

Cambridge Language Surveys

The Languages of Mainland Southeast Asia

N. J. Enfield

THE LANGUAGES OF MAINLAND SOUTHEAST ASIA

Mainland Southeast Asia is one of the most complex cultural and linguistic areas in the world. This book provides a rich and comprehensive survey of the history and core systems and subsystems of the languages of the region. Drawing on his depth of expertise in mainland Southeast Asia, N. J. Enfield includes more than a thousand examples from over a hundred of the languages of Cambodia, southern China, Laos, Malaysia, Myanmar, Thailand, and Vietnam, and brings together data and analyses that have not previously been available in one place. Chapters cover the many ways in which these languages both resemble and differ from each other, and the diversity of the area's languages is highlighted by an emphasis on minority languages, which outnumber the national languages by nearly a hundred to one. The result is an authoritative treatment of an important and fascinating linguistic area.

N. J. ENFIELD is Professor of Linguistics at the University of Sydney. His publications include *Linguistic Epidemiology* (2003), 'Areal linguistics and mainland Southeast Asia' (*Annual Review of Anthropology* 2005), *A Grammar of Lao* (2007), *Dynamics of Human Diversity* (2011) and *Languages of Mainland Southeast Asia* (2015, co authored with Bernard Comrie). He is working on a reference grammar of Kri, a Vietic language spoken in Laos.

CAMBRIDGE LANGUAGE SURVEYS

General editors

- P. Austin (School of Oriental and African Studies, London)
B. Comrie (University of California, Santa Barbara)
R. Lass (University of Cape Town)
D. Lightfoot (Georgetown University)
K. Rice (University of Toronto)
I. Roberts (University of Cambridge)
S. Romaine (University of Oxford)

This series offers general accounts of the major language families of the world, with volumes organized either on a purely genetic basis or on a geographical basis, whichever yields the most convenient and intelligible grouping in each case. Each volume compares and contrasts the typological features of the languages it deals with. It also treats the relevant genetic relationships, historical development and sociolinguistic issues arising from their role and use in the world today. The books are intended for linguists from undergraduate level upwards, but no special knowledge of the languages under consideration is assumed. Volumes such as those on Australia and the Amazon Basin are also of wider relevance, as the future of the languages and their speakers raises important social and political issues.

Volumes already published include

- Australian Languages* R. M. W. Dixon
The Dravidian Languages Bhadriraju Krishnamurti
The Languages of the Andes Willem Adelaar with Pieter Muysken
The Slavic Languages Roland Sussex and Paul Cubberley
The Germanic Languages Wayne Harbert
Sign Languages Diane Brentari
The Afroasiatic Languages Zygmunt Frajzyngier and Erin Shay
The Languages of Mainland Southeast Asia N. J. Enfield
Turkic Lars Johanson

Forthcoming titles

- The Balkan Languages* Victor Friedman and Brian Joseph

THE LANGUAGES OF MAINLAND SOUTHEAST ASIA

N. J. ENFIELD

University of Sydney



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314 321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi 110025, India

79 Anson Road, #06 04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108476331

DOI: [10.1017/9781108605618](https://doi.org/10.1017/9781108605618)

© N. J. Enfield 2021

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2021

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Names: Enfield, N. J., 1966 author.

Title: The languages of mainland Southeast Asia / N. J. Enfield.

Description: New York : Cambridge University Press, 2021. | Series: Cambridge language surveys | Includes bibliographical references and index.

Identifiers: LCCN 2020027669 (print) | LCCN 2020027670 (ebook) | ISBN 9781108476331
(hardback) | ISBN 9781108605618 (ebook)

Subjects: LCSH: Southeast Asia Languages.

