanga-an</i>	someone doubts about someone or something
<i>ka-pag-</i>	Nominalization	<i>ka-pag-tanggap</i>	a woman's confinement and recuperation after giving birth
<i>ka-pag-RDP-</i>	Nominalization	<i>ka-pag-so~sopyat</i>	a controversy, dispute
<i>mag-ka-</i>	Similarity	<i>mag-ka-picha</i>	two events are on the same day
Derivations involving the native durative paradigm			
<i>machi-pay-RDP-</i>	Associative	<i>machi-pay-po~pohaw</i>	someone stays awake the whole night with someone
<i>pay-X-en</i>	Causative	<i>pay-amonyit-en</i>	someone closes up a cut or a wound
<i>ma-pay-</i>	Causative	<i>ma-pay-chidong</i>	make something corrugated
<i>may-pa-</i>	Causative	<i>may-pa-diman</i>	someone is about to die
<i>maka-pay-</i>	Conditional ability	<i>maka-pay-bangon</i>	someone is able to wake up
<i>may-cha-</i>	Distributive	<i>may-cha-liproso*</i>	someone has leprosy
<i>may-cha-RDP-</i>	Distributive	<i>may-cha-bos~bosel</i>	a plant develops buds
<i>pay-cha-X-en</i>	Distributive	<i>pay-cha-pidy-en</i>	someone chooses and separates something
<i>may-cha-RDP-X-an</i>	Durative	<i>may-cha-ra-rak-an*</i>	someone or an animal does something the whole night
<i>may-cha-X-an</i>	Durative	<i>may-cha-sary-an</i>	someone or an animal does something from dawn to dusk
<i>pay-RDP-</i>	Intensive	<i>pay-sawa~sawat</i>	someone chatters about something
<i>ka-pay-cha-X-en</i>	Intensive, superlative	<i>ka-pay-cha-rakmah-en</i>	the worst of an injury or sickness
<i>pay-X-an</i>	Locative	<i>pay-ketket-an</i>	make a nest someplace
<i>pay-pay-pa-X-an</i>	Locative	<i>pay-pay-pa-ktas-an</i>	the place where someone roams around
<i>ka-pay-</i>	Nominalization	<i>ka-pay-alit</i>	equality
<i>ka-pay-RDP-</i>	Nominalization	<i>ka-pay-si~sidong</i>	cooperation
<i>ka-pay-sin-RDP-</i>	Pretense	<i>ka-pay-sin-si~singpet</i>	hypocrisy
<i>may-RDP-</i>	Process	<i>may-a~alat</i>	someone weaves an alat basket
<i>may-sin-</i>	Reciprocal	<i>may-sin-darop*</i>	two or more people or groups from different areas attack each other
<i>may-sin-RDP-</i>	Reciprocal	<i>may-sin-ti~tiro*</i>	two people hit, shoot, throw something at each other
<i>may-pay-</i>	Reciprocal	<i>may-pay-palang</i>	two or more people pull something back and forth from opposite ends
<i>may-pi-</i>	Repetition	<i>may-pi-rwa</i>	someone does or something happens twice
<i>may-CVy-</i>	Repetition	<i>may-roy~rongsoh</i>	to keep hammering

*stem is a loanword, constituting hybrid formation

There are also a few cases where both native and non-native durative prefixes can be used with the same verb, but appear to encode divergent meanings. An example is the Spanish word *kwarto* ‘room’, where *mag-kwarto* in (7a) means ‘make a room’, encoding dynamicity, while *nay-kwarto* in (7b) means ‘have a room’, encoding a stative sense.

- (7) a. Non-native *mag-* with non-native stem (Complex loanword)
Magkwarto ka so rakoh. (Maree et al. 2012: *kwarto*)
Mag-kwarto ka so rakoh
 DUR-room.SPA 2SG.NOM DET big
 ‘Make a big room.’
- b. Native *nay-* with non-native stem (Type 2 hybrid formation)
Naykwarto so anem bahay ko, ki dedekey. (Maree et al. 2012: *kwarto*)
Nay-kwarto so anem bahay ko ki de~dekey
 DUR-room.SPA DET six house 1SG.GEN but RDP~small
 ‘My house has six rooms, but they are small.’

Another example is in expressing direction/goal. The sentences in (8a) and (8b) involve the native Batanic word *songet* ‘forested area’. *Songet* also happens to be a place name in Babuyan Claro, and when derived to mean ‘to go to Songet’, it takes the non-native prefix *mag-* in combination with the directional *pa-*, as shown in (8a). In contrast, when referring to its general sense as ‘forested area’, the stem takes the native prefix *may-pa-*, as shown in (8b).²⁰

- (8) a. Non-native *mag-pa-* with a proper noun (Type 1 hybrid formation?)
Magpa-Songet dana sa. (elicited)
Mag-pa-Songet dana sa
 DUR-DIR-Songet.IVB already 3PL.NOM
 ‘They are already going to Songet.’

20 The same structure to mark direction/goal exists in Ivatan. However, there is no morphological distinction between general or specific locations as in Ibatan. Thus, in Ivatan, the form *may-pa-songet* can either be interpreted as ‘go to Sunget (a place in Mahatao, Batanes)’ or ‘go to the forested area’. However, the latter is the more common interpretation, as using the construction *may-pa-* to refer to proper nouns is not commonly used in Ivatan (based on personal communication with an Ivatan speaker).

- b. Native *may-pa-* with native stem (Native formation)
Maypasonget si anang mabekas. (elicited)
May-pa-songet *si anang mabekas*
 DUR-DIR-forested.area.IVB DET mother morning
 ‘Mother is going to the forested area in the morning.’

Ibatan also has instances of doublets, where a particular form is actually descended from two different sources. An example is the verb *boya* ‘to see, to meet, to watch’, where the Batanic languages and Ilokano share cognate forms. Ivatan *vuya*, Itbayaten *vooya*, and Ibatan *boya*²¹ are all cognates carrying the meaning ‘to see, to meet’. The Ibatan stem takes *may-*, as illustrated in (9a). The semantics of the word has also been expanded to include the meaning ‘to watch’, but in this particular sense, the form takes the non-native prefix *mag-*, as seen in (9b). This particular meaning of the form has been transferred from Ilokano, where the Ilokano word *buya*²² means ‘to watch’.²³ It is only the difference in meaning and the use of the non-native prefix that indicates that *mag-boya* is a complex loanword instead of a Type 1 hybrid formation.

- (9) a. Native *may-* with native stem (Native formation)
Mayboya tanchi andelak. (elicited)
May-boya *ta=anchi andelak*
 DUR-meet.IVB 1PL=FUT tomorrow
 ‘Let’s meet tomorrow.’
- b. Non-native *mag-* with non-native stem (Complex loanword)
Magboya kami so sine do Sabado. (elicited)
Mag-boya *kami so sine do Sabado*
 DUR-watch.ILO 1PL DET movie DET Saturday
 ‘We will watch a movie on Saturday.’

21 Ibatan reflects all instances of *v* in the other Batanic languages as *b*, thus the form *boya*. This is assumed to be a later change in Ibatan, arising from contact with Ilokano which retains the original PMP *b.

22 Ilokano *buya* and Ibatan *boya* are pronounced similarly, with both <u> and <o> pronounced as a high, back vowel. The only difference is orthography, where the vowel in Ibatan is represented as <o>.

23 In Ivatan, the verb ‘watch’ is *talamad*, as in *May-talamad aku su sine andelak* ‘I will watch a movie tomorrow’ (compare Ibatan *mag-boya* in (9b)). In Ibatan, however, *talamad* means ‘look down’. It is clear that the transfer of Ilokano *buya* ‘watch’ has affected this particular semantic network, where Ibatan *boya* has been extended to include the Ilokano meaning ‘watch’, and *talamad* has shifted to exclusively mean ‘look down’.

TABLE 10.10 Pairs of near-homophonous native and non-native forms in Ibatan

Source	Prefix	Stem	Definition
Native	<i>may-</i>	<i>babáng</i>	carry on the back
Ilokano	<i>mag-</i>	<i>bábang</i>	hesitate
Native	<i>may-</i>	<i>barót</i>	develop a boil
Ilokano	<i>mag-</i>	<i>bárot</i>	thread rattan strips
Native	<i>may-</i>	<i>sagót</i>	wear a loincloth
Ilokano	<i>mag-</i>	<i>ságot</i>	give a gift
Native	<i>may-</i>	<i>talón</i>	mound up, swell
Ilokano	<i>mag-</i>	<i>tálon</i>	make a rice paddy

This also relates to near-homophonous pairs of words that have arisen out of contact, where native Ibatan terms have come to share near-similar forms with Ilokano loanwords (only differing in terms of stress placement). Despite the similarity, however, the meanings and etymologies are kept distinct not only by maintaining the difference in the placement of stress, but also by the use of native and non-native prefixes, as illustrated in Table 10.10. The forms *babang*, *barot*, *sagot*, and *talon* occur with both native and non-native morphology, keeping the meanings and etymologies separate.

The cases described above clearly illustrate how the distribution of the durative paradigms in Ibatan, while relatively straightforward in the majority of cases (including doublets and near-homophonous terms that have different etymologies), can still be unpredictable for a small set of stems that constitute hybrid formations. As a final point, there are also instances where both the native and non-native durative prefixes appear to be used interchangeably (Table 10.11). It is not certain whether these are instances of stable variation in Ibatan, or if these constitute change in progress, where particular groups of speakers may tend to prefer the use of one particular paradigm over the other.

Thus, while the non-native durative paradigm has not yet been fully integrated into the morphological system of Ibatan given its limited distribution, not just in terms of the stems it occurs with but also the kinds of other structures it can combine with, it has added to the morphological complexity of Ibatan through contact-induced change. That is, Ibatan exhibits diversity of structures that are not seen in either Ilokano or its sister Batanic languages (see Section 2). This clearly runs in contrast with the usual claim in the literature that language contact results in a reduction of morphological complexity, and/or

TABLE 10.11 Forms that involve native and non-native prefixes in free variation

Source	Prefix	Stem	Definition
SPA	<i>mag-, may-</i>	<i>apilyido</i>	have the surname of
SPA	<i>mag-, may-</i>	<i>aritos</i>	wear earrings
UNCERTAIN	<i>mag-, may-</i>	<i>gipit</i>	wear a hairclip
ILO	<i>mag-, may-</i>	<i>gisgis</i>	brush teeth
ILO	<i>mag-, may-</i>	<i>ibbong</i>	become smelly
ILO	<i>mag-, may-</i>	<i>lobnak</i>	wallow
ILO	<i>mag-, may-</i>	<i>pakopak</i>	clap bamboo cymbal

convergence between the languages in contact (cf. Gumperz and Wilson 1971, Matras and Sakel 2007, Gardani et al. 2015, etc.), which is often explained as a “by-product of the trend to syncretise the inventory of constructions across the languages in a bilingual’s repertoire” (Matras 2015:54). The case of Ibatan demonstrates that equal emphasis should be put on the nature and kinds of complexity that may arise in contact-induced change (cf. Bakker et al. 2011, Meakins et al. 2019, etc.).

3.2 *Ongoing Cross-Linguistic Influence*

So far, we have seen the general distribution and usage of the parallel durative paradigms in Ibatan, informed by data from Maree et al. (2012). These patterns constitute apparent contact-induced change that has become more or less stable in Ibatan. Synchronically, however, further variation in the usage of the paradigms can be observed among individual speakers.

Van Coetsem (1988, 2000) argues that individual speaker-based psycholinguistic mechanisms are linked to particular contact outcomes. His framework centers on the psycholinguistic notion of language dominance, which underpins the individual’s agentivity in bi-/multilingual speech. Language dominance has to do with the person’s relative proficiency in the different languages in their repertoire, where the dominant language is oftentimes the language they are most proficient in, typically their first language. However, it must be noted that dominance is not static and can vary across a person’s lifetime. Therefore, a person’s dominance may shift to their second language, and this is dependent on factors beyond language proficiency, such as exposure, frequency of use, and domain/context of use, among many others (cf. Silva-Corvalán and Treffers-Daller 2016, Treffers-Daller 2019, etc.). Therefore, contact effects vary as a person becomes more dominant in the recipient language (RL).

Particular patterns of language dominance determine the application of what van Coetsem (1988, 2000) describes as borrowing transfer in RL agentivity and imposition transfer in SL agentivity. An individual tends to work within the resources of their dominant language. Thus, when dominant in the RL, they use RL resources but may borrow components, typically vocabulary, from their non-dominant SL (RL agentivity). In contrast, a person who is dominant in the SL has a tendency to impose SL materials, such as phonology and grammar, when they use their non-dominant RL (SL agentivity). In terms of contact-induced outcomes, therefore, borrowing transfer results in largely lexical borrowings, which are sporadic, while imposition transfer tends to result to a “catastrophic modification” of aspects of the RL by means of systematic structural innovations (van Coetsem 1988:25).

Taking this framework to understand synchronic cross-linguistic influence among Ibatan speakers, the variant use of the durative paradigms appears to correlate with language dominance. As presented in Section 1, there are three general groups of Ibatan speakers, namely Ibatan-dominant early bilinguals, Ilokano-dominant early bilinguals, and Ilokano-dominant late bilinguals, and they exhibit variation in their knowledge and use of the durative paradigms, based on a preliminary corpus of Ibatan speech collected during the author’s fieldwork, and supplemented by interviews with Ibatan speakers.

Ibatan-dominant early bilinguals exhibit the general pattern of the durative paradigms described in the previous section. A number of these speakers, in fact, show good awareness of internal structures and etymology, where they identify stems that occur with *mag-* as non-native, typically from Ilokano, and those that occur with *may-* as native stems, which they describe as “pure Ibatan.” This indicates that they have good knowledge of both Ibatan and Ilokano, and they clearly maintain the distinction between the two languages by means of the associated morphological structures.

In a similar vein, Ilokano-dominant early bilinguals (or those who have learned Ibatan and Ilokano in their childhood but prefer Ilokano as their everyday language) also appear to maintain the boundaries of the two languages. In Babuyan Claro, these speakers are known for code-switching between the two languages (where Ilokano is the matrix language), described by locals as *Ibakano*, a blend of Ibatan and Ilokano. Despite their relative dominance in Ilokano, however, they still follow the expected use of the durative paradigms, even in situations where they switch between Ibatan and Ilokano in an utterance, as illustrated in (10). Here, the Ibatan verb *may-tay-tagadan* ‘(remain) slack’, reflecting the expected use of the durative prefix, is maintained alongside a by and large Ilokano utterance.

- (10) Ilokano–Ibatan code-switching (Gallego ongoing: IVB1–20180830_04)
- a. ILO *Inserrek da man diay kwarto nga napan da nangcheck-upan kanianan ngem*
 ‘They put (him) in the room where he was checked up but ...’
- b. IVB *naw na nga may-tay~tagadan.*
 ‘(his mouth) just remained slack.’

In contrast, Ilokano-dominant late bilinguals, who have learned Ibatan in adulthood when they migrated to Babuyan Claro, tend to show structural imposition in their use of Ibatan. In terms of morphology, these speakers exhibit increased usage of the non-native durative paradigm, even with native stems that are expected to occur with the native paradigm. This is illustrated in sentences (11) and (12).

In (11), the Ilokano-dominant late bilingual speaker used the non-native prefix *pag-* for the native stem *chichwas* ‘search’ instead of the expected native prefix *pay-*. In other instances, the same speaker used the expected *may-* for native stems, as seen in (12). The variant use of the durative paradigms by Ilokano-dominant late bilinguals, illustrated in (11), are regarded by Ibatan-dominant speakers as errors, and have come to be a marker that sets apart this group of speakers. It is however important to highlight the temporary nature of these impositions. That is, as proficiency or dominance in the RL increases, these impositions tend to lessen in the speech of Ilokano-dominant late bilinguals.²⁴

- (11) Non-native *pag-* with native stem (Type 1 hybrid formation)
 Gallego (2019): IVB1–20180930_08
Pati iyaw no chitowa aywanaw ki nachipagchichwas.
Pati iyaw no chito=a aywan=aw ki nachi-pag-chichwas
 also DEI DET dog=LK pet=REF INV SOC-DUR-search.IVB
 ‘Even the pet dog searched (with him).’
- (12) Native *may-RDP-* with native stem *yonot* (Native formation)
Myan saw mayyoyonot kan yaw no chitwaw.
Myan sa=aw may-yo~yonot kan yaw no chito=aw
 EXT 3PL=REF DUR-RDP~go.along.IVB and DEI DET dog=REF
 ‘There they are, going along, including the dog.’

24 However, the small number of Type 1 hybrid formations indicate that some of these cases of imposition transfer have become regularized in Ibatan, but this is assumed to constitute a deeper layer of change that is distinct from this ongoing imposition transfer in Ilokano-dominant speech.

As van Coetsem (2000) argues, language dominance and speaker agentivity do play important roles in explaining individual patterns of cross-linguistic influence and outcomes of contact-induced change. However, this model needs to be further tested and refined. In particular, the notion of language dominance needs to be operationalized more carefully. As seen in this section, language dominance is gradient, and contact outcomes may vary even among non-dominant speakers. That is, certain Ilokano-dominant speakers of Ibatan (i.e. late bilinguals) tend to exhibit structural imposition as predicted by van Coetsem's SL agentivity, whereas others do not (i.e. Ilokano-dominant early bilinguals). Measuring language dominance in a way that captures such differences would allow us to better understand contact outcomes.²⁵

4 Explaining the Structural Consequences of Lexical Transfer in Ibatan

There are certain types of change such as contact-induced structural change that were once considered very rare phenomena in language contact (cf. Haugen 1950, Weinreich 1953, Matras and Sakel 2007, Gardani 2008, Gardani et al. 2015, etc.). However, there is now a growing body of literature that explores not just the evidence of such contact-induced outcomes, but also the tendencies and constraints that drive structural change.

Language-internal constraints pertain to the nature of the linguistic materials as well as the nature of the languages in contact. The latter involves structural compatibility or typological fit, where bound morphemes are more easily transferred from SL to RL if the two languages share parallel structures. In the case of Ibatan and Ilokano, the two languages are genetically related, and so they share not only parallel morphological structures but also similar forms for some of the verbal affixes. This must have played a significant role in facilitating the development of non-native morphology into Ibatan.

As for the nature of the linguistic material itself, it is argued that linguistic materials have varying degrees of structuredness or integration within the grammar, and this has an effect on ease of transfer. Morphemes which are more functionally opaque and abstract, hence more tightly integrated within the linguistic system, tend to be more resistant to transfer than those that have more

25 A quantitative analysis of the correlations between structural imposition in multilingual speech and language dominance provides empirical support to these claims. A corpus of Ibatan speech is currently being collected for the next phase of the author's research project (cf. Gallego ongoing).

concrete and transparent functions (Gardani et al. 2015:6). This idea is central in explaining the hierarchies which have been proposed in the early contact linguistic literature, where materials with more concrete functions and meanings, such as nouns and verbs, are argued to be more easily transferred than function words, and similarly, within the domain of morphology, derivational material over inflectional forms.²⁶ In Ibatan, it is clear that derivational morphology has been shaped by language contact, as seen in the development of the parallel durative paradigms, but inflectional paradigms reflect contact-induced features to a certain degree as well, as illustrated briefly in the domain of aspectual inflection in Section 2.2. This is indicative of the extent of contact-induced change in Ibatan, where it can be observed across all domains of the language, including ones which are said to be most resistant to transfer.

Moving beyond language-internal constraints that have been the main focus in the early language contact literature, more recent studies set up models that involve context-dependent and language-external explanations to account for the transfer of various linguistic materials. Focusing on morphology, Seifart (2015) represents morphological transfer as a cline, where on one end, non-native structure is restricted to non-native stems (constituting indirect transfer via complex loanwords), and where the other theoretical extreme are cases of hybrid formations (constituting direct transfer). Most cases of language contact would fall somewhere in between these two ends, where contact-induced structural change involves both direct and indirect processes, and the differences in each situation would be the ways in which these processes took place in the RL. To illustrate, the distribution of the non-native durative paradigm in Ibatan in complex loanwords and hybrid formations is indicative of the mechanisms that led to the development of such non-native structure in the language. These mechanisms often involve factors beyond linguistic structure. Seifart (2015) argues that direct transfer relies on the speakers' knowledge of the SL, whereas indirect transfer is governed by more complex processes, determined by schemas and local generalizations that revolve around the frequency of complex loanwords that carry the affix in question vis-à-vis corresponding simplex words.²⁷

26 However, it must be noted that the division between inflection and derivation is not always clear-cut. Some in fact argue that rather than constituting discrete categories, they instead form a continuum (see Bybee 1985, Dressler 1989, Haspelmath 1996, and Laca 2008). This gradience therefore adds further complexities in accounting for such hierarchies.