Classification: LCC PL3501 .E54 2021 (print) | LCC PL3501 (ebook) | DDC 409.59 dc23

LC record available at <https://lccn.loc.gov/2020027669>

LC ebook record available at <https://lccn.loc.gov/2020027670>

ISBN 978 1 108 47633 1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third party internet websites referred to in this publication
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

CONTENTS

<i>List of Figures</i>	page ix
<i>List of Maps</i>	xi
<i>List of Tables</i>	xii
<i>Preface</i>	xix
<i>Acknowledgements</i>	xxi
<i>Note on the Nature and Organization of the Book</i>	xxii
<i>List of Abbreviations</i>	xxiv

1 Context 1

1.1	The Mainland Southeast Asia Region	1
1.2	Mainland Southeast Asian Languages	3
1.3	Nomenclature, System Ontology, and Language Data Selection	4
1.4	History of MSEA	10
1.4.1	Prehistory	10
1.4.2	Hoabinhian Neolithic Tradition	12
1.4.3	Earliest Agriculture	13
1.4.4	Bronze and Iron Age States	15
1.4.5	Summary: Fall of First Millennium States, Rise of Second Millennium States	24
1.4.6	Away from States: Zomia	26
1.4.7	Dynamics of Ethnolinguistic Diversity in MSEA	28
1.5	On Modern Dominant Languages	32
1.6	Resources and Developments in MSEA Linguistics	35
1.6.1	Conferences and Publications	35
1.6.2	New Descriptive Work	36
1.6.3	New Methods	37

2 Language Histories and Classifications 38

2.1	Current Developments	40
2.1.1	Unprecedented Access to Linguistic Data	40

2.1.2	Unprecedented Access to Other Sources of Relevant Information	40
2.1.3	New Methods of Historical Linguistic Analysis	43
2.2	Austroasiatic	43
2.2.1	Austroasiatic Subgrouping	46
2.2.2	Munda Languages	47
2.2.3	Non Munda ('Mon Khmer') Subgrouping	49
2.3	Tai Kadai	51
2.3.1	Subgrouping of Tai Kadai Languages	53
2.3.2	The Tai Branch	59
2.4	The Chamic and Moklenic Branches of Austronesian	66
2.4.1	Chamic	69
2.4.2	Moklenic	70
2.4.3	Structural Change in Chamic and Moklenic	71
2.5	Hmong Mien	72
2.5.1	Comparative Historical Hmong Mien	73
2.5.2	Hmongic versus Mienic	74
2.6	Sino Tibetan	79
2.6.1	Sinitic	81
2.6.2	Tibeto Burman Subgroupings	85
2.6.3	Affixes in Historical Tibeto Burman	89
2.7	Proposed Connections among the Families	92
2.8	Problems and Challenges for Comparative Historical Research	94
2.9	Signed Languages	96
2.9.1	Home Sign Systems	97
2.9.2	Village Sign Systems	97
2.9.3	Alternate Sign Languages	99
2.9.4	National Level Sign Languages	100
2.9.5	Prospects	105
2.10	Classical Languages	106
2.11	Scripts and Writing Systems	107
3	Typological Overview	118
3.1	Preliminaries	118
3.2	Typological Overview	122
3.3	Morphological Profile	123
3.4	Syntax	128
3.5	Basic Phrasal Constituent Order Patterns	129
3.5.1	Relative Position of Verb and Object	131
3.5.2	Relative Position of Modifier and Head in Noun Phrases	132
3.5.3	Relative Position of Noun and Relative Clause	135

3.5.4	Relative Position of Possessor and Possessed	137
3.5.5	Relative Position of Adposition and Noun	138
3.5.6	Relative Position of Adjective and Standard of Comparison	139
3.6	Sentence Type Distinctions	141
3.7	Sentence Final Particles	144
3.8	Expressive Language	146
3.8.1	Ideophones/Expressives	146
3.8.2	Poetic Idioms (Four Syllable Expressions)	151
4	Phonological Systems	154
4.1	Consonants	154
4.2	Vowels	161
4.3	Phonotactics	166
4.4	Register	168
4.5	Tone	174
4.6	Tonogenesis	183
4.7	Tone Sandhi	195
4.8	Intonation	203
4.9	Prosodic Hierarchy and Phonological Words	204
5	Word Formation	209
5.1	Form Classes	210
5.1.1	Noun Verb Distinction Is Usually Clear	210
5.1.2	‘Adjectives’ Are Verbs, or Are Verb Like	211
5.1.3	‘Adpositions’ Are Often Nouns or Verbs	214
5.1.4	Verbs Can Act as ‘Adverbs’ without Derivation	216
5.2	Compounding	216
5.3	Psycho collocations	222
5.4	Reduplication	227
5.5	Affixation	242
5.5.1	Existential/Locative Marking and Impersonal Predicate	243
5.5.2	Lexical Category Derivation	243
5.5.3	Reciprocal/Collective Marking	246
5.5.4	Causative	247
5.5.5	Further Morphological Derivations	250
5.6	Tone in Word Formation	252
5.6.1	Tone Sandhi in Compound Formation	252
5.6.2	Regular Grammatical Distinctions	253
5.6.3	Forming Ideophones	255

6 Reference and Nominal Syntax	257
6.1 Nominal Modification	259
6.1.1 Simple Head Modifier Relations	259
6.1.2 Relative Clause Constructions	266
6.1.3 Possessive and Part Whole Constructions	277
6.2 Nominal Classification	282
6.2.1 Numeral Classifier Constructions	283
6.2.2 Class Terms	300
6.3 Personal Pronoun Systems	308
6.4 Demonstrative Systems	318
6.4.1 Two Term Systems	318
6.4.2 Three Term Systems	321
6.4.3 Four Term Systems	322
6.4.4 Five Term Systems	323
6.4.5 Larger Systems	324
6.4.6 Comment on Demonstrative Systems	325
7 Predication and Clausal Syntax	326
7.1 Marking Syntactic Functions	326
7.2 Verbal Marking	332
7.2.1 Negation	332
7.2.2 Aspect and Related Marking	335
7.2.3 Modality and Related Marking	340
7.3 Multi verb Constructions	342
7.3.1 Possible Number of Verbs in Series	345
7.3.2 Events Occurring in Series	346
7.3.3 Events or Event Facets Occurring as Elements of a Single Event	354
7.3.4 Complementation Strategies	362
7.4 Valency Changing Strategies	368
7.4.1 Participant Adding: Syntactic Causatives and Related Constructions	368
7.4.2 Single Participant in a Transitive Event: Reflexives and Related Constructions	375
7.4.3 Two Way or Distributed Action: Reciprocals and Related Constructions	384
<i>Postface</i>	393
<i>References</i>	396
<i>Author Index</i>	429
<i>Language Index</i>	431
<i>Subject Index</i>	435

FIGURES

- 1.1 Chronology, I: Mainland Southeast Asia during the last two Marine Isotope Stages (MIS 1, 2) *page 14*
- 1.2 ‘Schematic depiction of succession of dominant states’ 25
- 1.3 Chronology: Mainland Southeast Asia over the last 4,000 years 29
- 2.1 Mon Khmer classification according to Thomas and Headley 50
- 2.2 Austroasiatic (after Diffloth 2005: 79) 51
- 2.3 Austroasiatic (after Sidwell 2015: 179) 52
- 2.4 Tai Kadai languages 58
- 2.5 ‘Tentative tree diagram of Austronesian languages’ 68
- 2.6 Proposed Malayo Chamic grouping 69
- 2.7 ‘Malayo Sumbawan’ 70
- 2.8 Mienic subgrouping with four main subgroups 77
- 2.9 ‘Classification of Mjuenic dialects and sub dialects’ 78
- 2.10 Mienic, with three subgroups 78
- 2.11 Subclassification of Mienic languages 79
- 2.12 The ‘standard view’ of Sino Tibetan as consisting of two top level branches: Tibeto Burman and Sinitic 82
- 2.13 ‘Chinese “dialects” (languages)’ 84
- 2.14 ‘Schematic chart of Sino Tibetan groups’ 86
- 2.15 ‘The Tibeto Burman language family tree, reflecting the Sino Bodic and Western Tibeto Burman hypotheses’ 89
- 2.16 ‘Sign languages in Thailand classified by traditional language family trees’ 102
- 2.17 ‘Sign languages in Vietnam classified by a traditional language family tree’ 103
- 2.18 ‘Linguistic relations between sign languages in Thailand and Vietnam’ 104
- 2.19 Modern Burmese writing 110
- 2.20 Modern Lao writing 110
- 2.21 Modern Lao Song (Black Tai) writing 111