27 This derives from the concept of gradient morphology and the Word and Paradigm approach (see for instance Bybee 1995, Hay and Baayen 2005, Baayen 2008, and Ackerman, et al. 2009).

The contexts that underpin the contact situation, particularly the nature and intensity of social contact between the groups, determine the extent in which the SL affects RL structure (cf. Thomason and Kaufman 1988). For morphological transfer, this may sometimes result in what Kossmann (2010) describes as Parallel System Borrowing, which involves co-existent native and non-native forms in a language. In many cases, non-native morphology is restricted to loanwords, and are often unstable and irregular, but in other cases, these structures can achieve stability and even morphological productivity, and can become extended to native stems. Another related phenomenon is the transfer of sets of paradigmatically and syntagmatically related affixes. Seifart (2012, 2017) argues that this is in fact more frequent than the transfer of isolated forms, and this is known as the Principle of Morphosyntactic Subsystem Integrity. The morphological system of Ibatan evidently shows Parallel System Borrowing, where the non-native paradigm exists along with its native counterpart. Additionally, this morphological change in the language involves sets of related forms, as Seifart (2012, 2017) argues. These pieces of evidence point to the intensity of contact between Ibatan and Ilokano. However, as the two languages are genetically related and thus share a number of identical voice and aspectual affixes, it is extremely difficult to ascertain the full extent of this paradigmatic transfer of verbal morphology in Ibatan.

Curnow (2001) argues for the need to consider extra-linguistic information that goes beyond structural constraints in investigating the pathways of development of contact-induced change. Muysken (2010) takes a similar position, and proposes a scenario approach to language contact. Understanding contact phenomena from the aggregates of the multilingual individual, the community, and the larger geographical regions of the world provides stronger links between linguistic outcomes and the socio-historical contexts that underpin them. Essentially, Muysken (2010: 278) argues for an approach where “a specific linguistic result is linked to a historical setting, involving specific people (age, ethnicity, mix) with specific languages, languages interacting following specific scenarios, which are governed by well-defined processing constraints.”

In sum, the various constraints and mechanisms that govern language contact involve not only language-internal factors, but also language-external, context-based explanations. Thus, in seeking explanations for contact-induced outcomes, it is therefore necessary to take into account the contexts that underpin the particular contact-induced change under investigation. The dynamic setting of the Babuyan Claro community entails various mechanisms that drive contact outcomes, and these are reflected as layers of contact-induced change in Ibatan. In particular, the development of non-native morphology in the lan-

guage is facilitated not only through typological fit and structural compatibility, but the dynamic nature of multilingualism both at the levels of the individual and the community is also argued to be central in driving this type of change.

4.1 *Layers of Contact-Induced Change in Ibatan*

A context-based framework in analyzing contact-induced outcomes is proposed by Thomason and Kaufman (1988), which centers on the sociolinguistic context of the multilingual community (that is, the intensity and type of contact situation, which result in either language maintenance or shift). In situations of language maintenance, involving “borrowing interference”, the cline goes from light, moderate, to heavy contact, and in situations of language shift, involving “substratum interference” or “interference through shift”, the cline relates to the degree of interference from the source language, which depends on the size of the shifting group and the level of bilingualism of the community. Where the specific contact situation of the community is placed along the cline would determine the particular contact-induced outcomes, namely the transfer of non-basic vocabulary, or the transfer of more structured materials such as phonological, morphological, syntactic, and lexico-semantic features.

Thomason and Kaufman’s (1988) scale that focuses on widespread, community-level contact outcomes relates to the central concepts in van Coetsem’s (1988, 2000) speaker-based framework. That is, in situations of language maintenance, change is primarily seen in the lexicon, and this relates to the mechanisms involved in RL agentivity. In situations of shift, restructuring in the RL can happen via imposition of phonological and grammatical features from the SL, which is akin to the mechanisms governing SL agentivity.

Accounting for contact-induced language change involves linking the individual and the community, and understanding the transition from innovations to widespread change. Van Coetsem (2000) and Thomason and Kaufman (1988)’s models for language contact both put the psycholinguistic and sociolinguistic contexts of the multilingual individual and community at the heart of their frameworks. It then follows that communities with an extremely dynamic socio-political and linguistic landscape such as Babuyan Claro would reflect layers of change that are linked to changes in the patterns of multilingualism of the individual and the community. These phases in the history of Babuyan Claro are summarized in (1) and repeated in (13) below.

- | | | |
|------------|---------|---|
| (13) 1870s | Phase 1 | The arrival of the first Ibatan people |
| 1900s | Phase 2 | The emergence of the daya~laod networks |
| 1970s | Phase 3 | The rise of Ilokano |

1980s	Phase 4	The renewed vitality of Ibatan
ongoing	Phase 5	The influx of Ilokano immigrants
ongoing	Phase 6	The increasing influence of Filipino

The first Ibatans [RL agentivity]. The first group who permanently settled Babuyan Claro in 1869 originally came from Batanes but had been relocated to the Ilokano-speaking islands of Calayan and Camiguin (Maree 2005). It can be assumed that while they were there, they had considerable interaction with Ilokano speakers, but to what degree they learned Ilokano is uncertain. At this stage, it can be argued that loanwords, including complex ones, were introduced into Ibatan, but were fully adapted not just in terms of morphological structure (as Type 2 hybrid formations) but also in terms of phonology²⁸ given the likely individual-level dominance of Ibatan across these Ibatan-speaking first families.

The daya and laod networks [RL agentivity]. As more groups from both Batanic- and Ilokano-speaking backgrounds came to Babuyan Claro, the population on the island slowly grew. In the initial years of the community, ethnographic evidence shows that ethnolinguistic lines were kept more or less separate (Maree 1982), and this can be seen in the emergence of distinct social networks clustered in the geographic regions of *daya* ‘east’ and *laod* ‘west’ coinciding with the use of Ibatan and Ilokano respectively. However, the harsh environmental conditions on Babuyan Claro meant that the inhabitants relied on social contact across these networks. Interaction with Ilokano-speaking networks (*laod*) most likely facilitated the continued transfer of loanwords into Ibatan, which were then fully adapted into the language, under the assumption that the Ibatan-speaking networks maintained their dominance in Ibatan

28 To illustrate, the Ibatan word *absog* ‘bloated’, from Ilokano *bussog* ‘satiated, inflated’ (reconstructed as PAn *besuR ‘satisfied from having eaten enough, satiated’ (Blust and Trussel 2020), and forms doublets with the native Batanic *absog* ‘satiated’) underwent a unique Batanic sound change involving forms carrying the reflex of PAn *e (see Blust 2017 for further discussion). It is worth noting that this sound change is not productive in Ibatan anymore, and gives further support to the antiquity of these loanwords. A different explanation for this initial *a* in the Batanic languages is put forward by Reid (personal communication), where *a-* is analyzed as a retention of the old stative prefix *?a- (replaced by the newer forms *ma-* or *na-*), with subsequent loss of the original unstressed *e* in the Batanic languages. In either explanation, this initial *a-*, be it a result of sound change or a retention of the stative prefix, also applied in early loanwords in Ibatan, as seen in *absog* ‘bloated’.

since the arrival of their ancestors on the island. These fully adapted loanwords, which are older, widespread, and more socially integrated (cf. Poplack, et. al. 1988:72), are hence indicative of community-level dominance in Ibatan at this stage.

Ibatans with increased proficiency in Ilokano [SL agentivity]. The *daya* and *laod* networks largely correlate with speakers' language ideologies and use. While the setting in the early years of Babuyan Claro fostered a type of egalitarian multilingualism, where both Ibatan and Ilokano co-existed on a more or less equal footing, the rise in the status of Ilokano in the wider region, and consequently in Babuyan Claro, had profound effects on the patterns of multilingualism on the island around the 1970s. In addition to a significant portion of the population tracing their ancestry to Ilokano (and so maintaining Ilokano as their first language), there were more domains in which Ilokano was used to the exclusion of Ibatan, consequently threatening vitality. As a result, a number of Ibatan families have shifted to Ilokano as their everyday language. The Babuyan Claro community, including Ibatan-dominant speakers, certainly had increased exposure to and proficiency in Ilokano during this period. This either meant a shift in language dominance for some speakers, thereby becoming Ilokano-dominant, or a shift to (near-)symmetrical/balanced bilingualism for others, wherein they have (near-)equal dominance in both languages.

We can assume that this change in the nature of bilingualism drove a different kind of lexical transfer from that of the early stages of the community. That is, loanwords kept their SL morphology instead of being fully adapted into the grammar of Ibatan, driven by the increased proficiency of the speakers in Ilokano. At this stage, increased dominance in Ilokano may have entailed SL agentivity, and the maintained use of Ilokano morphology in Ibatan is indicative of imposition transfer. Moreover, the speakers' comparable proficiencies in the two languages, including a degree of awareness of morphological structures, must have facilitated the development of the adapted form *mag-* from the original Ilokano form *ag-*. That is, the speakers have analogized the Ilokano form *ag-* on the basis of the native counterpart *may-*. Since Ilokano *ag-* forms a paradigmatic relationship with the prefixes *nag-* and *pag-*, it is not difficult to analogize the form to be parallel with the native paradigm *may-*, *may-*, and *pay-*, thus leading to the current form *mag-*.

Younger generations of Ibatan-dominant speakers [RL agentivity?]. Further socio-historical changes in the Babuyan Claro community led to the renewed vitality of Ibatan from the 1980s. Ibatan has now regained its function as the

main language in Babuyan Claro, with Ilokano as the second language of the community and the lingua franca of the wider region. Younger generations of Ibatan speakers maintain their dominance in Ibatan, but keep considerable interaction with Ilokano speakers. This maintained social contact across the networks therefore allows for the mechanisms and processes that drive contact-induced change in Ibatan to persist.

Ilokano-dominant late bilingual speakers [SL agentivity]. Given the function of Ibatan as the main language of Babuyan Claro, Ilokano immigrants are learning Ibatan as their second language. As discussed in Section 3.2, the ongoing imposition transfer in the speech of Ilokano-dominant late bilingual speakers constitute the synchronic layer of contact-induced features we see in Ibatan. These features reflect a great deal of variation not only across individuals, but also within individual speakers. Synchronically, since such imposition transfer correlates with the speaker's (changing) language dominance, such can be transient and tend to be lost as the speaker's proficiency in Ibatan increases. These Ilokano immigrants constitute a small portion of the population, and their use of Ibatan tends to be dependent on the social networks they form in the community. That is, Ilokanos who form close ties with the *daya* network of mostly Ibatan-dominant speakers tend to learn Ibatan quickly, whereas those who are more affiliated with the *laod* network of Ilokano-dominant speakers tend to have lesser proficiency in Ibatan.

Ibatans with increased proficiency in Filipino [RL agentivity]. At present, the patterns of multilingualism in the Babuyan Claro community are shifting again, this time driven by the rising influence of Filipino. This is clearly reflected in how the younger generations of Ibatan speakers have become more proficient in Filipino. As Babuyan Claro became further integrated into the larger nation state, the Ibatans have more exposure to Filipino, not only as medium of instruction in schools, but also as the main language of print, broadcast, and social media. To compare, the older generations of Ibatans still have limited proficiency in Filipino, but a number of younger Ibatan-dominant speakers report preference towards using Filipino as their second language over Ilokano. As it happens, Filipino has forms identical to the non-native durative prefixes, and this must be reinforcing the current use and distribution of the paradigm in Ibatan. As expected, complex loanwords from Filipino, including nonce borrowings, occur with the non-native durative prefixes. Loanwords of foreign origin (typically English) are also introduced into Ibatan indirectly through Filipino, which have already been adapted with Filipino verbal morphology.

Partly through loanwords (and nonce borrowings), and via speakers with increased proficiency in Filipino, the non-native paradigm has come to be extended to include loanwords from other SLs. While it can be analyzed as a repurposing of the paradigm to accommodate non-Ilokano loanwords, the more accurate way to describe the influence of Filipino in this respect is reinforcing the function of the paradigm, given that Filipino shares exactly the same set of durative prefixes.

Linking phases and mechanisms. Changes in the socio-historical landscape of Babuyan Claro are clearly linked to changes in the nature of multilingualism on the island, which are then reflected as layers of contact-induced features in Ibatan. However, these apparent stages in the history of the community are by no means discrete. Even at present, different mechanisms of agentivity apply among different groups of speakers, yielding different outcomes: (1) for Ibatan-dominant speakers, RL agentivity resulting in lexical transfer, but keeping the boundary between Ibatan and Ilokano distinct through the expected use of the parallel paradigms; (2) for Ilokano-dominant early bilinguals, code-switching behavior with Ilokano as the matrix language; and (3) for Ilokano-dominant late bilinguals, SL agentivity resulting in the imposition of Ilokano structures in Ibatan speech, reflected in the variant use of structures.

This dynamic nature of multilingualism can also be seen in items that have been transferred multiple times into Ibatan. One clear example is the complex loanword *may-tarabako*, from Spanish *trabajo* ‘work.’²⁹ The degree of adaptation that applied on the loanword indicates that this is an early loan in the language. More recently, Maree et al. (2012) note that the younger generation now prefers to use the form *mag-trabaho*. This form, aside from the use of the non-native prefix *mag-*, exhibits a closer phonetic shape to the original Spanish word.³⁰ Such differences in how the word has been adapted into Ibatan shows agentivity at play; speakers with greater dominance in Ibatan are

29 Possibly transferred indirectly through Ilokano, as the two languages share the same adapted form *tarabako*.

30 This is also observable in Ilokano loanwords described in Footnote 28, where the more recent forms retain their original SL shape. To illustrate, Ibatan reflects doublet forms for ‘epileptic seizure’, *aksiw* and *kissiw*, both transferred from Ilokano *kissiw*, where the form *kissiw* is taken to be a recent loanword (not in Maree, et al. 2012, but evidently used by the speakers, particularly the younger generation, based on the author’s fieldwork), while *aksiw* is evidently an earlier loan reflecting greater phonological adaptation into Ibatan (with some speakers not aware of this older form).

more likely to adapt a form to their dominant Ibatan phonological structure, while those with greater proficiency in the sL³¹ tend to show less modification.

One thing that is apparent in the history of the Babuyan Claro community is that the speakers have continually kept Ilokano and Ibatan distinct. This etymological consciousness shows that the speakers are more or less aware of the differences between the languages in their repertoire, reflected most strikingly in how parallel morphological structures are used and maintained in Ibatan (not just in terms of derivational morphology discussed in this paper, but also in the domain of inflection, such as the aspectual marking described in Section 2.2). It also indicates how this must have been a conscious process for the Ibatans, as a way of flagging their mixed identity (Gallego 2020:107). This essentially relates to the phenomenon of *morphological compartmentalization* described by Matras (2015:48) for cases where (inflectional) morphology “is replicated along with lexical word forms from another language in situations in which speakers embrace and flag a bilingual identity.”

Ultimately, knowledge of sL structures is an essential part of how morphology is transferred and regularized in Ibatan. The large number of complex loanwords in the language suggests that the durative paradigm has been transferred indirectly. Seifart (2015) proposes that indirect transfer requires particular patterns in corpus frequencies involving pairs of complex and simplex loanwords, under the assumption that the speakers are analyzing non-native morphological structures on the basis of such patterns, but this does not seem to be the central mechanism for Ibatan. Given what we know of the nature of multilingualism in Babuyan Claro, the speakers are already clearly knowledgeable in Ilokano, and so, this must have played a crucial role in the development of non-native morphology in Ibatan. That is, good knowledge of Ilokano, along with the fact that the two languages are genetically related and typologically similar, allows for easier morphological analysis on the part of the speaker, which can then promote morphological productivity for non-native structures. Furthermore, this process entails a certain level of consciousness in the part of the speakers (cf. Thomason 2008, 2015), and that maintaining the distinction between Ibatan and Ilokano was an important motivation in this process.

At the same time, however, there are a few cases where the boundary between the two languages seems to be less clear. Hybrid formations are a clear indication of this. Some of these forms can be considered early loanwords

31 The sL is unlikely to be Spanish. Much of the Spanish lexicon in Ibatan is likely to have been transferred indirectly through Ilokano (and more recently Filipino).

into Ibatan (Type 2), and are indicative of speakers' shifting knowledge of what counts as loanwords, while others reflect impositions of SL structures (Type 1). While these forms comprise only a small subset of the distribution (5.29%), it is necessary to understand in more detail how such formations came to be stable in Ibatan, but this remains an open question.³²

4.2 *Further Questions*

It cannot be denied that outcomes of language contact and change exist within the socio-historical context of the community that use the languages. With context-based frameworks for language contact such as van Coetsem (1988, 2000) and Thomason and Kaufman (1988), contact outcomes are linked to mechanisms that govern language use. In this particular paper, understanding the structural consequences of the transfer of complex loanwords is not only approached as an outcome of language contact, but also through attested tokens of speaker-driven cross-linguistic influence. This case study thus allows us to test the various assumptions proposed in these context-based frameworks based on contemporary patterns of language use.

From what we have seen in Ibatan, there are evident gaps in the frameworks that need to be addressed. For instance, there is still much to know about the linguistic outcomes of symmetrical or balanced bilingualism, where the speakers have (near-)equal dominance in the two languages in their repertoire. Van Coetsem (2000) proposes the neutralization of transfer types, where outcomes linked to both imposition and borrowing transfer may be equally possible. While the literature on bilingualism argues that this is a rare type of bilingualism (cf. Grosjean 1985, etc.), it is still important to consider it within models of language contact to better understand its linguistic consequences. This issue is also deeply connected to the need for a nuanced operationalization of language dominance that goes beyond mere measurement of relative proficiency (cf. Silva-Corvalán and Treffers-Daller 2016, Treffers-Daller 2019, etc.). A gradient approach to language dominance that considers extra-linguistic factors both at the individual and community levels, such as level of exposure, frequency and domains of use, and age of acquisition among many others, definitely allows for a better understanding of the links between bilingual language use and the outcomes of contact-induced change.

32 The diffusion of change across the community is a question best explored within the methods of variationist sociolinguistics, which take into account frequency effects, the social value attached to the forms in question, patterns of speaker interaction, among others.

Additionally, modeling contact outcomes based on individual speaker behavior, while certainly insightful, does not directly address the propagation of change. This relates to Weinreich, Labov, and Herzog's (1968) transition problem in language change more generally, and for language contact more specifically, centers on the question of how to link together Muysken's (2010) aggregates of language contact. If change begins from the variation seen in individual patterns of speech, then what governs the spread of such innovations across the community (cf. Croft 2000)? In language contact, the nature of community bilingualism seems to play an important role (Thomason and Kaufman 1988), but cases where the data does not follow the expected results (as in instances of Type 1 hybrid formations) demand alternative explanations.

All these issues are relevant if we hope to reconstruct past contact scenarios based on contemporary ones. That is, we take the assumption that the mechanisms that apply synchronically must be the same ones that have applied in the past, and this is known as the Uniformitarian Principle. However, the main issue behind this principle is that we cannot assume that the social processes that operate in the present are actually comparable to those that operated in the past. For instance, many of the social concepts and models used to investigate particular linguistic phenomena, such as norms, standards, and prestige, may greatly differ across communities and across time periods (cf. Labov 1994:23, Bergs 2012:96). In reconstructing historical contact scenarios, speaker-based models such as van Coetsem (2000) are within the scope of the Uniformitarian Principle because we can assume that the mechanisms governing human cognition have not changed. At the same time, however, cognitive processes only present one side of the picture. That is, the psycholinguistic notion of language dominance also relies on extra-linguistic factors which are dynamic and are influenced by community-wide factors. There is thus the need to strengthen the current models and frameworks for language contact to better account for these considerations.

5 Conclusion

Because of the history of intense social contact between speakers of Ibatan and Ilokano for the past 150 years, the Ibatan language exhibits contact-induced features across various domains, including morphology, which is said to be dispreferred in language contact. The paper has focused on the structural consequences of lexical transfer in Ibatan, specifically the development of its non-native durative paradigm. While this has been primarily facilitated through complex loanwords, a small number of hybrid formations indicate that the

processes involved in this transfer are more complex, which are linked to overlapping mechanisms of agentivity that govern the multilingual individual and community across various stages in the development of Ibatan.