- 2.22 ‘Historical development of Indic scripts in Southeast Asia’ 113
4.1 Segmental structure of the Kri (AA) word 166
4.2 Hakha Lai (TB) tones 175
4.3 Modern Standard Chinese (SN) tones 176
4.4 Cantonese (SN) tones 177
4.5 Northern Vietnamese (AA) tones 178
4.6 White Hmong (HM) tones 178
4.7 Dong (TK) tones 179
4.8 Lao (TK) tones 180
4.9 The Xiamen (SN) tone circle 196
6.1 Accessibility hierarchy for the construction of relative clauses in
languages of the world 268

MAPS

1.1	Mainland Southeast Asia	<i>page</i> 2
1.2	‘Sundaland’	12
2.1	Approximate distribution of languages of the Austroasiatic family	44
2.2	Approximate distribution of Austroasiatic languages in Core MSEA	45
2.3	Approximate distribution of Tai Kadai languages	54
2.4	Approximate distribution of Austronesian languages	67
2.5	Approximate distribution of languages of the Chamic and Moklenic branches of Austronesian	68
2.6	Approximate distribution of Hmong Mien languages in MSEA	73
2.7	Approximate distribution of Sino Tibetan languages in MSEA	80
3.1	Order of adjective and noun in Tibeto Burman languages of northwestern MSEA and beyond	135

TABLES

- 1.1 Languages of MSEA: (a) numbers of languages in Core MSEA, by family; (b) numbers of languages in Greater MSEA, by family *page 4*
- 2.1 Lexical correspondences in Santali, Mon, and Muong 44
- 2.2 Proto Austroasiatic inventory of syllable initial consonants 46
- 2.3 Different patterns of merger of the Proto Austroasiatic three way initial bilabial plosive split, showing retention of implosive bilabial *ɓ survives in Old Mon (Monic), Bahnar (Bahnaric), and Katu (Katuic) 46
- 2.4 Different patterns of merger of the Proto Austroasiatic three way initial apical plosive split, showing retention of implosive apical *ɗ in Old Mon (Monic), Bahnar (Bahnaric), and Katu (Katuic) 47
- 2.5 Summary of Austroasiatic proto branch level initial consonant correspondences 48
- 2.6 Thirteen branches of Austroasiatic 53
- 2.7 Some basic vocabulary in Tai Kadai languages from distant branches 54
- 2.8 Lexical correspondences supporting the grouping of Kra languages (Gelao, Lachi, Laha, Paha Buyang, and Pubiao) as distinct from languages of other branches of Tai Kadai 56
- 2.9 Rice farming vocabulary, establishing a split between Gelao Hlai and the rest of Tai Kadai 57
- 2.10 Early Tai tone categories for ‘live’ syllables resulting from a historical split of three proto tones (A, B, and C) 61
- 2.11 Early Tai tone categories for ‘dead’ syllables, with long versus short vowels 62
- 2.12 Gedney’s complete system of twenty pan Tai tone boxes 63
- 2.13 Modern reflexes in Tai languages of syllables in boxes 1–3 under the A category in Proto Tai 65
- 2.14 Lexical evidence for a historical distinction between Hmongic and Mienic languages 75