Contact-induced structural change in Ibatan has resulted in what Kossmann (2010) describes as Parallel System Borrowing, where non-native structures co-exist with their native counterparts. This also relates to Seifart's (2012, 2017) Principle of Morphosyntactic Subsystem Integrity, where it is said that transferring sets of forms is arguably more common than transferring piecemeal. With the case of Ibatan, however, we cannot be fully certain to what extent this has affected morphology, in that many of the forms for verbal morphology are shared between Ibatan and Ilokano, given that the two are closely related languages under the Malayo-Polynesian family.

This is only one of the several issues that concern contact between genetically related and typologically similar languages (cf. Epps, Huehnergard, and Pat-El 2013). Another related matter is understanding how much typological similarity plays a role in language contact (cf. Seifart 2014). For the current study, the verbal morphology shared between Ibatan and Ilokano inherited from PMP and PAN seems to play a role in the transfer of the durative paradigm, in that the RL system can readily accept SL structures. However, perhaps the more relevant question is why this transfer occurred in the first place. Given that the structure already exists natively in Ibatan, why is there a need to develop and maintain a non-native counterpart?

It is then evident that structuralist and constraints-based approach to language contact, while useful in investigating the phenomenon, needs to be supplemented by information grounded on the socio-historical contexts of the speakers. This compartmentalization of morphology, described by Matras (2015) for cases where native and non-native structures are kept distinct in a language, is said to reflect how the speakers flag their bilingual identity. For the Ibatans, they indeed acknowledge their mixed ancestry and history, and they clearly maintain the boundary between Ibatan and Ilokano, even in the early years of the community. This therefore is one of the different factors that motivate the emergence and maintenance of a parallel non-native paradigm in the language.

Acknowledgements

The author thanks Bethwyn Evans and Jennifer Hendriks for their comments on an earlier draft of the manuscript.

Appendix: Glossing Abbreviations

1	1st person	IRR	Irrealis
2	2nd person	IVB	Ibatan
3	3rd person	IVV	Ivatan
AV	Actor voice	LK	Linker
CV	Circumstantial voice	LV	Locative voice
DEI	Deictic	NOM	Nominative
DET	Determiner	NTRL	Realis neutral
DIR	Directional	PFV	Realis perfective
DIST	Distributive	PL	Plural
DUR	Durative	PV	Patient voice
EXT	Existential	RDP	Reduplication
FUT	Future	REC	Reciprocal
GEN	Genitive	REF	Anaphoric reference
ILO	Ilokano	SG	Singular
INC	Inchoative	SOC	Social
INV	Inversion marker	SPA	Spanish
IPFV	Realis imperfective		

Bibliography

- Ackerman, Farrell, James Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In *Analogy in grammar: Form and acquisition*, edited by James Blevins and Juliette Blevins, 54–82. Oxford: Oxford University Press.
- Baayen, Harald. 2008. Corpus linguistics in morphology: Morphological productivity. In *Corpus linguistics: An international handbook (Handbooks of linguistics and communication science 29)*, edited by Anke Lüdeling and Merja Kytö, 899–919. Berlin: Mouton de Gruyter.
- Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall, and Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26, 5–42.
- Bergs, Alexander. 2012. The Uniformitarian Principle and the risk of anachronisms in language and social history. In *The handbook of historical sociolinguistics*, edited by Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre, 80–98. Oxford: Blackwell.
- Blust, Robert and Stephen Trussel. 2020. *The Austronesian comparative dictionary*. <http://www.trussel2.com/ACD/>. (Accessed 12 January 2020.)

- Blust, Robert. 1991. The Greater Central Philippine hypothesis. *Oceanic Linguistics* 9, 104–162.
- Blust, Robert. 2017. Regular metathesis in Batanic (Northern Philippines)? *Oceanic Linguistics* 56 (2), 491–504.
- Blust, Robert. 2019. The resurrection of Proto-Philippines. *Oceanic Linguistics* 58 (2), 153–256.
- Blust, Robert. 2020. Response to comments on “The resurrection of Proto-Philippines”. *Oceanic Linguistics* 59(1/2), 450–479.
- Bybee, Joan. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam: Benjamins.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425–455.
- Campbell, Lyle. 1993. On proposed universals of grammatical borrowing. In *Historical Linguistics 1989*, edited by Henk Aertsen and Robert Jeffers, 91–109. Amsterdam: John Benjamins.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.
- Curnow, Timothy. 2001. What language features can be ‘borrowed’. In *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, edited by Alexandra Aikhenvald and R.M.W. Dixon, 412–436. Oxford: Oxford University Press.
- Dressler, Wolfgang. 1989. Prototypical differences between inflection and derivation. *Zeitschrift für Sprachwissenschaft und Kommunikationsforschung* 42, 3–10.
- Ebarhard, David, Gary Simons & Charles Fennig (eds.) 2022. *Ethnologue: Languages of the World, Twenty-fifth edition*. Dallas: Summer Institute of Linguistics International. <http://ethnologue.com>. (accessed 2022-11-25).
- Epps, Patience, John Huehnergard, and Na’ama Pat-El (eds.) 2013. *Journal of Language Contact* 6.
- Gallego, Maria Kristina. 2020. Ibatan of Babuyan Claro (Philippines)—Language Contexts. *Language Documentation and Description* 17, 87–110.
- Gallego, Maria Kristina. 2022a. Lexical sources for Ibatan. In *The legacy of Consuelo J. Paz: A Festschrift*, edited by Wylan dela Peña, 71–89. Quezon City: University of the Philippines Center for International Studies.
- Gallego, Maria Kristina. 2022b. “The stratigraphy of a community: 150 years of language contact and change in Babuyan Claro, Philippines.” PhD diss., Australian National University. URL: <https://openresearch-repository.anu.edu.au/handle/1885/275585>
- Gallego, Maria Kristina. ongoing. *Consequences of contact: Documenting Ibatan within the multilingual landscape of Babuyan Claro*. London: SOAS, Endangered Languages Archive. URL: <http://hdl.handle.net/2196/00-0000-0000-0014-09E8-3>
- Gardani, Francesco. 2008. *Borrowing of inflectional morphemes in language contact*. Frankfurt: Peter Lang.

- Gardani, Francesco. 2012. Plural across inflection and derivation, fusion and agglutination. In *Copies versus cognates in bound morphology*, edited by Lars Johanson and Martine Robbeets, 71–97. Leiden: Brill.
- Gardani, Francesco, Peter Arkadiev, and Nino Amiridze (eds.) 2015. *Borrowed morphology*. Berlin: De Gruyter Mouton.
- Grosjean, Francois. 1985. The bilingual individual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development* 6, 467–477.
- Gumperz, John Joseph and Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In *Pidginization and creolization of languages*, edited by Dell Hymes, 151–167. Cambridge: Cambridge University Press.
- Haspelmath, Martin and Uri Tadmor (eds.) 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 1996. Word-class changing inflection and morphological theory. In *Yearbook of morphology*, edited by Gerd Booij and Jaap van Marle, 43–66. Dordrecht: Kluwer.
- Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26, 210–231.
- Hay, Hennifer and Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9, 342–348.
- history and typology of western Austronesian voice systems*, edited by Fay Wouk and Malcolm Ross, 17–62. Pacific Linguistics 518. Canberra: Pacific Linguistics.
- Kossmann, Maarten. 2010. Parallel System Borrowing: Parallel morphological systems due to the borrowing of paradigms. *Diachronica* 27 (3), 459–487.
- Labov, William. 1994. *Principles of linguistic change, Volume 1: Internal factors*. Oxford: Blackwell.
- Laca, Brenda. 2008. Derivation. In *Language typology and language universals*, edited by Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, Wolfgang Raible, 1214–1227. Berlin: Walter de Gruyter.
- Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. Doctoral dissertation, University of Hawaii at Manoa, Hawaii, United States of America.
- Liao, Hsiu-chuan. 2020. A reply to Blust (2019) “The resurrection of Proto-Philippines”. *Oceanic Linguistics* 59(1/2), 426–449.
- Maree, Judith, Orlando Tomas, and Rundell Maree. 2012. *Ibatan to English dictionary, with English, Filipino, Ilokano, and Ivatan indices*. Manila: Summer Institute of Linguistics.
- Maree, Judith. 2005. *The Ibatan: A genealogy of the people of Babuyan Claro Island*. Manila: Summer Institute of Linguistics.
- Maree, Rundell. 1982. Ibata’an cultural origins and social organization. MA thesis, University of Texas Arlington.

- Maree, Rundell. 2007. *Ibatan: A grammatical sketch of the language of Babuyan Claro Island*. Manila: Linguistic Society of the Philippines.
- Matras, Yaron and Jeanette Sakel, eds. 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin: De Gruyter Mouton.
- Matras, Yaron. 2015. Why is the borrowing of inflectional morphology dispreferred? In *Borrowed morphology (Language Contact and Bilingualism 8)*, edited by Francesco Gardani, Peter Arkadiev, and Nino Amiridze, 47–80. Berlin: De Gruyter Mouton.
- Meakins, Felicity, Xia Hua, Cassandra Algy, and Lindell Bromham. 2019. Birth of a contact language did not favor simplification. *Language* 95 (2), 294–332.
- Muysken, Pieter. 2010. Scenarios for language contact. In *The handbook of language contact*, edited by Raymond Hickey, 265–281. Oxford: Blackwell.
- Poplack, Shana, David Sankoff, and Cristopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26 (1), 47–104.
- Reid, Lawrence A. and Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5 (2), 433–490.
- Reid, Lawrence. 1992. On the development of the aspect system in some Philippine languages. *Oceanic Linguistics*, 31 (1), 65–91.
- Reid, Lawrence. 2020. Response to Blust “The resurrection of Proto-Philippines”. *Oceanic Linguistics* 59(1/2), 374–393.
- Riesberg, Sonja. 2014. *Symmetrical voice and linking in western Austronesian languages*. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2002. The history and transitivity of western Austronesian voice and voice-marking. In *The history and typology of western Austronesian voice systems*, edited by Fay Wouk and Malcolm Ross, 17–62. Canberra: Pacific Linguistics.
- 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. 2020. Comment on Blust, “The resurrection of Proto-Philippines”. *Oceanic Linguistics* 59(1/2), 366–373.
- Rubino, Carl. 2000. *Ilocano dictionary and grammar*. Hawaii: University of Hawaii Press.
- Seifart, Frank. 2012. The Principle of Morphosyntactic Subsystem Integrity in language contact: Evidence from morphological borrowing in Resígaro (Arawakan). *Diachronica* 29(4), 471–504.
- Seifart, Frank. 2014. Does structural-typological similarity affect borrowability? A quantitative study on affix borrowing. *Language Dynamics and Change* 5, 92–113.
- Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91(3), 511–532.
- Seifart, Frank. 2017. Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics* 7(3), 389–431.

- Silva-Crovalán, Carmen and Jeanine Treffers-Daller (eds.) 2016. *Language dominance in bilinguals: Issues of measurement and operationalization*. Cambridge: Cambridge University Press.
- Smith, Alexander. 2017. The Western Malayo-Polynesian problem. *Oceanic Linguistics* 56(2), 435–490.
- Thomason, Sarah and Terrence Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Thomason, Sarah. 2008. Social and linguistic factors as predictors of contact-induced change. *Journal of Language Contact* 2 (1), 42–56.
- Thomason, Sarah. 2015. When is the diffusion of inflectional morphology not dispreferred? In *Borrowed morphology (Language Contact and Bilingualism 8)*, edited by Francesco Gardani, Peter Arkadiev, and Nino Amiridze, 27–46. Berlin: De Gruyter Mouton.
- Treffers-Daller, Jeanine. 2019. What defines language dominance in bilinguals? *Annual Review of Linguistics* 5. 375–393.
- Van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.
- Van Coetsem, Frans. 2000. A general and unified theory of the transmission process in language contact. Heidelberg: Winter.
- Weinreich, Uriel, William Labov, and Marvin Herzog, 1968. Empirical foundations for a theory of language change. In *Directions for historical linguistics*, edited by Winfred Lehmann and Yakov Malkiel, 95–195. Austin: University of Texas Press.
- Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. The Hague: Mouton.
- Whitney, William Dwight. 1881. On mixture in language. *Transactions of the American Philosophical Association* 12: 5–26.
- Wolff, John 1973. Verbal inflection in Proto-Austronesian. In *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, edited by Andrew B. Gonzalez, 71–91. Philippine Journal of Linguistics Special Monograph Issue No. 4. Quezon City: Linguistic Society of the Philippines.
- Wouk, Fay and Malcolm Ross (eds.) 2002. *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University.
- Zorc, David. 2020. Reactions to Blust's "The resurrection of Proto-Philippines". *Oceanic Linguistics* 59(1/2), 394–425.

The Effects of Language Contact on Lexical Semantics: The Case of Abui

George Saad

1 Introduction

This chapter documents incipient and on-going changes in the Abui verbal lexicon as caused by contact with Alor Malay (AM).¹ Using the *apparent-time construct* from sociolinguistics (e.g. Bailey et al. 1991), it provides detailed documentation of the ongoing semantic shift in three different Abui event domains by comparing their usage across four age-groups. It contributes to our understanding of how semantic shift gradually takes place through contact. It also shows how contact affects verbs in different ways, highlighting why some semantic shifts may be more advanced than others. It answers the question, “What can variation among age-groups in the use of the ‘visual perception’, ‘falling’, and ‘change of state’ verbs tell us about the semantic changes taking place in Abui?”

In language endangerment settings, it is common for the lexicon of the endangered language to shrink (Aikhenvald 2020). One of the ways this happens is that the lexicon loses low frequency words which depict highly specific meanings. Some of the meaning spaces or contexts occupied by these words may then be swallowed up by a semantically related yet more highly frequent word. This process of a word absorbing the space of a semantically related word is known as *generalization* (Blank 1999; Traugott and Dasher 2001).

Generalization is a common semantic change that takes place in healthy (monolingual) language settings as well, but in language endangerment scenarios, which involve reduced input of the endangered language, less frequent words fall into disuse and become displaced by more frequent words at a more rapid pace.

Indeed, while language internal factors, such as frequency and polysemy, may play a role, language external factors, such as the structure of the donor

1 This chapter is based on Chapter 5: Variation and change in verb usage in Saad, George. 2020b. “Variation and Change in Abui: The Impact of Alor Malay on an indigenous language of Indonesia” PhD dissertation, Leiden: Leiden University.

language in a language contact scenario, may also accelerate this process. If the dominant (donor) language uses one word to express what the endangered (recipient) language traditionally expressed as two or more words, then contact is more likely to result in the endangered language favoring one word and dropping the other. This has been shown by countless bilingualism and heritage language studies (Gathercole and Moawad 2010; Polinsky 2008; Jarvis and Pavlenko 2008; Backus, Seza Doğruöz, and Heine 2011; Weinreich 1953). This is usually attributed to the fact that conceptual representations associated with the distinction have not been mapped out (Jarvis & Pavlenko, 2008). Furthermore, evidently, not all words are affected in the same way; each word has a life trajectory of its own.

Abui, like many languages of eastern Indonesia, is under threat from the regional Malay variety, in this case, Alor Malay. This is causing the Abui lexicon of younger speakers, labeled here as Light Abui, to show early signs of generalization when compared to the Abui lexicon of older speakers, labeled as Traditional Abui.² There is a positive correlation between age and exposure to Abui; the younger the speaker, the less exposure they have to Abui and thus the more likely they are to exhibit generalization.

The goal of this chapter is to document in detail the semantic shift taking place in the Abui lexicon by observing how various age-groups express three domains: *visual perception*, *falling*, and *change of state*. These event domains were specifically selected for investigation because, in an Abui corpus collected from 66 speakers, they a) were the most commonly used, b) showed the most advanced signs of generalization, and c) fell under the same translation equivalent category. In other words, for each domain, one Alor Malay lexical item corresponded to at least two Abui items. These event domains are considered strong candidates to be at the forefront of generalization.

As demonstrated in this paper, Light Abui shows clear signs of generalization, replacing the two or more Traditional Abui forms with one form. This semantic change is not categorical but continuous, as exhibited through the variation in the use of these forms by the three younger age-groups. In addition, each verb domain tells a story of its own. This study investigates language change under the assumptions of the *apparent-time construct* (Bailey et al. 1991): differences in young people's speech are heralded as being indicative of incipient language change. At the same time, the Abui case also challenges this construct because Abui exhibits a phenomenon recently described for Indone-

2 The term is 'Light Abui' is taken from the Australianist model of describing the contact-induced variety spoken by younger speakers when compared to the more traditional variety spoken by older speakers (O'Shannessy 2005).

sia known as ‘delayed/adult vernacular production’ (Anderbeck 2015, 27; see also Peddie 2021): children only produce the vernacular when they reach young adulthood. This suggests that only a real-time longitudinal study can affirm whether variation observed today will lead to change.

This paper is structured as follows: Section 2 discusses the sociolinguistic setting. The methodology, which discusses a production task is laid out in section 3. The three event domains are discussed in section 4; for each verb domain, a description is given of the use in Traditional Abui, Light Abui, and Alor Malay. A general discussion is offered in section 5, followed by a conclusion in section 6.

2 Sociolinguistic Setting

Abui is a Timor-Alor-Pantar (Papuan) language spoken by around 17,000 people central and west-central Alor, eastern Indonesia (Kratochvíl 2007); see Figure 11.1. It is the largest indigenous language spoken on the Alor archipelago and also the earliest and most well-described (Du Bois 1944; Nicolspeyer 1940; Stokhof 1984; Kratochvíl 2007; Kratochvíl and Delpada 2008; Kratochvíl 2011; 2014; Saad 2020a). The Abui language, and especially the Abui spoken in the village of Takalelang, is under threat from the regional *lingua franca*, Alor Malay, and to a lesser extent the national *lingua franca*, Indonesian. The two *lingua francas* sit on a basilect-acrolect cline (Baird, Klamer, and Kratochvíl in prep.; Paauw 2008). This is mostly evident in speakers under the age of 40 (born after 1975) and particularly visible among speakers below the age of 25 (born after 1990). One of the main reasons propelling this shift was a migration of inhabitants from mountain villages to the northern coast (see present day Takalelang in Figure 11.1) which brought them more in contact with members of different ethno-linguistic backgrounds as well as a regime favouring the use of Indonesian in institutions such as churches, health centers, and schools. As such, there were strong movements within some of these institutions to force parents to start raising their children in Alor Malay/ Indonesian and ban the use of Abui among children at home and at school.