2.15	Modern reflexes of Proto Hmong Mien tight (*NC) versus loose (*N C) prenasalized obstruents	75
2.16	Evidence for a correspondence between historical prefixes (also found in Tibeto Burman) and modern voiceless initials in Hmongic languages	76
2.17	'Sino Tibetan comparisons'	81
2.18	Selected core vocabulary of Chinese varieties	83
2.19	Proto Lolo Burmese initial simple plosives	90
2.20	Modern reflexes of Proto Lolo Burmese initial bilabial stops	90
2.21	Lexical correspondences illustrating modern reflexes of Proto Lolo Burmese initial bilabial stops	91
2.22	Modern reflexes of Proto Lolo Burmese voiced and voiceless initial bilabial stops, with and without the *s and *m prefixes	91
2.23	Lexical correspondences illustrating modern reflexes of Proto Lolo Burmese voiced and voiceless initial bilabial stops, with and without the *s and *m prefixes	92
2.24	Data on selected sign languages of core MSEA	100
2.25	'Neologisms and lexical specifications' based on Pali/Sanskrit in Thai	107
2.26	'Cham alphabet letters: Consonants'	109
2.27	Features common to cases of the independent development of orthography in 'Zomian' minority languages	115
3.1	Overview of sound system features in MSEA languages	122
3.2	Overview of morphosyntax semantics system features	123
3.3	Features of Cantonese sentence final particles (SFPs)	145
3.4	A selection of Sgaw Karen (TB) ideophones	147
3.5	Some ideophones in Chrau (Austroasiatic)	150
4.1	Simple syllable initial consonants in Khmu (AA)	156
4.2	Khmu (AA) initial clusters	156
4.3	Syllable final consonants in Khmu (AA)	157
4.4	Set of consonants possible in the initial position of main syllables in Kri (AA)	157
4.5	Final segments of stressed syllables in Kri (AA)	157
4.6	Syllable initial consonants in Lao (TK)	158
4.7	Syllable final consonants in Lao (TK)	158
4.8	Syllable initial consonants in Sgaw Karen (TB)	159
4.9	Syllable initial consonant clusters in Sgaw Karen (TB)	159
4.10	Syllable final consonants in Sgaw Karen (TB)	160
4.11	Set of consonants possible in the initial position of main syllables in Moklen	160
4.12	Set of consonants possible in the final position of main syllables in Moklen	160

- 4.13 Initials of Dànanshān Miao, Guizhou (HM) 161
 4.14 Cambodian (AA) vowel phonemes 162
 4.15 Dànanshān Miao (HM) vowel phonemes 162
 4.16 Cantonese (SN) vowel phonemes 162
 4.17 Shanghainese (SN) vowel phonemes 163
 4.18 Sgaw Karen (TB) vowel phonemes 163
 4.19 (Eastern) Khmu (AA) vowel phonemes 164
 4.20 Lao (TK) vowel phonemes 164
 4.21 Modern Western Cham (AN) vowel phonemes 164
 4.22 Moken (AN) vowels 165
 4.23 Bahasa Malaysia (AN) vowels 165
 4.24 Moklen (AN) vowel phonemes 166
 4.25 Some Kri (AA) (near) minimal pairs contrasting in register alone 169
 4.26 Long vowel phonemes of Kri (AA) showing vowels paired by register 170
 4.27 Long vowel phonemes of Kri (AA) laid out in a standardized nine place system 170
 4.28 Selection of Rengao (AA) long vowel phonemes and their realization in tense versus lax register 171
 4.29 Hypothetical scenario based on the Rengao situation shown in Table 4.28 172
 4.30 Vowel phonemes of Cambodian (AA) 172
 4.31 The relation between modern Khmer vowels (shown in Table 4.30) and the earlier two register system in Old Khmer 173
 4.32 Takhian Thong Chong registers 174
 4.33 Illustration of falling, rising and low tones on smooth syllables as realized after *ka* ‘my’ in Hakha Lai (TB) 176
 4.34 The three classes of syllable initial consonant in Lao 182
 4.35 Modern Lao (TK) examples of reflexes of traditional Tai tones 183
 4.36 Vietnamese (AA) and Chinese (SN) lexicon 184
 4.37 Eight traditional Chinese tone categories 185
 4.38 Traditional Vietnamese tone categories 186
 4.39 Stage 1 of tonogenesis, pre tonal 187
 4.40 Stage 2 of tonogenesis, pitch contours being phonologized 187
 4.41 Stage 2 of tonogenesis, showing syllable types distinguished by voicing of syllable initial segment 187
 4.42 Stage 3 of tonogenesis, initial contrasts lost (here, all Cs are voiceless), incidental pitch differences arising from previous ±voicing now make lexical contrast 188
 4.43 Comparison of Vietnamese tone category C words with Mon, Mnong, and Kri cognates (all AA) 188

- 4.44 Comparison of Vietnamese tone category C words with Khmer cognates (both AA) 189
- 4.45 Comparison of Vietnamese tone category B words with Khmu, Riang, and Kri cognates (all AA) 189
- 4.46 Comparison of Vietnamese tone category A words with Khmu cognates (both AA) 189
- 4.47 The six Vietnamese tones, plotted as a function of two non tonal parameters of syllables in the proto language 190
- 4.48 Haudricourt's model of tonogenesis applied to White Hmong 190
- 4.49 'Karen tone box' 191
- 4.50 Three values for terminance in Kri (AA) 192
- 4.51 The six Kri (AA) syllable types defined by intersections of terminance and register distinctions 192
- 4.52 The six Vietnamese tones plotted as a function of two non tonal parameters of syllables in the proto language 192
- 4.53 Associations between syllable type and pitch in Moklen (AN) 193
- 4.54 Moklen (AN) pitch based minimal pairs 194
- 4.55 Examples of Xiamen (SN) tone sandhi following the tone circle pattern 196
- 4.56 Pattern of tone replacement in progressive tone sandhi in White Hmong 197
- 4.57 Biao Min Yao (HM) tones 198
- 4.58 Hakha Lai (TB) tones, with values represented by start and end points 199
- 4.59 Five combinations of tone in Hakha Lai (TB) that do not undergo tone sandhi change 200
- 4.60 The four combinations of tone in Hakha Lai (TB) that undergo tone sandhi change 200
- 4.61 Wuming Zhuang (TK) tones 201
- 4.62 Two patterns of tone alternation in Wuming Zhuang (TK) 201
- 4.63 Wuming Zhuang (TK) tones re described in terms of register (first letter) and contour within that register 202
- 4.64 Two patterns of tone alternation in Wuming Zhuang (TK) re stated using a register/contour description of tone 202
- 4.65 Elements of the phonological/prosodic hierarchy 205
- 4.66 Kinds of unstressed syllables in Lao (TK) distinguished by their identity on the implicational scale minor > light > unstressed 206
- 4.67 Comparison of phonological and other grammatical features of Munda and Mon Khmer (non Munda) Austroasiatic languages 208
- 5.1 Kratochvíl's five way distinction in 'basic types of morphological construction' in Mandarin Chinese (SN) 209

- 5.2 Example psycho collocations in Bangladesh Khumi (Kuki Chin, TB) 223
- 5.3 Example psycho collocations in Pwo Karen (TB) 223
- 5.4 Example psycho collocations in Yongning Na (Mosuo) (TB) 224
- 5.5 Example psycho collocations in Southern Min (SN) 224
- 5.6 Example psycho collocations in Wa (Paraok) (AA) 224
- 5.7 Example psycho collocations in Vietnamese (AA) featuring reference to *lòng* ‘entrails, bosom’ 225
- 5.8 Example psycho collocations in Thai (TK) featuring reference to *cay* ‘heart’ 225
- 5.9 Example psycho collocations in Hmong (Mong Leng, Hmong Mien) mostly featuring reference to the liver 226
- 5.10 Example psycho collocations in Colloquial Eastern Cham (AN) featuring reference to the liver, body, head, and belly 226
- 5.11 Example psycho collocations in Mon (AA) 227
- 5.12 Example psycho collocations in Khmer (AA) 227
- 5.13 Example psycho collocations in Burmese (TB) 227
- 5.14 Four types of chameleon affix in Vietnamese (AA) 237
- 5.15 Expressive reduplication in Lao (TK) with stem alternation, back vowel changed to front vowel 241
- 5.16 Expressive reduplication in Lao (TK), front vowel changed to schwa 242
- 5.17 Doublets in White Hmong (HM) arising from Sandhi effects 252
- 6.1 Dominant properties of relativization in non Sinitic MSEA languages 267
- 6.2 Type I classifier constructions in MSEA and beyond 285
- 6.3 Type II classifier constructions in MSEA and beyond 285
- 6.4 Khmu (AA) classifiers 286
- 6.5 Mandarin classifiers: Categories and inventory sizes 291
- 6.6 Class nouns in White Hmong (Hmong Mien) 301
- 6.7 Some variable features of personal pronouns across a sample of MSEA languages 307
- 6.8 Personal pronouns in Cantonese (SN) 308
- 6.9 Personal pronouns in Mandarin (SN) 309
- 6.10 Personal pronouns in Hakka (SN) 309
- 6.11 Personal pronouns in Lahu (TB) 310
- 6.12 Personal pronouns in Xuyong Hmong (HM) 310
- 6.13 Personal pronouns in Green Miao (HM) 310
- 6.14 Personal pronouns in Mulao (TK) 311
- 6.15 Personal pronouns in Chadong (TK) 311
- 6.16 Personal pronouns in Chrau (AA) 312
- 6.17 Personal pronouns in Phan Rang Cham (AN) 312