In order to investigate how contact with Alor Malay has affected Abui language use, a distinction is made between Traditional Abui, spoken by *kalieta* ‘elders’ (age: 40+) and Light Abui, spoken by three age-groups: *moqu* ‘(pre)adolescents’ (age: 9–16), *neeng abet/ maayol maak* ‘young adults’ (age: 17–25 years), *kalieta* ‘adults’ (age: 26–34). Traditional Abui is spoken by elders who were raised by their parents as Abui L1 speakers, and only learned Malay when entering school. In contrast, the three groups speaking Light Abui all received some Alor Malay in their language socialization, with the group of

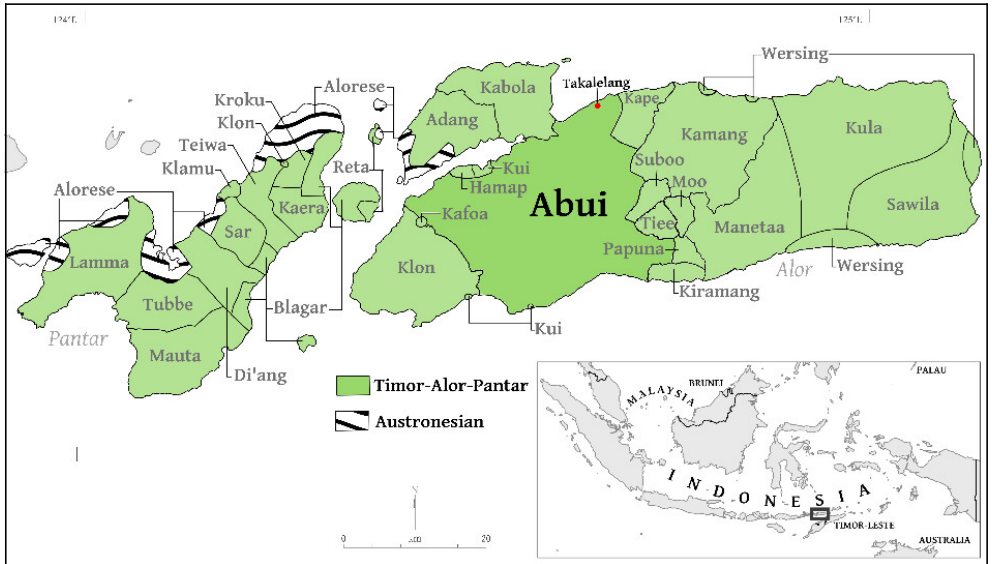


FIGURE 11.1 Map of indigenous languages of Alor and Pantar

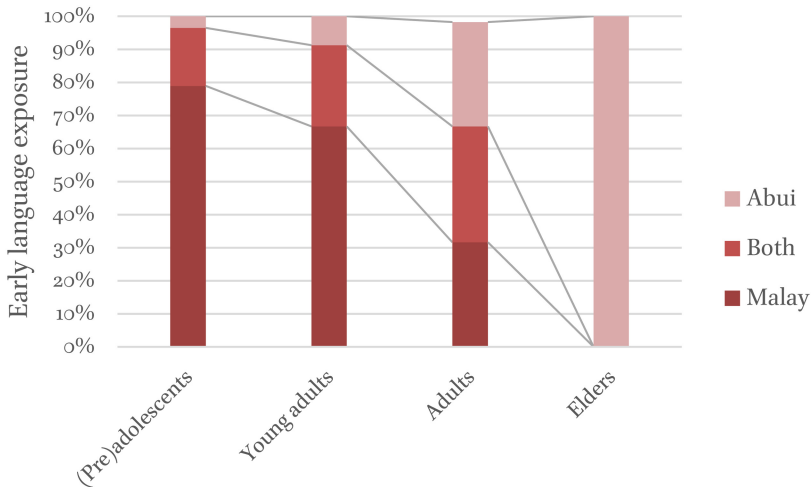


FIGURE 11.2 Early language exposure among the three Light Abui groups

moqu (pre)adolescents (age: 9–16) having the most exposure to Alor Malay and the least to Abui. A questionnaire was carried out to a total of 66 speakers within these groups to try to determine their early language exposure history (for more detailed information on how this was done, see Saad (2020b, 116–127)). The results are depicted in Figure 11.2, illustrating whether speakers self-reported being raised in Abui, both, or Alor Malay. The group of elders

unanimously reported being raised in Abui by the age of 10, while the three Light Abui groups showed varying degrees of multilingualism in their upbringing. There is a clear gradual increase in exposure to Abui with ageing. Most notable, however, is the similar low exposure to Abui among (pre)adolescent and young adults. We will return to these figures in the discussion in section 5.

The three Light Abui groups were constructed on the basis of life-stage (for a more elaborate discussion, see Saad 2020b, 116–127). These categories correspond roughly to emic Abui age-constructs which have been observed since the time of Du Bois (1933) and continue to do so today. They were backed up using data from ethnographic interviews with around 10 Abui speakers which outlined the details of these age-groups (Saad 2020b, 116–127).

One of the interesting features about Light Abui is the pattern of Abui language acquisition which has been termed ‘delayed/adult vernacular production’ (Anderbeck 2015, 27; see also Peddie 2021). Children grow up overhearing Abui from their older peers, but only really become active speakers during or after adolescence. This phenomenon has only recently been described but appears to be much more widespread in Indonesia and Melanesia (Anderbeck 2015; Saad 2020a; Peddie 2021).

A summary of the age-groups is provided in Table 11.1 (Saad 2020b, 96–99) (adapted from Saad 2020b, 128). The age-boundaries themselves are rough estimates of these life-stages and allow for the objective and empirical sampling of the community.³

TABLE 11.1 Age-groups used in this study

Age-group	Range	Life-stage	Language history
<i>Moqu</i> ‘(Pre) adolescents’	9–16	Still learning essential, daily chores. Speak AM to peers, parents, and adults. Are addressed in AM, except by grandparents. Understand Abui, but do not speak it on a frequent basis.	Were raised exclusively in AM by parents. Spoke AM to everyone.
<i>Neeng abet/maayol maak</i> ‘Young adults’	17–25	Sexually mature and preparing for marriage. Speak AM with peers and Abui with adults/elders.	Were raised mostly in AM by parents. Spoke AM to everyone.

3 The lower age limit of 9 was set because speakers below nine were unwilling or felt uncomfortable being recorded. With regards to adults and elders, the Abui category *kalieta* describes

TABLE 11.1 Age-groups used in this study (*cont.*)

Age-group	Range	Life-stage	Language history
<i>Kalieta</i> 'Adults'	26–34	Typically married and/or bear child(ren). Speak AM and Abui with other adults. Speak AM with children.	Raised by both Abui and AM by parents. Spoke a mix with peers.
<i>Kalieta</i> 'Elders'	40–75	Married. Can participate in ritualized negotiations. Speak Abui with peers and parents. Speak Alor Malay with children.	Raised exclusively in Abui. Learned some form of Malay when entering school at age ~ 6–12

3 Methodology

In order to collect comparable data across age-groups, a video elicitation task, known as the Surrey Stimuli (Fedden, Brown, and Corbett 2010; Fedden and Brown 2017) was used. The Surrey Stimuli task was carried out with 66 speakers, whose details are outlined in Table 11.2.

The Surrey Stimuli video elicitation task involved showing speakers 40 short video-clips exhibiting a variety of events (Fedden et al. 2014). These included, among many others, the three event domains of visual perception, falling, and change of state. While all the responses were being transcribed and annotated, they were also being double checked by older, native speakers for grammaticality and felicitousness. After this process, it was clear that there was considerable age-related variation in the choice of verbs for certain events: some verbs appeared to be generalized to other contexts. While many other verbs also showed variation among speakers, the three event domains of visual perception, falling, and change of state were selected for in-depth investigation for two reasons. First, these domains contained the verbs that were the most frequently used in the production task, while the other types of verbs were used sporadically and thus did not fulfill sampling criteria. Second, these event domains were present in clips eliciting two polarities of a given semantic feature (e.g. both \pm CONTROL as opposed to just +CONTROL). In other words, this

both, but there are clear differences in the language histories of these groups as well as, to some extent, their status in the community.

TABLE 11.2 Breakdown of participants

Groups	Abui name	Age-range	M	F	Total
'(Pre)adolescents'	<i>Moqu</i>	9–16	9	10	19
'Young adults'	<i>Neeng abet/ maayol maak</i>	17–25	10	9	19
'Adults'	<i>Kalieta</i>	26–34	10	9	19
'Elders'	<i>Kalieta</i>	40–75	4	5	9
Total		9–75	33	33	66

made it possible to study whether, for example, *-ien-* was used appropriately in its target context 'see [-CONTROL]' as well as whether *(-)wahai* was used appropriately in its target context 'look at [+CONTROL]', instead of just one of these polarities. This allowed for the testing of directionality of generalization. Every verb was judged as being a match if it was used in its appropriate context and a mismatch if it was used in a different context. These are presented in Table 11.3.

TABLE 11.3 Coding of event domains

Event domain	Context	Match	Mismatch
Visual perception	'see [-CONTROL]'	<i>-ien-</i>	<i>(-)wahai</i>
	'look at [+CONTROL]'	<i>(-)wahai</i>	<i>-ien-</i>
Falling	'fall over [-ELEVATION]'	<i>-quoil-, -kaai</i>	<i>(el ong) hayeei</i>
	'fall from above [+ELEVATION]'	<i>(el ong) hayeei</i>	<i>-quoil-, -kaai</i>
Change of state	'wake up [-COP]'	<i>-minang, -tein-</i>	<i>-ru-</i>
	'get up [+COP]'	<i>-ru-</i>	<i>-minang-, -tein-</i>

4 Three Event Domains in Abui and Alor Malay

This section discusses the three event domains in Abui that were selected for investigation: *visual perception*, *falling*, and *change of state*. Given that contact is argued to play a role, a description is also given of the subsequent translation equivalents in Alor Malay. The main differences between Abui and Alor Malay in these three domains are that Abui uses a narrow system while Alor Malay uses a broad system (Gathercole and Moawad 2010). This means that

Abui uses at least two verbs to lexically distinguish two given contexts, while Alor Malay simply uses one verb for both of these contexts.⁴

Sections 4.1, 4.2, 4.3 discuss these three event domains in detail. Each of these sections has three parts. In the first part, a description is given of the distinctions lexicalized in Traditional Abui. Each of the three event domains is split according to [\pm FEATURE] and examples of the use of each verb in its designated context are given. In order to fully understand the distribution of these verbs in the lexicon, this section includes example sentences as well as a presentation of polysemy and token counts of word frequency in a large Abui corpus of spontaneous and elicited texts. In the second part, there is a description of the developments in Light Abui, comparing the four age-groups' token frequencies. The third part gives a description of the translation equivalents in Alor Malay.⁵

4.1 *Verbs of Visual Perception*

Abui follows a cross-linguistically common trend of distinguishing between 'seeing' and 'looking at'. Given that vision has been shown psychologically to be the dominant human sense (Alais and Burr 2004; Stokes and Biggs 2014), many studies have shown that a large number of languages have adapted to this by a) using visual perception verbs more frequently than verbs for other types of perception and b) lexically differentiating different types of visual perception (Levinson and Majid 2014; Viberg 1983; Winter, Perlman, and Majid 2018).

What is of interest here is the lexical differentiation between different types of visual perception. Cross-linguistically, it is extremely common for languages to use a *dynamic system* where they encode a distinction between the experience verb 'see' and the activity verb 'look at' (Levinson and Majid 2014; Viberg 1983). *Experience* refers to 'a state (or inchoative achievement) that is not controlled', while *activity* here refers to 'an unbounded process that is consciously

4 These distinctions are found in other Alor-Pantar languages, such as Kamang, for example. Sometimes, the Abui forms are also cognate with the Kamang forms, though this is not always the case. Compare Kamang *kawaila* 'fall over' vs. *mu'tan* 'fall from above' (Schapper and Manimau 2011, 224; 249) and Abui *-quoil-* 'fall over' vs. *hayeei* 'fall from above'.

5 These three event domains represent a small sample of domains where Abui uses a narrow system, while Alor Malay uses a broad system. Another example includes the verbal domain of 'eating': Abui, *nee* 'eat (soft food)' and *takai* 'chew/ eat (hard food)'; Alor Malay, *makan* 'eat, chew on'. There are of course numerous examples where Alor Malay uses a narrow system, while Abui uses a broad system. One example is Abui *buuk* 'drink; smoke' and Alor Malay *minum* 'drink' and (*isap*) *rokok* 'smoke' (Kratochvíl p.c.). However, not too many of these examples were found in the corpus of Surrey Stimuli data.

TABLE 11.4 Verbs of perception

Event domain	Semantic feature	[± Feature] Sense	Traditional Abui	Light Abui	Alor Malay
Visual perception	[±CONTROL]	[-CONTROL] 'see' [+CONTROL] 'look at'	<i>-ien-</i> (-) <i>wahai</i>	(-) <i>wahai</i>	<i>lihat</i>

controlled by a human agent' (Viberg 1983, 123). With these characteristics in mind, the feature [±CONTROL] is used to differentiate these two verbs.

In Traditional Abui, the context 'see [-CONTROL]' is expressed by the experience verb *-ien-*, while the context 'look at [+CONTROL]' is expressed by the activity verb (-)*wahai*. In Light Abui, the form (-)*wahai* becomes generalized. In Alor Malay, there is only one form: *lihat*; see Table 11.4.

4.1.1 Traditional Abui

Example (1) illustrates the use of the experience verb *-ien-* in a 'see [-CONTROL]' context. It is a response to a clip from the Surrey Stimuli (discussed in 3 showing a man walking by, failing to 'see the banana' on the floor and then stepping on it. The experience verb *-ien-* is used to describe the event of 'not seeing the banana'.

(1) 'see [-CONTROL]'

Neeng nuku laak-i me mai balei h-ien naha.
 man one walk-PFV come.IPFV COND banana 3.PAT-see NEG
 'As a man passed by, he didn't see the banana.' [ss.4of.24]

The use of the activity verb (-)*wahai* 'look at [+CONTROL]' is shown in (2). Example (2) is a response to a clip where a man is sitting and actively 'looking at the cheese'.

(2) 'look at [+CONTROL]'

Neeng nuku do mit ba keju he-wahai.
 man one PROX sit LNK cheese 3.LOC-look.at
 'A man is sitting and looking at the cheese.' [ss.4of.24]

An important point to make about the word *-ien-* is that it is more polysemous than the verb (-)*wahai* 'look at'. It may denote other verbal meanings such as 'find, know, understand', as well as nominal senses 'eye' and 'backside'.

TABLE 11.5 Frequency of visual perception verbs (Kratochvíl corpus)

Verb	Sense	Tokens	% of total number of verbs (N = 6450)
<i>-ien-</i>	All senses	434	6.72%
<i>-ien-</i>	- 'see'	84	0.51%
(-) <i>wahai</i>	All senses	226	3.50%
(-) <i>wahai</i>	- 'look at'	226	3.50%

In the Kratochvíl corpus, the form *-ien-* with all its senses included appears 434 times (6.72% out of a total verb count of 6450). This is almost double the amount that (-)*wahai* appears in (226 tokens, 3.50%). However, *-ien-* with the strict sense of 'seeing' actually appears less frequently (84 tokens, 0.51%) than the verb (-)*wahai* (226 tokens, 3.50%). These figures are presented in Table 11.5. What this shows is that strictly in the domain of visual perception, (-)*wahai* is more frequent than *-ien-* but less polysemous.

4.1.2 Light Abui

Both (pre)adolescents and young adults generalized the form *-wahai* 'look at' to contexts required *-ien-* 'see', while adults did not. This is shown in Table 11.6 in the column labeled proportion of mismatches. The proportion of mismatches illustrates how often a speaker used the mismatch verb, (-)*wahai* 'look at', in a 'see [-CONTROL]' context when the form *-ien-* was expected. In the 'Proportion' column, the denominator shows how many times a group produced a 'see [-CONTROL]' context, while the numerator shows how many times a group used (-)*wahai* 'look at' instead of *-ien-* 'see'.⁶

TABLE 11.6 Proportion of mismatches for *-ien-* 'see [-CONTROL]' target

Group	Speakers	Proportion	SD
(Pre)adolescents	19	8/11 (73%)	.47
Young adults	19	14/17 (82%)	.39
Adults	19	1/13 (8%)	.28
Elders	9	0/4 (0%)	.0

6 Recall from § 4.3.1 that the amount of 'contexts' produced is dependent on both the stimu-

As can be seen, (pre)adolescents (8/11; 73%) and young adults (14/17; 82%) had a high number of mismatches, compared to adults (1/13; 8%) and elders (0/4; 0%), who had close to none. These differences are significant (see Saad 2020a, 289 for statistical tests).

There were no signs that speakers showed the opposite pattern, namely that they would instead use the *-ien-* verb in a (-)*wahai* 'look at' context. This suggests that the verb (-)*wahai* 'look at' is becoming generalized and displacing the form *-ien-* and that the feature [CONTROL] is being lost in the domain of visual perception.

4.1.3 Alor Malay

As opposed to Abui, Alor Malay does not lexically encode a distinction between visual activity and visual experience, a tendency which is considered cross-linguistically rare (Viberg 1983). Alor Malay uses *lihat* as a generic term for both 'see' and 'look at'. This is shown in (3a–b), which presents responses to stimuli, which in Abui elicited the verbs *-ien-* and (-)*wahai* respectively (see (1) and (2)).⁷

(3) Alor Malay

a. 'see'

Laki-laki satu jalan datang ni=yang dia tidak lihat
 man one walk come PROX=RE 3SG NEG visually.perceive
pisang.
 banana
 'As a man passes along, he does not see the banana.' [SS.4OF.AM]

b. 'look at'

Laki-laki duduk ko lihat keju.
 man sit LNK visually.perceive cheese
 'A man is sitting and looking at some cheese.' [SS.4OF.AM]

In summary, Abui lexicalizes visual perception verbs according to the feature [\pm CONTROL]. The verb *-ien-* 'see' refers to an uncontrolled visual experience, while (-)*wahai* 'look at' refers to a controlled visual activity. The verb *-ien-* in its

lus shown and the construction a speaker uses. Because speakers were free to describe the clips in ways they saw fit, not all speakers produced constructions that could be used for this particular study. This is why the denominators differ per group.

7 Like Abui, Alor Malay does not mark tense grammatically; however, it may indicate tense through temporal adverbs. Throughout this paper, in the absence of temporal adverbs, the default tense used is the present tense.

specific sense denoting ‘see’ occurs less frequently than the verb (-)*wahai* ‘look at’. However, *-ien-* is much more polysemous and may be used in various grammatical contexts; when taking into account its other senses, it appears almost twice as much as the verb (-)*wahai* ‘look at’. Finally, Alor Malay has one only verb *lihat* for the generic act of visual perception.

4.2 Verbs of Falling

In the event domain of ‘falling’, traditional Abui verbs are specified for the feature [\pm ELEVATION], lexically distinguishing between the synonyms *-quoil-* and *-kaai*, both denoting ‘falling over [$-$ ELEVATION]’ and *hayeei* denoting ‘falling from above [$+$ ELEVATION]’. The main difference between the two polarities is that *-quoil-* and *-kaai*, ‘falling over [$-$ ELEVATION]’ are used for nouns which are upright and then fall over, such as a person or a tree falling over. Conversely, *hayeei* ‘[$+$ ELEVATION] fall from above’ is used for nouns which have landed on a surface lower than their initial starting point. This typically includes coconuts falling from trees, balls falling from the sky and people falling from motorbikes. In Light Abui, speakers generalize the verb *hayeei* to all contexts. In Alor Malay, the term for all types of falling is *jatu*. These distinctions are depicted in Table 11.7.

TABLE 11.7 Verbs of falling

Event domain	Semantic feature	Context [\pm FEATURE]	Traditional Abui	Light Abui	Alor Malay
Falling	[\pm ELEVATION]	[$-$ ELEVATION] ‘fall over’ [$+$ ELEVATION] ‘fall from above’	<i>-quoil-</i> , <i>-kaai</i> (<i>el ong</i>) <i>hayeei</i>	<i>hayeei</i>	<i>jatu</i>

4.2.4 Traditional Abui

Abui has two synonymous verbs expressing the sense of ‘falling over [$-$ ELEVATION]’, *-quoil-* and *-kaai*, as shown in (4a–b). In both of these examples, the man is walking along a flat plain and then falls over, hence the use of either of these two verbs.

- (4) ‘fall over [$-$ ELEVATION]’
 - a. *Neeng nuku laak-i me mai da-quoil-i.*
 man one walk-PFV come COND 3.REFL.PAT-fall.over-PFV
 ‘As a man came along, he fell over.’ [ss.4of.69]

b. *Neeng nuku laak~laak-i ba me kabelang-di ba*
 man one RDP~walk-PFV LNK come trip-INCH.PFV LNK
da-kaai.

3.REFL.PAT-fall.over

'A man came scurrying along, tripped, and fell over.' [ss.3of.41]

In contrast, in example (5), the verb *hayeei* 'fall from above' is used to describe an event where a banana falls from above onto the flat surface of a standing log.

(5) 'fall from above [+ELEVATION]'

Balei san nuku bataa tuku tahang hayeei.

banana ripe one wood CLF on.top fall.from.above

'A ripe banana fell on top of a log.'

Another important difference between the 'fall over' verbs, *-quoil-* and *-kaai*, on the one hand and the 'fall from above' verb *hayeei* on the other is in their polysemy. The verbs, *-quoil-* and *-kaai*, are not at all polysemous, while *hayeei* is, having a richer array of senses than just 'fall from above'. Its core sense 'fall from above' has been extended to other domains, including: 1) 'something bad befalling someone', 2) '(get) hit' 3) 'close a door', 4) '(arrive) until a certain point'.

In addition to and in spite of its polysemy, in absolute terms, it is also much more frequent, as shown in Table 11.8. It accounts for 6.81% of all the 6450 verbs in the Kratochvíl corpus, while the 'fall over' verbs, *-quoil-* and *-kaai* combined, occur in only 22 tokens, accounting for only 0.34% of the total number of verbs. Even when we exclude the additional senses, *hayeei* in its strict sense 'fall from above' still occurs in 171 tokens (2.65%), which still greatly outnumbers *-quoil-* and *-kaai* combined.

This points to the prevalence of, not only the lexical item *hayeei* with respect to either *-quoil-* (439 vs. 16) and *-kaai* (439 vs. 6), but also the sense 'fall from above' with respect to the sense 'fall over' (171 vs. 22).

TABLE 11.8 Frequency of falling verbs (Kratochvíl corpus)

Verb	Sense	Tokens	% of total number of verbs (N = 6450)
<i>-quoil-</i>	'fall over'	16	0.25%
<i>-kaai</i>		6	0.09%
<i>hayeei</i>	All senses	439	6.81%
	- 'fall from above'	171	2.65%

4.2.5 Light Abui

In Light Abui, there is a strong preference for the form *hayeei* ‘fall from above to be generalized and used in ‘fall over [-ELEVATION]’ contexts typically warranting the verbs, *-quoil-* or *-kaai*. (Pre)adolescents and young adults exhibited generalization 87% (20/23) and 100% (29/29) of the time, while adults did so 48% (14/29), all of which are statistically significant with regards to the Traditional Abui of elders (see Saad 2020b, 299–300 for statistics). The figures are shown in Table 11.9.

TABLE 11.9 Production data: Proportion of mismatches for *-quoil-/kaai* ‘fall over [-ELEVATION]’ target

Group	Speakers	Proportion	SD
Elders	9	0/9 (0%)	.00
Adults	19	14/29 (48%)	.51
Young adults	19	29/29 (100%)	.00
(Pre)adolescents	19	20/23 (87%)	.34

There was no evidence to suggest that Light Abui speakers generalized in the opposite direction, namely using the forms *-quoil-/kaai* ‘fall over [-ELEVATION]’ where *hayeei* ‘fall from above [+ELEVATION]’ was required (Saad 2020b, 300). Thus, there is a clear pattern: (Pre)adolescents, young adults, and adults generalize the nontarget form *hayeei* ‘fall from above [+ELEVATION]’ to ‘fall over [-ELEVATION]’ contexts. This suggests that the verb *hayeei* is becoming generalized and displacing the forms *-quoil-/kaai* and that the feature [ELEVATION] is being lost in the domain of falling.

4.2.6 Alor Malay

In Alor Malay, there is only one lexical item available for ‘fall’, *jatu*, which is unspecified for elevation: the senses ‘falling over’ as in (6a) and ‘falling from above’ as in (6b) are both expressed by the same verb, *jatu*; compare (4b)–(5).

(6) Alor Malay

- a. *Laki-laki satu ada jalan datang dia terantuk ko langsung*
 man one PROG walk come 3SG trip LNK immediately
jatu.
 fall
 ‘As a man passes by, he trips and falls.’ [SS.AM.40F]

- b. *Pisang jatu di atas kayu.*
 banana fall LOC top wood
 'A banana falls on top of a log' [ss.11m.am.3]

In summary, Abui lexicalizes falling verbs according to the feature [\pm ELEVATION]. The synonyms *-quoil-* and *-kaai* '[-ELEVATION] fall over' refer to a falling event where an entity which is already partially on the ground falls completely to the ground. In contrast, the verb *hayeei* '[+ELEVATION] fall from above' refers to a falling event where the entirety of an entity is at a higher starting point and falls onto a lower landing point. In addition to these componential differences, *hayeei* 'fall from above' is also more polysemous than *-quoil-* and *-kaai* 'fall over'. As such, in absolute terms it is also much more frequent as well. Moreover, if we only consider the strict sense of *hayeei* 'fall from above' and exclude its other senses, then it is still more frequent than *-quoil-* and *-kaai* 'fall over'. Finally, Alor Malay uses one verb *jatu* 'fall' to encode the generic act of falling.

4.3 Verbs of Change of State

The third event domain discussed here is 'change of state'. In this domain, Traditional Abui lexicalizes distinctions in both event semantics and argument structure. With respect to event semantics, Traditional Abui lexicalizes verbs based on the feature of [\pm CHANGE OF POSTURE] (occasionally also shortened to [COP]). The principle distinction in verbs of change of state we are concerned with is between the two senses: 'wake up [-CHANGE OF POSTURE]' and 'get up [+CHANGE OF POSTURE]'.⁸ Specifically, the sense 'wake up' involves a change of state from sleeping consciousness to waking consciousness *without* a change of posture. On the other hand, the sense 'get up' involves a change of state by moving into an upright posture, without necessarily a change in consciousness. These are summarized in Table 11.10.

4.3.7 Traditional Abui

The 'wake up [-CHANGE OF POSTURE]' sense further lexicalizes verbs according to argument structure, with the root *-tein-*⁸ being used for transitive clauses of 'waking someone up' and *-minang-*⁹ being used in intransitive clauses of 'someone waking up by themselves'. The 'get up [+CHANGE OF POSTURE]' sense, on the other hand, uses one verb stem *-ru-*⁹ for both transitive and intransitive clauses, with the choice of agreement prefix (*ha-* or *da-* for third

8 In some parts of this paper where space is limited, change of posture is abbreviated to [COP].

9 The root *-ru-* may or may not involve a change in conscious state.

TABLE 11.10 Change of state verbs

Semantic feature	Senses [±FEATURE]	Traditional Abui	Light Abui	Alor Malay
[±CHANGE OF POSTURE]	[-CHANGE OF POSTURE] 'wake up'	<i>-minang-</i> (INTR) <i>-tein-</i> (TR)	<i>da-rui</i> (INTR) <i>ha-rui</i> (TR)	<i>bangun</i> (INTR) <i>kasi bangun</i> (TR)
	[+CHANGE OF POSTURE] 'get up'	<i>-rui-</i>		

person) determining transitivity: for transitive verbs, the *ha-* '3.PAT' inflection indexes a P argument, while for intransitive verbs, the *da-* '3.REFL.PAT' inflection indexes S arguments.

These distinctions in both event semantics and argument structure are exemplified in examples (7a–b)–(8). Example (7a) illustrates the use of the verb *-tein-* 'wake up TR [-CHANGE OF POSTURE]'. Here, the child is woken up by the father but is not physically raised up; instead, he remains lying on the ground, hence the component [-CHANGE OF POSTURE]. Example (7b) illustrates the use of the intransitive form of the sense 'wake up INTR [-CHANGE OF POSTURE]', expressed by the form *-minang-*. Here, the man woke up by himself while he was seated against a wall and he subsequently remained seated, also involving a lack of change of posture.¹⁰

(7) 'wake up [-CHANGE OF POSTURE]'

a. Transitive

Neeng moqu nuku anei taa ya he-maama di
 man child one ground sleep.IPFV SEQ 3.AL-father 3.AGT
me ha-tein-a.
 come.IPFV 3.PAT-wake.up-IPFV

'A small boy is sleeping on the ground, his father comes along and wakes him up.' [ss.4of.24]

¹⁰ As a result of these suppletive forms, there are restrictions on pronominal markers. The use of the reflexive *da-* '3.REFL.PAT' on *-tein-* is ungrammatical, as in **dateina* 'woke himself up'. Similarly, the use of the nonreflexive *ha-* '3.PAT' as in *haminangda* 'woke someone else up' is also ungrammatical.

b. Intransitive

Neeng nuku tadei haba oro marak-di ba
 man one sleep.PFV but DIST startle-PFV LNK
da-minang-di.

3.REFL.PAT-be.conscious-PFV

'A man was asleep (leaning against something), but got startled and woke up.' [ss.43f.25]

In (8a–b), the verb *-ru-i-* 'get up' entails the component [+CHANGE OF POSTURE]. It is derived from the root *ru-i-* 'be erect'. In the transitive 'get up' example, (8a), Ata was lying down, looking at his phone; then Simon came and dragged him up, causing him to be upright; the pronominal prefix *ha-* indexes a third person P argument. In the intransitive example, (8b), the man was just sitting against the wall, and then got up and left; the reflexive pronominal prefix *da-* indexes a third person S argument. Both (8a–b) imply a change of posture, hence the use of *-ru-i-* 'get up'.

(8) 'get up [+ CHANGE OF POSTURE]'

a. Transitive

Simon di Ata ha-ru-i-di

S. 3.AGT A 3.PAT-erect-INCH.PFV

'Simon raised Ata.' [FN.43F]

b. Intransitive

Neeng nuku mit-di da-ru-i-di

man one sit-INCH.PFV 3.REFL.PAT-erect-INCH.PFV

'A man was seated (and then) got up.' [ss.3of.41]

Another important difference between the verbs *-tein-/-minang-* 'wake up' and *-ru-i-* 'get up' is that the verb *-ru-i-* is more polysemous. It may occur in a larger number of grammatical contexts and it can index both animate and inanimate targets. It may be used for causing humans to get up as well as objects, such as houses, planks, or motorbikes. It can also be used to index intangible nouns, such as 'history', 'stories', or 'discussion points'. When the argument is inanimate, new senses are derived, comparable to 'resurrect', 'set straight', or 'raise'. The verbs *-tein-/-minang-* are more restricted in that they typically only index animate arguments.¹¹

¹¹ The verb *-minang-* must always index an animate argument. However, it can addition-

TABLE 11.11 Frequency of change of state verbs (Kratochvíl corpus)

Verb	Sense	Tokens	% of total number of verbs (N = 6450)
<i>-tein-</i>	'wake up'	4	0.06 %
<i>-minang-</i>	All senses	10	0.15 %
	- 'wake up'	2	0.03 %
<i>-rui-</i>	All senses	68	1.05 %
	- 'get up'	59	0.91 %

In terms of frequency data from the Kratochvíl corpus, Table 11.11 illustrates that the 'wake up' verbs *-tein-* (4 tokens, 0.06 %) and *-minang-* (2 tokens, 0.03 %) are much less frequent than the *-rui-* 'get up' verb (59 tokens, 0.91 %). The verb *rui-* 'get up' occurs 68 times (1.05 %) if we include all senses.

4.3.8 Light Abui

In Light Abui, there is a strong preference for the form *-rui-* 'get up [+CHANGE OF POSTURE]' to be generalized over to 'wake up [-CHANGE OF POSTURE]' contexts, and for the verbs *-tein-* (TR)/*-minang-* (INTR) to drop out; see Table 11.12. This was statistically significant for (pre)adolescents (22/27; 81 %) and young adults (20/34; 59 %). In addition, adults also showed some propensity for generalization (9/31; 29 %). Statistics are shown in Saad (2020b, 301).

Table 11.12 shows a clear pattern: (Pre)adolescents and young adults generalize the nontarget form *-rui-* 'get up' to 'wake up [-CHANGE OF POSTURE]' contexts. However, all groups use the target form in 'fall from above [+CHANGE OF POSTURE]' contexts. This suggests that the verb *-rui-* 'get up' is becoming generalized and displacing the forms *-tein-/minang-* and that the feature [CHANGE OF POSTURE] is being lost in the domain of change of state.

4.3.9 Alor Malay

In Alor Malay, only one verb exists for the relevant change of state event domain. The Alor Malay term *bangun* lumps together the two senses lexically differentiated in Abui, 'wake up' and 'get up'. It says nothing about whether

ally add another argument using the locative prefix to derive the meaning 'remember something' (lit. 'become conscious of something'). In this respect, it is also polysemous, having the meaning 'wake up' and also 'remember something'.

TABLE 11.12 Production data: Proportion of mismatches for *tein-/minang* 'wake up [-CHANGE OF POSTURE]' target

Group	Speakers	Proportion	SD
(Pre)adolescents	19	22/27 (81%)	.40
Young adults	19	20/34 (59%)	.50
Adults	19	9/31 (29%)	.46
Elders	9	0/12	.00

the sleeping animate being has moved upright or opened their eyes or not. With respect to argument structure, there are also formal differences: Alor Malay transitive clauses involve the use of the causative marker *kasi* 'give' in a serial verb construction, while intransitive clauses simply use the verb *bangun*.

Examples (9a–b) illustrate the use of the intransitive *bangun* 'wake up/get up'. The examples are taken from responses in Alor Malay to the same elicitation stimuli presented to speakers in Abui in examples (7b)–(8b). The verb *bangun* 'wake up, get up' is used to express the two senses lexically differentiated by Abui and corresponds to Abui *-minang-* and *-ru-* respectively.

(9) Alor Malay

a. 'wake up (INTRANSITIVE)'

Dia kaget bangun habis ada lihat kiri kanan.

3SG shocked get.up SEQ PROG look left right

'He got startled and woke up; then, he was looking left and right.'

b. 'get up (INTRANSITIVE)'

Dia bangun ko jalan.

3SG get.up LNK walk

'He gets up and leaves.'

Turning now to the transitive usage, examples (10a–b) illustrate the use of *kasi bangun* 'wake s.o up/erect s.o/sth', composed of the causative *kasi* 'give' and *bangun* 'wake up, erect, get up'. Example (10a) addresses the 'wake up sense' which implies a lack of change of posture, while (10b) illustrates the 'get up' sense which implies a change of posture.

(10) Alor Malay

a. 'wake up (TRANSITIVE)'

Anak kecil satu ada tidur, dia punya bapa ni ada jalan
 child small one PROG sleep 3SG POSS father PROX PROG walk
datang ko kasi bangun dia.

come LNK give get.up 3SG

'A small child is sleeping, his father comes along and wakes him up.'

b. 'get up (TRANSITIVE)'

Simon kasi bangun Ata ko duduk.

S. give get.up A. LNK sit

'Simon lifts Ata up and then sits.'

A breakdown of the forms in Abui and Alor Malay are presented in Table 11.13.¹²

To conclude, Abui lexically differentiates verbs based on [\pm CHANGE OF POSTURE]. The verbs *-tein-* (transitive) and *-minang-* (intransitive) refer to a change of state event where an entity enters a waking state of consciousness with no change of posture. The verb *-rui-* (both transitive and intransitive) refers to a change of state event involving a change of posture. The verb *-rui-* is also both more frequent and more polysemous than the verbs *-tein-* and *-minang-*. Alor Malay is indeterminate to the feature and uses one verb *bangun* polysemously.

TABLE 11.13 Change of state verbs in Abui and Alor Malay

Sense	Language	Transitive	Intransitive
'wake up'	Abui	<i>ha-tein-</i>	<i>da-minang-</i>
	Alor Malay	<i>kasi bangun</i>	<i>bangun</i>
'get up'	Abui	<i>ha-rui-</i>	<i>da-rui-</i>
	Alor Malay	<i>kasi bangun</i>	<i>bangun</i>

4.4 *Summary: Differences between Abui and Alor Malay*

So far, we have seen the results presented for the proportion of mismatches across four age-groups, for the three verbal domains. In all three domains, generalization is clearly widespread, highlighting the loss of the features [CONTROL], [ELEVATION], [CHANGE OF POSTURE], respectively.

¹² The *ha-* inflection is used in transitive clauses to index a P argument, while the *da-* inflection is used in intransitive clauses to index an S argument.

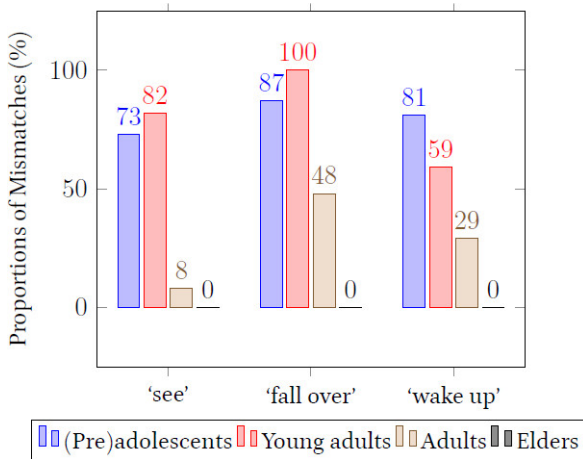


FIGURE 11.3 Proportion of mismatches for 'see', 'fall over', 'wake up'

Figure 11.3 visualizes the results of the preceding paragraphs, by using the mean percentages. Comparing the three contexts, it seems that 'fall over' shows the highest proportion of mismatches.

5 Discussion

This paper asked the question, "What can variation among age-groups in the use of the 'visual perception', 'falling', and 'change of state' verbs tell us about the semantic changes taking place in Abui"?

As predicted, there was variation in the use of the verbs among the four age-groups. The patterns observed in Light Abui for (pre)adolescents, young adults, and, to a lesser extent, adults, point to an increase in generalization. For all three domains, (pre)adolescents (9–16 years) and young adults (17–25 years) exhibited high percentages of generalization, while adults (26–34 years) exhibited generalization in one of the three domains. Elders (40–75 years), being the control group, consistently used the verbs in their appropriate contexts.

The variation across these age-groups points to an increase in frequency of the generalized forms rather than a categorical change. Nonetheless, this distribution suggests that the specific semantic distinctions encoded by the features [CONTROL] in events of visual perception, [ELEVATION] in events of falling, [CHANGE OF POSTURE] in events of change of state are gradually becoming irrelevant features in these three event domains. In this sense, the process of generalization is leading to the loss of these lexical features. Inter-

estingly, two of the forms being replaced, namely *-quoil-* ‘fall over’ and *-tein-* ‘wake (someone) up’ have cognates in other AP languages, such as Kamang, while additionally, *-tein-* ‘wake (someone) up’ is a reflex of *pTAP *tani* ‘wake’. The form *-quoil-* is likely to also be a reflex of a proto form given prevalence of cognates in other AP languages, such as Blagar, Kabola, Suboo, Reta (Kaiping, Edwards, and Klamer 2019), and Kamang (Schapper and Manimau 2011).

There was a clear pattern of which verbs were selected for generalization. Examining the question of why certain verbs were generalized and not others, it is likely that frequency and polysemy play a role. The generalized verbs were all more frequent than their polar counterpart and often also more polysemous, except in the visual perception event domain. Both frequency and polysemy are argued to be important lexical semantic factors that might determine the outcome of semantic change and additionally be extra sensitive in bilinguals. This is in line with Winter, Perlman, and Majid (2018) who found that words which are more frequent are often ‘re-used to express other concepts’ (p. 7). Frequency is also linked to polysemy: higher frequency words are more likely to be used in a variety of contexts, which will then lead to the acquisition of additional senses (Calude and Pagel 2011; Winter, Perlman, and Majid 2018; Zipf 1945).

As suspected, age proved to be a strong predictor of generalization. This is unsurprising, given that age is a defining feature of the transitional bilingualism found in the speech community. Age is linked to both *history* and *life-stage* which together have implications for exposure and language use (see Eckert (2017) for discussion of notions of *history* and *life-stage* in language variation and change). Specifically, *history* relates to early exposure to and use of Abui, while *life-stage* relates to current exposure to and use of Abui.

These two notions could help explain general differences between the three Light Abui groups. (Pre)adolescents and young adults behaved very similarly, while adults did not show as much generalization as the younger two groups. This is probably related to the history of input, as depicted in Figure 11.2. (Pre)adolescents and young adults had similar language acquisition history of both being raised predominantly in Alor Malay (see Figure 11.2). Present-day adults, however, were the first cohort of speakers whose parents transitioned from raising their children in Abui to raising them in Alor Malay. As shown in Figure 11.2, in the group of adults, 68% of speakers reported having received either a mix of the two languages or exclusively Abui as a child. This is (more than) double the amount reported by (pre)adolescents (22%) and young adults (34%), who were raised predominantly in Alor Malay; see also Saad (2020b, 121–125).

At the same time, it is important to address the role of *life-stage* in explaining why (pre)adolescents and young adults show very similar rates of generalization.¹³ It was discussed in section 2 that speakers of Light Abui exhibit delayed/adult vernacular production (Saad 2020b; Anderbeck 2015), meaning that speakers begin speaking Abui more actively during young adulthood (~17 years). This could be predicted to reverse the effects of generalization, as many studies show that an increase in exposure and proficiency during adulthood could allow an L2 learner to learn the distinctions and produce the appropriate verbs (Abutalebi 2008; Green 2003; Jarvis and Pavlenko 2008). However, based on their similar rates of generalization in this study, it has been shown that the life-stage of young adulthood does *not* reduce the rate of generalization. Indeed, this finding may be slightly at odds with other work, which provides considerable evidence that an increase in exposure and proficiency during adulthood could allow an L2 learner to learn these distinctions. This points to the late L2 learner being able to develop a lexico-semantic system with its own conceptual system and rely less on the L1 (Abutalebi 2008; Green 2003; Jarvis and Pavlenko 2008).¹⁴

It is also important to address why the group of adults showed signs of generalization at all, given that, among all the Light Abui groups, they had the most early exposure and current exposure to Abui. The fact that we observe this significant difference with the Traditional Abui of elders implies that the semantic changes documented in this paper, despite being most widespread and advanced among (pre)adolescents and young adults, probably originated in the group that is now adults. This suggests that the variation was likely already taking place around thirty years ago. At the same time, in opposition to the claim that this change may have originated thirty years ago, one can also not rule out the fact that this may have been a later change. If we assume this, then it may be possible that young adults (who generalize across the board) might have initiated this change, and then subsequently also influenced adults, despite being younger than them. This could be a possibility if we assume that the generalized forms are not necessarily stigmatized and that adults and young adults spend time together.

13 Note that it is only possible to assess the effect of life-stage on generalization between (pre)adolescents and young adults because they share similar history of exposure (22% and 34%); see Figure 11.2. It is more difficult to judge the effect of life-stage between young adults and adults because they have different histories of exposure (34% and 67%).

14 Proficiency was not directly tested in this study. However, adults' self-reports on their fluency of Abui score higher than those of young adults.

Two important questions arise, with the first pertaining to whether these innovations are contact-induced. These innovations are argued here to indeed be contact-induced based on the fact that the Traditional Abui-speaking elders do not generalize while the Light Abui-speaking groups of (pre)adolescents, young adults, and adults do generalize. In addition, the dominant language, Alor Malay, only uses one form to encode each of these three events. It has been shown in a number of previous studies that speakers whose L₁ uses a broad system, like Alor Malay, and who are learning an L₂ that uses a narrow system, like Abui, will have difficulties using the verbs correctly. They are thus likely to overgeneralize one of the forms. This was found both for bilingual speech communities and for second language learning contexts (e.g. Ameen et al. 2009; Gathercole and Moawad 2010; Pavlenko and Driagina 2008; Weinreich 1953; Backus, Seza Doğruöz, and Heine 2011).

A follow-up question is whether the contact phenomenon at hand is a case of simplification due to reduced input or due to transfer (in the form of lexical calquing) from Alor Malay into Abui.¹⁵ One of the difficulties in arguing for lexical calquing is that the types of semantic changes discussed here are also commonly attested in the absence of contact (e.g. Blank and Koch 1999; Campbell 2013; Traugott and Dasher 2001). Lexical calques are often easier to identify when they involve more rare combinations of words, corresponding to the donor language, as for example in the German word *fernsehen* ‘television’ (lit. ‘remote vision’) which is a literal translation of English *television* (Matras 2009). Nonetheless, at this point, one argument can be made in favor of lexical calquing in the domain of visual perception, where the form *-ien-* ‘see’ is being replaced by the form *(-)wahai* ‘look at’. It is cross-linguistically rare to have only one visual perception verb (Levinson and Majid 2014; Viberg 1983) so the fact that generalization is taking place could strongly suggest transfer from Alor Malay.¹⁶ Silva-Corvalán (1993) typically argues that simplification and overgeneralization involve internal tendencies but are accelerated by bilingualism. It is argued here that both lexical calquing (transfer) and reduced input are acting in a cumulative way to account for the patterns of generalization. To really tease the two apart, one would need to investigate verbs which involve a broad system in Abui and a narrow one in Alor Malay. In addition, one would also need to examine verbs that have the same level of specificity in Abui and Alor

15 Lexical calquing is defined here as ‘copying the polysemies of the model language into the recipient language’ and is considered a synonym of ‘loan translation’ (Ross 2013, 19).

16 It was not possible to get much information on whether creoles encode these distinctions. If many creoles do encode this distinction, this would strengthen the claim that generalization here is due to lexical calquing.

Malay, such as Abui *-buk* 'cradle (without cloth)' and *-wik* 'cradle (with cloth)' which correspond neatly to Alor Malay *koko* 'cradle (without cloth)' and *gendong* 'cradle (with cloth)'.

One important question that is unlikely to be addressed conclusively in the absence of a real-time longitudinal study is whether these innovations will lead to fully-fledged changes, as predicted by the apparent time construct. In other words, will the high rates of generalization found especially in the groups of (pre)adolescents and young adults persist with these individuals as they enter more senior life-stages and thus lead to language change? Speculating on the basis of the synchronic data, this does appear to be the case. It is predicted that the current group of (pre)adolescents will keep generalizing when they grow older and that this variation will indeed lead to change. Insights from another study and observations from the current data support this hypothesis. Firstly, Gathercole and Moawad (2010) found that words which conceptually contained very similar senses, applicable to the verbs in the three event domains (e.g. 'fall from above' vs. 'fall over'), had a much higher chance of being generalized than verbs which were conceptually more different to one another. This predicts that, at least for the three event domains described, generalization is likely to persist. In addition, the current cohort of young adults produced a high proportion of mismatches in all three domains, showing their high tendency to generalize. They did so having had similar levels of input compared to (pre)adolescents and also showing no decrease in their rate of generalization. In addition, (pre)adolescents will continue receiving input from their adjacent older age-group which favors the generalized forms. Finally, even the age-cohort above young adults, *adults*, produced enough mismatches to show evidence that they also generalize in one of the domains (falling). This shows that some of the innovations described here are so far advanced that they even occur in the speech of a group that has had higher levels of exposure to Abui than the current group of (pre)adolescents may ever have. Taken together, all of this predicts that when the current group of (pre)adolescents enters young adulthood and adulthood, they will continue to generalize.

6 Summary and Conclusion

This study investigated the distribution, causes, and implications of lexical variation in three event domains. Much of the variation was explained by age, and thus also by exposure to Abui. Traditional Abui, spoken by the group of elders, was used as the baseline variety, since Abui is the considered the L1 of this group, having only learned Alor Malay after the age of 7. With regards to

Light Abui, the rate of generalization was highest among (pre)adolescents and young adults. Some generalization was also found in adults, hinting that this is the group in which these innovations first appeared. I argue that the three verbs (-)wahai 'look at', hayeei 'fall from above', rui 'get up' which originally only referred to those specific senses, are becoming the generic verbs for 'visually perceive', 'fall', and 'get up/ wake up' and that the specific verbs -ien- 'see', -quoil/-kaai- 'fall over', -minang- 'wake up' and to a lesser extent -tein- 'wake (s/o) up' might become obsolete. If this variation leads to semantic change, then there is sufficient evidence that L1 transfer from Alor Malay into Abui has also taken place.

There are several exciting avenues for further research. The first one would include a follow-up panel study in eight-years-time, when members of the current age-groups would have advanced to the adjacent age-group. This would allow for a more robust testing of age-grading vs. apparent time, offering a more conclusive answer to the question of whether the current variation will lead to change. In addition, future work can focus on other verbs that appear to be undergoing generalization, such as the perception verbs 'hear' and 'listen'. Moreover, it could be worthwhile to tease out the effect of transfer from Alor Malay by looking at translation pairs that are congruent across languages such as Ab. -buk vs. AM. koko 'cradle (with cloth)' and Ab. -wik vs. AM. gendong 'embrace (without cloth)' in addition to looking at pairs which are 'broad' in Abui and 'narrow' in Alor Malay. This can determine to what extent direct transfer is taking place. Finally, future work can also try to extrapolate the findings of this speech community to speech communities of closely related Alor-Pantar languages to address the topic of how small-scale variation can lead to linguistic diversity.

Acknowledgements

I would like to acknowledge the unwavering support and insight provided by the Abui community of Takalelang and in particular Dori Lanma, Simon Daniel Lanma, Ansel Delpada, Benny Delpada, Frengki Delpada, Lorens Malbiyeti, Vivi Maufani, and Abner Yetimauh. In addition, I would like to express my words of gratitude to František Kratochvíl, Luís Miguel Rojas-Berscia, Kate Bellamy, Maria Carmen Parafita Couto, Pieter Muysken, Ad Backus, Marian Klamer, and Francesca Moro for offering invaluable feedback for the chapter. Data collection and write-up were supported by the NWO grant, 277-70-012 'Reconstructing the past through the languages of the present: The Lesser Sundas', led by Marian Klamer at Leiden University. Preparation for public-

ation was supported by ‘Sinophone borderlands—interaction at the edges’, CZ.02.1.01/0.0/0.0/16_019/0000791, awarded to Ondřej Kučera and ‘Verb class analysis accelerator for low-resource languages—RoboCorp’, 20–18407S, awarded to František Kratochvíl, both of which were hosted at Palacký University Olomouc.

Bibliography

- Abutalebi, Jubin. 2008. “Neural Aspects of Second Language Representation and Language Control.” *Acta Psychologica* 128 (3): 466–478.
- Aikhenvald, Alexandra Y. 2020. “Language Contact and Endangered Languages.” In *The Oxford Handbook of Language Contact*, edited by Anthony P. Grant. Oxford: Oxford University Press: Oxford. <https://doi.org/10.1093/oxfordhb/9780199945092.013.11>.
- Alais, David, and David Burr. 2004. “The Ventriloquist Effect: Results from near-Optimal Bimodal Integration.” *Current Biology* 14 (3): 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>.
- Ameel, Eef, Barbara C. Malt, Gert Storms, and Fons Van Assche. 2009. “Semantic Convergence in the Bilingual Lexicon.” *Journal of Memory and Language* 60 (2): 270–290.
- Anderbeck, Karl. 2015. “Portraits of Language Vitality in the Languages of Indonesia.” In *Language Documentation and Cultural Practices in the Austronesian World: Papers from ICAL 12 (A-PL 019)*, edited by I Wayan Arka, L.N. Seri Malini, and I.A.M. Puspani, 19–47. Canberra: Pacific Linguistics.
- Backus, Ad, A. Seza Dođruöz, and Bernd Heine. 2011. “Salient Stages in Contact-Induced Grammatical Change: Evidence from Synchronic vs. Diachronic Contact Situations.” *Language Sciences* 33 (5): 738–752. <https://doi.org/10.1016/j.langsci.2011.02.004>.
- Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand. 1991. “The Apparent Time Construct.” *Language Variation and Change* 3 (3): 241–264.
- Baird, Louise, Marian Klamer, and František Kratochvíl. in prep. “Alor Malay: A Grammar Sketch of an Emerging Eastern Variety.” Manuscript. Leiden University.
- Blank, Andreas. 1999. “Why Do New Meanings Occur? A Cognitive Typology of the Motivations for Lexical Semantic Change.” In *Historical Semantics and Cognition*, edited by Andreas Blank and Peter Koch, 61–89. Berlin: Mouton de Gruyter.
- Blank, Andreas, and Peter Koch. 1999. *Historical Semantics and Cognition*. Berlin: Mouton de Gruyter.
- Calude, Andreea S., and Mark Pagel. 2011. “How Do We Use Language? Shared Patterns in the Frequency of Word Use across 17 World Languages.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1567): 1101–1107. <https://doi.org/10.1098/rstb.2010.0315>.
- Campbell, Lyle. 2013. *Historical Linguistics: An Introduction*. 3rd edition. Edinburgh: Edinburgh University Press.

- Du Bois, Cora. 1944. *The People of Alor: A Social-Psychological Study of an East Indian Island*. Minneapolis: The University of Minnesota Press.
- Eckert, Penelope. 2017. "Age as a Sociolinguistic Variable." In *The Handbook of Sociolinguistics*, edited by Florian Coulmas, 151–167. Oxford: Blackwell Publishing Ltd.
- Fedden, Sebastian, and Dunstan Brown. 2017. "Participant Marking: Corpus Study and Video Elicitation." In *History and Typology*, edited by Marian Klamer, 405–448. Berlin: Language Science Press.
- Fedden, Sebastian, Dunstan Brown, and Greville Corbett. 2010. "Conditions on Pronominal Marking: A Set of 42 Video Stimuli for Field Elicitation. University of Surrey." <http://dx.doi.org/10.15126/SMG.25/1>.
- Fedden, Sebastian, Dunstan Brown, František Kratochvíl, Laura C. Robinson, and Antoinette Schapper. 2014. "Variation in Pronominal Indexing: Lexical Stipulation vs. Referential Properties in Alor-Pantar Languages." *Studies in Language* 38 (1): 44–79. <https://doi.org/10.1075/sl.38.1.02fed>.
- Gathercole, Virginia C. Mueller, and Ruba Abdelmatloub Moawad. 2010. "Semantic Interaction in Early and Late Bilinguals: All Words Are Not Created Equally." *Bilingualism: Language and Cognition* 13 (4): 385–408.
- Green, David W. 2003. "The Neural Basis of the Lexicon and the Grammar in L2 Acquisition." In *The Interface between Syntax and the Lexicon in Second Language Acquisition*, edited by Roeland van Hout, Aafke Hulk, Folkert Kuiken, and Richard Towel. Amsterdam: John Benjamins.
- Jarvis, Scott, and Aneta Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.
- Kaiping, Gereon, Owen Edwards, and Marian Klamer. 2019. "LexiRumah 2.1.1. Leiden: Leiden University Centre for Linguistics." 2019. <https://lexirumah.model-ling.eu/lexirumah/>.
- Kratochvíl, František. 2007. "A Grammar of Abui: A Papuan Language of Alor." PhD dissertation, Leiden: Leiden University.
- Kratochvíl, František. 2011. "Transitivity in Abui." *Studies in Language* 35 (3): 588–635.
- Kratochvíl, František. 2014. "Differential Argument Realization in Abui." *Linguistics* 52 (2): 543–602. <https://doi.org/10.1515/ling-2013-0072>.
- Kratochvíl, František, and Benidiktus Delpada. 2008. *Abui-Indonesian-English Dictionary*. 1st edition. Kupang: Unit Bahasa dan Budaya.
- Levinson, Stephen C., and Asifa Majid. 2014. "Differential Ineffability and the Senses." *Mind & Language* 29 (4): 407–427.
- Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.
- Nicolspeyer, Martha Margaretha. 1940. "De Sociale Structuur van Een Aloreesche Bevolkingsgroep." PhD dissertation, Leiden: Rijksuniversiteit te Leiden.
- O'Shannessy, Carmel. 2005. "Light Warlpiri: A New Language." *Australian Journal of Linguistics* 25 (1): 31–57.

- Paauw, Scott H. 2008. "The Malay Contact Varieties of Eastern Indonesia: A Typological Comparison." PhD dissertation, Buffalo: State University of New York.
- Pavlenko, Aneta, and V. Driagina. 2008. "Advancing in Russian through Narration." Pennsylvania State University: Calper Publications.
- Peddie, Emma. 2021. "Delayed Vernacular Production in Melanesia: A Case Study-Based Analysis." MA (Hons.) thesis, Edinburgh: University of Edinburgh.
- Polinsky, Maria. 2008. "Without Aspect." In *Case and Grammatical Relations: Studies in Honor of Bernard Comrie*, edited by Greville Corbett and Michael Noonan. Amsterdam/Philadelphia: John Benjamins Publishing.
- Ross, Malcolm. 2013. "Diagnosing Contact Processes from Their Outcomes: The Importance of Life Stages." *Journal of Language Contact* 6 (1): 5–47. <https://doi.org/10.1163/19552629-006001002>.
- Saad, George. 2020a. "Abui." In *The Papuan Languages of Timor, Alor and Pantar: Volume 3: Sketch Grammars*, edited by Antoinette Schapper, 267–346. Boston/Berlin: Mouton De Gruyter.
- Saad, George. 2020b. "Variation and Change in Abui: The Impact of Alor Malay on an Indigenous Language of Indonesia." PhD dissertation, Leiden: Leiden University.
- Schapper, Antoinette, and Marten Manimau. 2011. *Kamus Pengantar Bahasa Kamang-Indonesia-Inggris [Introductory Kamang—Indonesian—English Dictionary]*. 1st edition. Kupang, Indonesia.
- Silva-Corvalán, Carmen. 1993. "On the Permeability of Grammars: Evidence from Spanish and English Contact." In *Selected Papers from the 21st Symposium on Romance Languages*, edited by William J. Ashby, Marianne Mithun, Giorgio Perissinotto, and Eduardo Raposo E., 19–43. Amsterdam: John Benjamins.
- Stokes, Dustin, and Stephen Biggs. 2014. "The Dominance of the Visual." In *Perception and Its Modalities*, edited by Dustin Stokes and M. Matthen, 350–378. Oxford: Oxford University Press.
- Stokhof, Willem Arnoldus Laurens. 1984. "Annotations to a Text in the Abui Language (Alor)." *Bijdragen Tot de Taal-, Land- En Volkenkunde* 140: 106–162.
- Traugott, Elizabeth Closs, and Richard B. Dasher. 2001. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Viberg, Åke. 1983. "The Verbs of Perception: A Typological Study." *Linguistics* 21 (1): 123–162. <https://doi.org/10.1515/ling.1983.21.1.123>.
- Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.
- Winter, Bodo, Marcus Perlman, and Asifa Majid. 2018. "Vision Dominates in Perceptual Language: English Sensory Vocabulary Is Optimized for Usage." *Cognition* 179: 213–220.
- Zipf, George Kingsley. 1945. "The Repetition of Words, Time-Perspective, and Semantic Balance." *The Journal of General Psychology* 32 (1): 127–148.

Index

- Abui 8, 10, 15, 18, 79, 84, 87, 187, 225, 231, 234, 243, 253, 392–417
Abui Atimelang 239
Abui Fuimelang 86, 239
Abui Petleng 239
Abui Takalelang 66, 68, 70, 74, 86, 239, 242, 243, 251, 394
Abui Ulaga 73, 81, 239
Light Abui 393–396, 399–401, 403, 404, 407, 409, 412–415, 417
Traditional Abui 393, 394, 399, 400, 403, 404, 406, 407, 414–416
- Acehnese 39, 41–42, 50, 51
- Adang 79, 215, 223–225, 231–237, 248–251, 255–258
Adang Lawahing 66, 68, 70, 71, 73, 81, 83, 86, 87, 226, 229–232, 236, 239, 249
Adang Otvai 66, 68, 70, 71, 73, 81, 83, 86, 87, 89, 226, 229–236, 239, 249
- adaptation 338, 359, 363, 379–383
- Adonara 77, 80
- adstrate 309, 337
- adult
language shift 13–14, 394
second language learning 11
vernacular production 393, 396, 414
- affix(es)
adjectival 307, 314
borrowing 311, 312, 335
nominative 307
- affixation 314
- age 11, 393, 394, 397, 413, 416
-boundaries 396
-constructs 396
-grading 417
-group 392–394, 396, 397, 399, 411, 412, 416, 417
- agentivity of the speaker 3, 372, 375, 386
recipient language (RL) 373, 378, 379–382
source language (SL) 373, 378, 380–382
- agglutinative 356, 365
- Ainbai 265, 286
- algorithm 214, 219–221
- Alor Malay (AM) 8, 14, 60, 392–403, 405–407, 409–411, 413, 415–418
- Alor-Pantar Strait 215, 217, 218, 224, 230, 232, 234, 238, 241, 245, 248, 251, 253, 255, 259
- Alor-Pantar 2, 158, 159, 163–167
- Alor 61, 76, 215–218, 231–257
East Alor 64, 66, 68, 70, 71, 74, 75, 77, 78, 79, 81–84, 86, 87, 89, 90–92
South Alor 64, 66, 68, 70, 73, 78, 79, 81, 86, 87, 89, 91, 92
- Alorese 7, 13–14, 59, 71, 81, 83, 213–259
Alorese Alor-Besar 71, 223, 224, 227, 228, 232, 234n5, 235–252, 254
Alorese Alor Kecil 231–249, 252, 254
Alorese Bana 228, 229, 236, 239, 243, 246, 249, 250, 252, 254, 255
Alorese Baranusa 226, 232, 233, 239, 241–243, 249, 252, 254
Alorese Beang Onong 239, 242, 243, 246, 249, 251, 252, 255
Alorese Buaya 234n5, 236, 239, 241, 243, 244, 246, 248, 251, 252, 254
Alorese Dulolong 231–252, 254
Alorese Helangdohi 82, 228, 229, 236, 239, 249, 251, 252, 254
Alorese Kayang 239, 242, 244, 246, 249, 251, 252, 255
Alorese Marica 71, 82
Alorese Munaseli 82, 227–252, 254
Alorese Pandai 82, 223–226, 230, 231, 236, 239–243, 246, 248–252, 254
Alorese Ternate 222, 234, 236, 239–252, 254
Alorese Wailawar 82, 236, 239, 241, 244, 246, 249, 251, 252, 254
- Amabi 104
- Amanab 268, 273, 276, 279
- Amanatun 104
- Amanuban 104, 119
- Amarasi 107–109
Kotos Amarasi 104, 107, 119, 125, 132
Ro'is Amarasi 104, 119
- Amfo'an 104
- analogy 310, 319, 324, 333, 363, 380
- ancestor 1, 159

- ancestry, unknown 6, 120, 135, 145, 159
- ancient
 loanwords 60, 64–70, 256
 times 15
- apparent-time 392, 393, 416, 417
- archaeological evidence 2, 43, 87
- Aru 64n7
- aspect
 habitual 355
 imperfective 355
 perfective 355
 progressive 355
- aspiration 28, 33n11
- associative 367–368
- Atoni 62n5, 102
- Austroasiatic 26
- Austronesian (AN) 1, 2, 5, 8, 101–106,
 108–110, 111, 113, 117, 120, 122, 123–
 126, 132–134, 136, 180–182, 185–186,
 188, 191–195, 197–200, 203–206, 348,
 351
 origin 2, 148n1, 151, 159, 168, 172, 174, 175,
 249
 family 181, 314
 lexicon 2, 121–124
- automatic lexical similarity detection 214,
 219, 255, 256
- Auyi 268, 273, 276
- Ba'a 103, 105, 119
- Babar islands 64n7, 184
- Babar languages 193
- Babuyan Claro 349, 351–354, 369, 373, 374,
 377, 378–383
- back-formation 37n13
- Baikeno 104
- bangles 75
- bark cloth 75, 87
- barter trade 68, 74
- Barupu 269, 295
- Batanic 349, 351, 353, 358–359
- Batuley 64n
- Baumata 104
- beeswax 77
- Bengali (Be.) 25, 42, 44, 50
- Berik 295
- Biak 10, 46–47
- Biboki 104
- Bilbaa 103, 119
- bilingual(s) 10–13, 94, 172, 308, 318, 329, 333,
 337, 351–353, 373–375, 382, 413
 children 218, 257
 dominant 13, 14, 352, 373, 375, 382
 early 351, 352, 373, 375, 381, 382
 late 352, 373–375, 381, 382
 speakers 15, 92, 94, 171, 381
 speech 175, 415
 speech communities 415
- bilingualism 11, 12, 14, 92–96, 142, 174, 184,
 257, 258, 294, 307, 308, 311, 312, 334, 335,
 351, 378, 380, 384, 385, 393, 415
- adult 294
- asymmetric 13, 14, 218, 258
- balanced or symmetrical 380, 384
- child 12
- English-Tagalog 312
- transitional 14, 413
- Bima-Lembata 142
- Binongko 79
- Blagar 66, 71, 186–187, 221–225, 228–250,
 253, 255, 258, 259
 Blagar Bakalang 223, 230–239, 243, 245,
 246, 249, 252, 253
 Blagar Bama 68, 69, 73, 81, 82, 86,
 87, 89, 221, 223, 230, 233–239, 245–
 249
 Blagar Kulijahi 68, 69, 73, 82, 83, 86, 87,
 225, 230, 232, 235–247, 252
 Blagar Manatang 68, 69, 73, 86, 89
 Blagar Nule 68, 69, 73, 82, 83, 86, 89,
 221–225, 228–232, 235–239, 243–249,
 252
 Blagar Pura 68, 69, 73, 82, 86, 221, 222,
 228–239, 243–248
 Blagar Tuntuli 68, 69, 73, 78, 81, 82, 86,
 87, 89, 223, 229, 230, 236–239, 243–246,
 250
 Blagar Warsalelang 68, 69, 73, 82, 86, 87,
 229, 230, 235–246
- body part terms 9, 80, 89, 123, 169, 200,
 201, 202, 204, 247, 254, 275, 276, 280,
 291
- Bokai 103, 119
- Bomberai 58, 77
- Bonfia 199
- Border 16, 18, 268, 278, 279, 282, 283, 293
- Borneo 43, 44, 46, 49, 76
- borrowed elements 11, 13, 14

- borrowing 101–102, 106–109, 113, 120–125,
132–134, 136, 174, 181–206
see also transfer
amount of 12–13
ancient 60
Austronesian to Papuan 185, 193, 197–
198
direct 307, 310, 311, 312, 335, 336
direction of 4, 5, 185–186, 191, 194, 198,
199, 205–206, 272, 274, 275
grammatical, structural, syntactic 60,
95, 171, 172, 213, 258
indirect 307, 310, 311, 312, 319, 335, 336,
337
irregular 88
level of 13–15
lexical 1, 3, 10, 11, 12, 13, 15, 16, 25, 26, 27,
37, 38, 41, 42, 60, 93, 181, 185, 206, 213,
214, 217–220, 233, 236, 240, 255, 256,
258, 307, 308, 310, 335, 373
morphological 307, 310, 332, 335
multiple borrowing events 200, 204, 205
nonce 62, 381
of derivational morphology 3
Papuan to Austronesian 185–186, 188–
189, 191, 199, 205
permissiveness toward 12
pre-modern 60, 101–102, 113, 120–125,
132–134, 136
resistant 61, 122–123, 222
re-borrowing 134, 199, 200, 206, 237
scale 10, 348n
separate borrowing events 192
source 181, 186–191, 193–194, 197–200,
203, 205
source language unknown, no longer
extant 181, 188–189, 192, 195, 205
timeframe of 27, 35–40
BP 6, 15, 16, 58, 59, 60, 64, 65, 66, 215
see also Before Present time
British 7
Brunei 76
Bugis 35, 50, 84
Bunak, Bunaq 59, 66, 67, 69, 73, 74, 76, 80,
83, 89, 91, 92, 94–96, 186–189
 Bunak Bobonaro 71, 76, 81
 Bunak Maliana 77, 78, 81, 83, 84
 Bunak Suai 71, 76, 81
Buton 76
captured 79
Castilian Spanish 307
causative 94, 225, 367–368
Central Flores languages 141, 143, 169
Central Lembata 10, 72, 172
Central Masela 193
Central Philippine 314
Central Timor 188–189
 languages 143
 subgroup 102n
Central Eastern Malayo-Polynesian 103
Certificate of Ancestral Domain Title 352
Cham 36–37
Chamic 26, 51
Chamorro 319n7
China 87
Chinese 13, 218, 360, 364
 mestizos 308, 334
clan lineage 77
clipping, of first syllable 28
closed-class items 11
coast(s) 7, 42, 43, 74, 75, 76, 77, 81, 82, 92,
215, 241, 266, 268, 287, 351, 394
coastal people, coastal populations 68, 75,
77, 213, 253
code-switching 13–14, 142, 172, 174, 175,
373
cognate set(s) 4, 66, 107, 124–126, 146, 147,
225, 230, 238, 252, 253, 270, 285, 289,
302–303
coinage 107–110, 120–121
see also ex-nihilo root creation
colonial
 language 15
 powers 6, 7, 15
 slave trade 76
 times 51, 66, 84, 218
combs 75
common descent 4
complexification 12, 94, 365
 morphological 372
compound 107–108
compounding 314
conditional ability 368
constraint
 language-internal 375–377
constructional
 calquing 11
 transfer 93

- contact-induced 415
 (language) change 1, 2, 3, 10, 92, 129–
 135, 171, 257, 334, 348, 353, 354, 358, 363,
 365, 371, 372, 375, 376, 377, 378, 381, 384,
 386
 outcomes 4, 348, 349, 353, 373, 375, 377,
 378
 similarities 1
- contact 1, 112, 123–126, 129–135, 392, 394,
 398, 415
 casual 10–11, 13–14, 58
 date 15
 intense, intensity of 11–12, 13–14, 94, 95,
 307, 378
 long-term 94, 95, 213
 models of 10
 multi-purpose 94
 pre-modern 57, 129
 process 14
 scenario, setting, situation 10, 12–13, 181,
 185, 191, 204–206, 335
 superficial 14, 16, 95, 96, 256
 zones 92
- contemporary times 16
- convergence 294, 296, 300, 372
- copying 3, 171
- Cordilleran 349n4
 see also Northern Luzon
- crop 85
- cross-linguistic influence 372–373, 375
- Dadu'a 65, 67, 69, 71, 76, 80, 87
- dataset 6, 214, 218, 219, 256, 312, 313, 318, 319,
 320, 321, 324, 328, 329, 333, 336, 337
 historical 317, 318, 320, 322, 325, 328, 329,
 333
 size 7, 9
 synchronic 7
 type 7
- dating 6
- Dawan 62n, 102
- daya* networks 351, 379–381
- debt bondage 76
- Deing 65, 221, 226, 230, 231, 235–239, 242,
 243, 247, 249, 255
- Dela-Oenale 104, 119, 125, 134
- Dela 103, 108n5
- deliberate language change 172
- Dengka 103–105, 119, 125, 132
- dental stop 28
- derivation 107–110
 adjectival 309, 321
 agentive 326–328
 lexical 313, 314, 323, 327, 337
 nominal 310, 312–317
 strategy 314, 315, 333
 Tagalog 307, 312, 316, 326, 338
 type 315
- derivational
 affix 11
 morphemes, integration of 18, 309
 morphemes, transmission of 18
 morphology 11, 13, 14, 107–110, 376
- Dhao 103, 112
- dialect 182, 184, 186, 196, 197, 198, 200, 213,
 214, 219–221
 dialectal differences 182, 196
 dispersal 16
- diffusion 2, 125, 129–131, 384n
 linguistic diffusion 204
- diminutive
 suffix(es) 313, 327, 328, 333, 336
- directional/goal 369
- distributive 355–358, 366, 368
- Diu 103
- Dobel 199
- domains, lexical 36–37
- dominant
 group (of speakers) 293
 language 3, 18, 372–373
 see also language dominance
- Dominican 85
- donor
 language 92, 95, 219, 220, 222, 224, 226–
 233, 236, 238, 240, 243, 244, 248, 255,
 259, 309, 311, 335, 336
 region 60, 90–92
- doublet 370
- Dravidian 7, 8, 25, 37n13
- Dumo 269
- durative 348–349, 355–360
- Dusur 269, 284, 295
- Dutch 7, 25, 41, 60, 77, 78, 199, 310, 310n4
 VOC 77
- dynamicity 369
- early language exposure 394, 395
- earrings 75

- East Alor 82
 East Asian 16
 East(ern) Timor 26, 48, 181–182, 192–193
 Edictor 146, 222, 225
 elision, of word-final vowels 29, 36, 46
 Elseng 290, 298
 Emplawas 193
 Ende 76, 112
 English 25, 26, 41, 108, 110, 134–135, 310n4,
 311, 312, 315, 319, 320–325, 327, 330, 331,
 338–342, 353n9, 358–360
 Erai 193
 etymology, etymologies 5, 108, 181, 185, 186,
 194, 198, 200, 205, 270, 282, 353, 364,
 373
 etymons, etyma 78, 180, 185–190, 197–200
 European colonizers 7
 event domain 392–394, 397–400, 403, 406,
 409, 412, 413, 416
 ex-nihilo root creation 107–110, 120–121
 see also coinage
 exogamy 217, 258
- Fataluku 59, 66, 70, 71, 77, 81, 89, 180, 182,
 186–187, 189–191, 193
 Fatule'u 104
 faulty etymologies 26–27
 Filipino 308, 312–314, 320, 349, 352–353,
 364, 381–382, 383n31
 see also Tagalog
 alphabet 326
 first language 94, 372, 380
 Flores-Lembata 174
 languages 140–142
 region 59, 88, 90–92, 217
 subgroups 141, 148
 Flores 76, 92, 112, 158, 159, 163–167, 217
 fluency 11
 focus 355
 see also voice
 foreigners 77
 Franciscan travel account 75
 Frata languages 190–191, 193
 free variation 364, 372
 function words 11, 376
- Galela 46–47
 Galolen 65, 69, 71, 72, 76, 80, 84, 184
 Gayo 50
- gender
 grammatical 316
 lexical 29
 marginal 309
 genealogical connection 58
 generalization 14–15, 392, 393, 397, 399,
 401–405, 409, 411–417
 geographical
 area 287
 distribution 26, 28, 41
 spread 6
 grammatical calquing 92
 grammaticalization 95
 Greater Central Philippine 349n3
 Gresi 273
- Habun 184
 Hamap 66, 226, 231, 232, 248
 Hamap Moru 73, 83, 87, 239
 Hawaiian 26
 Hawu 73n, 106, 110n12
 head
 -initial 95
 -final 95
 Helong 62, 63, 68, 106–107, 110n12, 117, 131–
 132
 heritage language 12
 Hewa 71
 High variety 15
 Hindustani 25, 31–32, 41–42, 44, 48, 50
 hinterland 76
 hispanisms 308n2, 309
 historical records 2
 Holocene 64
 homeland(s) 61, 298
 homophonous 200
 near- 371
 hybrid formation(s) 311–313, 317–324, 328–
 331, 333, 335, 337, 339–342, 354, 365–371
 Type 1 363–364
 Type 2 363–364
 hybridization 312, 325, 338
- I'saka 269, 271, 293, 295, 299
 Ibanag 359n, 360, 364
 Ibatan 9, 13, 18, 348–349
 Iberian 84
 Peninsula 307
 Idate 69, 76, 80, 82, 84, 85

- Iha 77
 Ilokano 9, 13, 18, 348–349
 Imonda 268, 273, 276, 278, 279, 282, 283,
 296, 298
 imposition 3, 4, 13–14, 378, 380–384
 structural 374–375
 inalienable 81n
 inchoative 355
 see also punctual
 India 75
 Indian English 110, 134–135
 Indo-Aryan 8, 25–26, 29–32, 36–39, 49–
 50
 Indo-European 2, 7
 Indonesia 393
 Eastern 393, 394
 Indonesian 15, 60, 80, 88, 215, 218–220, 394
 inheritance 270, 271
 inland people, inland communities 68, 77,
 92
 see also mountain people
 innovation(s), innovative 192–193, 197, 205,
 295, 299, 302
 Insana 104
 integration 375
 of derivational morphemes 18
 phonological 33–35, 46
 semantic 33–34
 strategy of 280
 intensive 368
 interference 3
 borrowing 378
 substratum 378
 through shift 378
 irregularity 189n4, 192, 195
 Island SE Asia 2, 15
 isomorphism 295
 Ivatan 349, 359–360
 Ivasay 349
 Isamorong 349

 Java 37, 43
 Javanese 6, 8, 13–14, 16, 25, 28–32, 35, 39–43,
 48–49

 Kabola 225, 226, 229, 231, 233, 239, 242, 248,
 251
 Kabola Monbang 66, 71, 73, 81, 83, 87,
 89, 230, 250

 Kaera 66, 68, 69, 71, 73, 78, 80, 81, 86, 87, 221,
 224, 227, 228, 233, 236–250, 255, 258
 Kaera Abangiwang 230
 Kafoa 70, 73, 78, 81, 83, 87, 187–188, 226, 231,
 240, 242, 243, 253
 Kairui 180, 182, 184, 187, 203–204
 Kamang 78, 79, 83, 84, 86, 89, 186–188, 231,
 240, 245
 Kamang Atoitaa 66, 68, 70, 74, 83, 87, 89
 Kambera 85, 112, 134
 Kannada 25, 32n10
 Karo Batak 41–42, 50
 Kaure (family) 295
 Kawaimina languages 8, 13–14, 16, 180–206
 Kedang 71, 72, 142, 150, 151, 160, 161, 169, 172,
 174
 Kei 77, 199
 Keka 103
 Kemak 63, 69, 70, 72, 76, 77n, 80, 82, 84, 87,
 89, 102n, 188, 197
 Kemtuik 273, 275, 278, 290
 Keo 112
 Ketun 104
 Khmer 36–37
 kidnapping 77
 Kilmeri 7, 13–14, 268, 270, 271, 273–287, 293,
 295, 296, 298, 299, 301–303
 kinship 9, 62
 Kiraman(g) 66, 68, 70, 73, 81, 86, 87, 240,
 250, 253
 Kisar 182, 193, 199
 Kisar-Luangic languages 192
 Klamu 64n, 65, 67, 69, 71, 85
 Klon 81, 87, 186–188, 225, 241, 243, 246, 255
 Klon Bring 66, 70, 73, 78, 83, 240
 Klon Hopter 68, 70, 71, 73, 79, 83, 86, 225,
 238, 241, 250
 Kolana 77
 Kopas 104
 Korbafo 103, 125
 Kuala Lumpur 7
 Kui 66, 68, 70, 77, 81, 86, 187, 238, 240, 242,
 243, 246, 250, 253
 Kui Labaing 73, 78, 81, 87, 230, 250
 Kula 66, 70, 71, 72, 86, 186, 234n4
 Kula Lantoka 68, 74, 79, 223, 224, 243,
 249
 Kusa-Manea 104, 119
 Kwomtari (family) 266

- labour power 76
- Lakalei 184
- Lamahala 77
- Lamaholot 8, 13, 16, 59, 71, 72, 80, 143, 144,
149, 150, 160, 161, 169, 174, 175, 216,
234n4, 246, 252
subgroups 140, 141, 162, 168
- Landu 103, 119
- language contact 140, 175, 375
see also contact
aggregates of 385
models of 10, 376, 378, 384
outcomes of 10, 171, 384
scenario 392
- language dominance 372–375, 384–385
pattern of 348
- language history 396, 397
- language locations 17
- language maintenance 378
- language mixing 172
see also mixed code
- language of interethnic communication
66
see also lingua franca
- language proficiency 372
- language shift 134–135, 174, 257, 375
- laod networks 351, 379–381
- Lelain 103
- Lelenuk 103
- Lembata 74n16, 85, 92
- lenition 29
- Lesser Sundas 58, 106, 117
- Leti 33–35, 193
- lexeme 180–181, 185–186, 189, 190–191, 193,
195, 197–198, 205–206
of unknown origin 5
set 4, 61, 146, 147, 161
similar 4, 146
- lexical
borrowing 1, 93, 101–102, 106–109, 113,
120–125, 132–134, 136, 180–181, 198, 206,
307, 308, 310, 335
see also lexical influence
calques 12, 92, 415
contrasts 1
differentiation 399
entwinement 16, 181, 206
feature 412
field(s) 273, 288
influence 89, 214, 256–258
see also lexical borrowing
item 393, 404, 405
stratum/strata 105–107, 116–117, 120–124
stratigraphy 2, 16
survey 61
transfer 180, 206
variation 416
- lexicon 93, 120–124, 348, 378, 399
of older speakers 393
of younger speakers 393
shared 4, 16, 117, 161, 191, 193–198, 203,
205–206
verbal 392
- LexiRumah 7, 59, 61, 146, 218
- life stage 11, 396, 397, 413, 414
- lineage 79
- lingua franca 15, 94, 184, 218, 257, 349, 381,
394
- linguistic
area 2, 141
diversity 417
documentation, documentary gap/mater-
ial 7, 145, 184, 191, 205–206
material 348, 375–376
- Liquiça 77
- loan(word)(s) 60, 61, 95, 181, 184, 197–198,
203, 205, 214, 218–258
ancient 64–70
co-exist(ence) 271, 277, 280–281, 283,
287–288
complex 311–313, 317–319, 322–324, 328–
330, 333–337, 354, 360, 364–365, 383
concept of 3, 272
dispersal 16
early 379n, 382
hispanized English 319
integration 1, 280
mutual 18
number of 283
pre-modern 70–89
proportion of 9, 353
relative age of 270
simplex 328, 333, 335, 337, 376, 383
Spanish complex 313, 317–319, 321, 324,
328, 330, 337
sporadic 61
spread of 1, 217, 223
translation 92, 415n

- locative 355, 368
 Lole 103–105, 119, 134–135
 Luang 193
 Luangic languages 193
 Luzon 349, 352–353
- Madagascar 26, 43–44, 50
 Maka languages 16, 181–206
 pre-Maka 189
 Makalero 59, 76, 180–199
 Makasae 16, 59, 66, 67, 70, 71, 74, 80, 81, 87,
 180–206
 Makassar 35, 49–50
 Makassarese 76, 84
 Makatian 103
 Malacca 7, 66
 Malagasy 28–29, 33n, 43–45
 Malay 6, 8, 10, 13–16, 25–28, 30–40, 41–50,
 60, 80, 82–83, 198–199, 214, 215, 218,
 222, 225, 227, 236, 251
 court 7
 empire 7
 language of trade 7
 literature 7
 Malayalam 25, 32n10, 39–40
 Malayo-Polynesian, MP 103, 113–117, 120–121,
 124, 126, 128, 133, 136, 351
 Malayo-Polynesian expansion 140, 6
 Malaysia 7
 Maluku 29, 49, 76, 158, 192, 193, 199
 Southwest 103, 106
 Mambae 63, 69, 85, 102, 188
 Central Mambae 65, 76
 North-Western Mambae 65, 76, 84, 89
 Southern Mambae 65, 72, 76, 80, 84
 Western Mambae 84
 Mandarin Chinese 12
 Manem 273, 276, 293
 Mangarai 76
 Maranao 49, 50
 marriage 79, 88, 95, 256, 257
 Mekwei 273
 Melanesia 11, 396
 mestizos
 Chinese 308, 334
 Spanish 308
 metathesis 189n4, 191, 199, 203n13
 Meto 62n, 102, 104, 109n, 125
 see also Uab Meto
- Mexican Spanish 307, 308n, 322, 327, 342
 Middle-Indo-Aryan 25, 28–32, 34–39, 46–
 48, 50
 Midiki 69, 87, 180, 182, 184, 187, 193, 200–
 201, 203, 204
 migration 18, 268, 298, 300
 millet 85
 Minangkabau 43–44, 50
 Mindanao 76
 minority language 12
 Miomafo 104
 mixed code 172, 174
 see also language mixing
 mixed lexicon 175
 Mlap 273, 275
 modern times 16
 Moklenic 26, 36–39
 Molo 104, 107, 119, 125, 132
 Moluccas 7, 66
 mood
 irrealis 355
 neutral 356
 realis 355
 morphological compartmentalization 383
 morphologically complex loanwords 48–49
 morphologically
 analyzable 6
 complex 4
 productive 377, 383
 simplex 4
 morphology
 derivational 14, 18, 107–110, 376, 383
 morphosyntax 354, 355
 mountain people 77
 MP *see* Malayo-Polynesian
 multilingualism 12, 184, 218, 293, 395
 egalitarian 380
 nature of 378, 382–383
 pattern of 363, 378, 380
- national language 7, 15, 314
 native 348
 affixal inventory 326, 331
 affixes 4, 354
 stems 4, 311, 312, 324, 337, 354
 Naueti 65, 67, 69, 71, 76, 80, 87, 89, 180–206
 pre-Naueti 195
 Nedebang 186–187, 215, 223, 225, 234, 238,
 240

- Negrito languages 2, 141
 Neolithic 64, 87
 neutralization of transfer 384
 New Guinea 2, 18, 46, 50, 57
 New Spain 307
 New-Indo-Aryan 25, 27–35, 44, 47–48, 50
 newcomers 79
 Nimboran 7, 16, 18, 268, 271, 273–280, 290, 293, 296, 298
 Nominal 357, 361–362
 non-Austronesian (non-AN) 5, 8, 9, 15, 101, 102, 110, 115, 118, 120, 124, 133–136, 140
 vocabulary 141, 159, 160, 168, 169, 170
 languages 142
 non-native 348
 affixes 4, 354
 stems 4, 354
 North Halmahera 26, 46
 North Maluku 44
 Northern Luzon 349, 358, 359n, 363n
 see also Cordilleran
 noun(s)
 agentive 316–318, 321, 324, 327, 331, 338
 common 310, 312, 314, 321, 325
 proper 321, 325
 numeral(s) 9, 11, 240
 quinary 92n
 systems 92n
 nursery forms 4, 109, 120
 Nusa Tenggara 49, 50

 Oecusse 77
 Oenale 103, 107, 125, 132
 Oepao 103, 119
 Oirata 70, 81, 182, 187, 189, 190, 193
 Old Cham 30n, 36, 51
 Old Javanese 25n, 28n, 29–34, 37–39, 42–45, 49–50
 Old Khmer 37, 44
 Old Mon 36, 38–39, 51
 Old Sundanese 40
 onomatopoeia 109, 120
 oral
 source 298
 tradition(s) 74, 263, 268
 origin, unknown 6, 120, 135, 145, 159
 orphans 79
 Otomí 308

 Pagi 274, 276, 279, 281, 283, 290, 293, 299, 301–303
 Pali 25, 33, 36, 51
 Pantar-Straits 92
 Pantar 61, 85, 214–218, 224–243, 248, 251–259, 395
 Papua
 West 58
 Southwest 77
 Papuan 1, 57, 394
 languages of Timor *see* Timor Alor
 Pantar languages
 lexicon 2, 258
 Papuna 66, 70, 73, 78, 86
 paradigm 348
 paragoge 198
 parallel lexemes 172, 173
 see also parallel vocabularies, parallel lexicons
 parallel lexicons 93
 see also parallel lexemes, parallel vocabularies
 Parallel System Borrowing 377, 386
 parallel paradigms 358, 360
 parallel vocabularies 145
 see also parallel lexemes, parallel lexicons
 PCEMP *see* Proto Central Eastern Malayo-Polynesian
 PCMP *see* Proto Central Malayo-Polynesian
 Peninsular Spanish 322
 Persian 26, 30
 PET *see* Proto Eastern Timor
 PFRATA *see* Proto Frata
 Philippine-type morphosyntax 355
 Philippines 1, 7, 26, 29, 33, 36, 39, 43, 50, 76, 84, 308, 309, 312, 314, 329n, 334, 340, 348–349
 phonemes
 new(er) 302
 old(er) 301, 302
 phonological
 adaptation 271, 272, 276
 assimilation 133–135
 innovation 27, 29
 system 110–119, 6
 transfer 12, 13, 93, 110–119, 133–135, 378
 Pigafetta, Antonio 66, 75

- PMAKA *see* Proto Maka
 political control 76
 Portugal 77
 Portuguese 7, 25, 29n, 41, 48, 60, 77
 possessor prefixes 81n
 pottery 74
 pre-colonial political economy 76
 pre-modern 6, 15, 16
 prefix 315, 327
 verbal 18, 225, 251, 348
 prefixation 314
 pretense 368
 Principle of Morphosyntactic Subsystem
 Integrity 377, 386
 PRM *see* Proto Rote-Meto
 PTAP *see* Proto Timor-Alor-Pantar
 process 368
 pronouns 11, 218, 295
 see also pronoun systems
 Proto Alor-Pantar, proto AP 59, 162, 240,
 245, 250
 Proto Austronesian, proto AN 59, 62, 113,
 124, 240
 Proto Bunak 59
 Proto Central Eastern Malayo-Polynesian,
 PCEMP 106, 115n, 116–117
 Proto Central Flores 93, 112
 Proto Central Malayo-Polynesian, PCMP
 106
 Proto Central Timor 62, 63
 Proto Eastern Timor, PET 59, 188, 190–193,
 205
 Proto Flores-Lembata 62, 141, 148, 151, 168,
 169, 174, 218
 proto form 4, 228, 238, 241, 245
 Proto Frata, PFRATA 191–192
 Proto Helong 106–107
 proto language 4, 234
 Proto Maka, PMAKA 190–195
 Proto Malayo-Polynesian, proto MP, PMP
 59, 62, 63, 101, 104–118, 120–121, 124–136,
 184–185, 191–192, 194–195, 197, 199–200,
 203–204, 356
 Proto Meto 105, 119
 Proto Nuclear Alor-Pantar 230
 Proto Nuclear Rote 105, 119
 Proto Oceanic 237
 Proto Rote-Meto, PRM 8, 9, 13, 62, 63, 101,
 102, 104–136, 170, 145, 170
 Proto Timor-Alor-Pantar, PTAP 59, 61, 63,
 186–191, 193, 200, 205, 240, 413
 Proto Timor-Babar 62, 63, 117
 Proto West Rote-Meto 104n, 119, 120n, 126,
 131–132
 Proto Western Malayo-Polynesian, PWMP
 199
 psycholinguistic mechanisms 372
 punctual 355
 see also inchoative
 purist 12
 Puyuma 43
 PWMP *see* proto Western Malayo-Polynesian

 Quechua 308

 Raklungu 204
 recipient language 3, 7, 14, 18, 220, 307n,
 310, 311, 335, 338, 372
 see also RL
 recipient/dative agreement/object 278, 279,
 296
 reciprocal 355, 365–366, 368
 reconstruction(s), reconstructibility 145,
 147, 150, 157, 185–186, 188n2, 191, 193
 reduplication 314, 315, 316
 reflex, reflexes, reflection 185–186, 188–195,
 200
 regional language(s) 7, 105–106, 110–113,
 184, 393
 relative chronology 298, 300
 relexification 16
 repetition 368
 restructuring 378
 Reta 186–187, 224–234, 240, 241, 246, 252, 253
 Reta Pura 66, 68, 69, 73, 78, 86, 89, 221,
 224, 244–251
 Reta Ternate 66, 68, 69, 73, 82, 89, 221,
 244, 247, 250
 retroflex stop 28
 Rikou 103, 107, 119, 125, 132
 ritual 84
 RL 3, 310, 311, 312, 335, 336, 372
 see also recipient language
 Rongga 112
 Rote-Meto 16, 101–110, 112, 117, 120–121, 125–
 126, 129, 131–133, 135–137, 141, 170, 169,
 170
 Rote 33–35, 50, 102–105, 108, 109n

- Sa'ani 182
 sailing proximity 92
 sandalwood 77
 Sanskrit (Skt) 8, 13–14, 16, 25–26, 28–38,
 41–51, 134n, 323, 328
 Sar 65, 67, 69, 71, 80, 86, 228, 234, 235, 237,
 258
 Sar Adiabang 230, 236, 238, 240, 241
 Sar Nule 230, 238, 240
 sarongs 75
 Sawila 66, 68, 70, 71, 72, 77, 78, 80, 86, 186–
 188, 224, 243, 249
 script
 Indic 25n, 50
 sea currents 92
 second language 93, 94, 95, 352, 363, 372,
 381, 415
 secondary transmission 26–28, 49
 Selaru 103
 Selice Romani 12
 Seluwasan 103
 semantic change, semantic shift 14, 18, 229,
 253, 392, 393, 413, 417
 semantic domain, semantic field 1, 12–14,
 15, 61, 95, 122–124, 169, 122–124, 214, 254,
 257
 agriculture and vegetation 12, 14, 15, 222,
 228, 256, 259
 animals 14, 15, 64–68, 123, 200, 202, 205,
 222, 233, 256, 257
 basic actions 14, 123, 222, 223, 257, 259
 body parts 14, 80–83, 123, 169, 200, 202–
 204, 222, 239, 247, 254
 clothing and grooming 12
 eating 291
 existence/posture 281, 291
 geography 14
 governance 1, 14
 house 12, 222
 humans 200, 203–204
 hunting 291, 299
 kinship 14, 200, 203–204, 222, 239, 254,
 291
 law 12, 14, 222
 marriage 15, 88–89, 256, 257
 material culture 291
 minerals and metals 14
 motion 14, 123, 222, 239
 mythology 14
 natural kinds, kind-referring terms 291
 nature 14
 numerals 14, 279, 241, 254
 plants 14, 169, 123, 200, 202–203, 205,
 256, 257
 religion 1, 12, 14
 royal titles 14
 social and political relations 12, 15, 123,
 222, 227
 societal structures 14, 15, 75–80, 257
 subsistence and trade 14, 15, 68–70, 83–
 87, 257
 technology 1, 14, 15, 222, 223, 257, 259
 textile technology 70–75, 257
 tools 14, 15, 123, 222, 257
 toponyms 14
 trade 1, 256, 257
 semantic shift 27, 33–35, 188, 189, 200, 219,
 243–244, 271, 274–276, 278, 284–286,
 290, 300, 392–393
 see also semantic change
 Sengi 276
 Sentani 16, 269, 271, 281–283, 293, 295–296,
 298
 serial verb construction 95
 servants 79
 shift 93, 94, 96, 394
 see also language shift
 shift-induced change(s) 11, 92
 shifted language 94
 sibilant 28
 Sika 83, 84, 93, 142, 150, 151, 160, 168, 169,
 172
 Sika-Hewa 84
 similarities 1
 similarity set 4, 146, 147
 simplex-complex pairs 316, 318, 323, 330,
 333, 335, 336
 simplification of morphology 94
 simplified morpho-syntax 12, 93
 Singapore 7
 Sinhala 25, 33
 Siraya 43
 Skou 16, 269, 284–287, 289, 294, 295, 296,
 298
 SL 3, 5, 358
 see also source language
 slave(s) 76, 77, 78
 debt slaves 79

- slave raiders 77
- slave trade 76
- small-scale speech communities 92
- Solomon Islands 2
- Solor 75, 77
- sorghum 85
- sound change(s) 124, 132, 148, 149, 168, 270, 273, 282, 289, 290, 298, 301, 302
 - irregular 126–129
- sound imitations 4
- sound correspondence(s) 6, 141, 160, 161, 168, 175, 118–119, 124–126, 274, 277, 281, 290, 298
- source
 - language 3, 7, 14, 16, 292, 311, 312, 337, 338, 342, 358
 - unwritten 6
 - word 4
 - lexeme(s) 273, 283
- South America 84
- South Dravidian 25–26, 32, 36–39, 41, 51
- South East Barito 43
- Southeast Asia 13–14, 76, 84, 87
- Spanish 7, 8, 18, 322
 - colonial rule 307
 - complex loanword 313, 317, 318, 319, 321, 324, 328, 330, 337
 - diminutive suffix 313, 327, 328
 - Mexican Spanish 307, 308n, 322, 327, 342
- split
 - into subgroups 160, 169, 170, 175
- Sri Wijaya empire 66
- staple food 85
- stative 369, 379n
- stem(s) 354
 - adjectival 314, 315, 317, 321, 323, 326, 331
 - native 311, 312, 325, 337
 - nominal 315, 316, 317, 319, 321, 324, 329, 330
 - verbal 315, 316, 317n, 321, 326
- strangers 79
- stratigraphic analysis 16
- structural compatibility 375
- structuredness 375
- structures
 - integrated 11
 - loose 11
- Suboo 66, 68, 70, 73, 74, 84, 86, 87, 89
- subsistence 84
- substrate 93, 115–119, 124, 133–135, 145, 159, 170
- suffix(es) 181, 189, 198, 200, 202–204, 206
 - adjectival 18, 316
 - agentive 18, 310, 313, 317, 322, 326, 336
 - hybridization of 338
 - noun-forming 307
 - Spanish diminutive 313, 327, 328
- Sulawesi 66, 76, 79, 49
- Sumatra 37, 41–43, 49–51
- Sumba 85, 106, 112, 134
- Sumo 288
- superlative 368
- Swadesh list 9, 95, 121–122, 214, 256
- syncretise 372
- syntactic
 - copying 12
 - restructuring 92
- syntax 92
- Tabla 269, 282, 290, 298
- taboo 74
- Tagalog 7, 8, 13, 15, 18, 28–31, 33n11, 43–45, 307–342, 349n
 - see also* Filipino
 - Leipzig Corpus 313
 - texts 313
- Taikat 268, 273, 276, 279, 281, 285, 293
- Taiwan 26, 43, 50
- Talae 103
- Tamil 8, 13, 25, 31–33, 38–42, 50
- Tanimbar 76, 77, 103
- Tao 349
 - see also* Yami
- target language 292
- Tausug 39–40, 49
- tax 77
- Teiwa 186–187, 215, 223–228, 231–245, 252, 258
 - Teiwa Adiabang 65, 71, 73
 - Teiwa Lebang 67, 69, 71, 80, 86, 230, 242
 - Teiwa Nule 67, 69, 71, 86
- Tela-Masbuar 193
- Termanu 103, 107–108, 110n, 119, 125, 132
- Ternate 7, 46–47
- tertiary transmission 27

- Tetun 29n7, 48, 60, 64n, 70, 72, 76, 79, 88,
110n, 184, 194, 197–200, 203
Tetun Dili 69, 76, 80, 82, 84, 85, 87
Tetun Suai 65, 67, 69, 74, 82, 85
- textile(s) 74
Indian textiles 75
patola 75
- Thai 36–39, 51
- Tidore 7, 66
- Tii 103–105, 119, 125, 132, 134
- Timaus 104
- Timor 2, 59, 61, 66, 75, 76, 79, 85, 88, 90,
101–102, 105–107, 110, 112, 117, 121–124,
126, 136, 158, 159, 162–167, 180–185, 190,
198n9, 200n11, 206
- Timor Leste, Timor-Leste 60, 180–182, 184
- Timor-Alor-Pantar (TAP) language(s) 3,
7, 8, 13–14, 120, 171, 175, 180–206, 394,
413
see also Timor-Alor-Pantar (TAP) family
- Timor-Alor-Pantar family 57, 143, 120, 213,
215, 394, 413
Eastern Timor 180, 182, 185, 188, 191–193,
197, 203
- Timor-Alor-Pantar 26, 32, 48
- Timor-Babar subgroup 62, 102–103, 106, 117,
137, 143, 182, 184
East Timor 77, 182, 184, 205
North Timor 82
- Timorese 62n, 102
- Tiyei 66, 68, 70, 73, 74, 83, 84, 86, 87, 89
- Toba Batak 28, 30–32, 42–43, 50
- Tokodede 63, 64n, 69, 72, 76, 80, 82, 83, 84,
87, 102n
- Tor (family) 276, 295
- trade 95, 218, 256, 257
trade language 7, 94
trade network 84
- traders 77, 287
- trajectories, of borrowing 32, 41–50
- Trans New Guinea family 58
- transfer 3, 4, 258, 307n, 331, 335, 336
see also borrowing
direct 376
direction of 275, 277
imposition 13–14, 373, 380
indirect 365, 376, 380
lexical 348, 354, 380, 382
neutralization of 384
old 289
structural 348
- Transition 378, 385
- Tubbe 67, 69, 80, 85, 87
- Tugun 193
- Tulu 25, 32n10
- typological
contrasts 1
fit 375, 378
similarity 11, 112, 386
- Uab Meto 62, 93, 102
see also Meto
- Uniformitarian Principle 385
- unknown origin, unknown ancestry 6, 120,
135, 145, 159
- variation 297, 392, 393, 412–414, 416, 417
stable 371
- variationist sociolinguistics 384n
- verbs
of change of state 393, 397, 398, 406–
411
of falling 393, 397, 398, 403–406
of visual perception 393, 397, 398, 399–
403
- vocabulary 93
basic 9, 11, 14, 121–122, 159, 169, 214, 254,
287, 309
core 9, 292
non-basic 10–11, 14, 122–124
- voice
see also focus
actor 355
benefactive 355n
circumstantial 355
locative 355
patient 355
undergoer 355
- voicing 28, 32–33
- Waima'a 65, 67, 69, 71, 72, 74, 76, 80, 82, 84,
180–206
- Wallacea 101, 106, 136–137
- Wanderwörter 14, 199, 287, 291, 298
- Waris 268, 275, 278, 279, 282, 285, 296, 298,
301–303
- wars 77
war prisoners 79

- weaving 75
 - cloth 75
 - technology 74
 - songket 75
- Welaun 63, 102n, 132n, 189, 197, 204
- Wersing 66, 70, 71, 78, 80, 83, 87, 186–187, 246, 247
 - Wersing Maritaing 68, 72, 86, 223, 243, 246, 250
 - Wersing Taramana 72, 244, 246, 250
- West Papua 58
- West Tarangan 10
- Western Pantar 187, 224, 226, 228, 229, 242, 243, 251, 255
- Western Pantar Lamma 224, 226, 227, 240, 241, 244
- Western Pantar Tubbe 221, 247–249
- Wetan 193
- Wetar 184
 - Wetar languages 192–193
- Word and Paradigm 365, 376n
- word formation
 - Tagalog 310
- word order 94, 171, 221, 222
- written
 - historical records 15
 - traditions 6
- Wutung 269, 284, 286, 294, 295, 296
- Yakan 33–35, 49
- Yami 349
 - see also* Tao