6.18	Personal pronouns in Zhuang (TK)	312
6.19	Personal pronouns in Jahai (AA)	313
6.20	Personal pronouns in Kri (AA)	314
6.21	Some kin categories that are consequential for selection of person reference form in Kri (AA)	314
6.22	Kri (AA) basic kin terms	315
6.23	Khmer (AA) personal reference paradigm	316
6.24	Burmese (TB) personal reference paradigm	317
6.25	Cantonese (SN) demonstrative pronouns	318
6.26	Mulao (TK) demonstrative pronouns	319
6.27	Zhuang (TK) demonstrative pronouns	319
6.28	Hainan Cham (AN) demonstrative pronouns	319
6.29	Phan Rang Cham (AN) demonstrative pronouns	319
6.30	White Hmong (HM) demonstrative pronouns	320
6.31	Lao (TK) demonstrative determiners	320
6.32	Lao (TK) demonstrative adverbs	320
6.33	Semelai (AA) basic demonstratives	321
6.34	Sedang (AA) demonstratives	321
6.35	Hakka (SN) demonstratives	321
6.36	Aizhai Miao (HM) demonstratives	322
6.37	Green Miao (HM) demonstratives	322
6.38	Chrau (AA) demonstratives	322
6.39	Akha (TB) demonstratives	322
6.40	Lai (TB) demonstratives	323
6.41	Lahu (TB) demonstratives	323
6.42	Kri (AA) demonstratives	323
6.43	Jahai (AA) demonstratives	324
6.44	The semantics and pragmatics of Jahai (AA) demonstrative categories	324
7.1	Some idealized grammatical categories of verbal marking that ground a proposition by linking (1) the ‘narrated event’ to (2) the coordinates of the speech event	331
7.2	Cantonese (SN) path of motion verbs	355
7.3	Cantonese (SN) directional SVCs	355
7.4	Elements of the Lao (TK) Manner Path Direction construction	356
7.5	Semelai (AA) associated motion multi verb constructions	358

PREFACE

To understand the nature of human language, we need to know the properties that languages can have, the distribution of those properties in the world, and the reasons for that distribution. This book brings together information relevant to these questions in relation to one geographical area: mainland Southeast Asia (MSEA). In line with the remit of the Cambridge Language Surveys series, we present a technical survey of the languages of this area, with two main points of thematic focus: first, the histories of the languages and their speakers, and second, the structural properties of the languages from sound systems to the make up of phrases and clauses.

There is of course much more to language: verbal art and oratory, ritual speech, multilingualism, dialectology, politeness and impoliteness, discourse structure, toponymy, personal names, kinship terminologies, ethnobiological classification, narrative style and practice, language ideology, et cetera. These areas await a more sociologically and anthropologically oriented volume, tapping into a wealth of past and current research in these domains. For example, on the sociolinguistics of language endangerment, and associated issues including language protection and revitalization, see Phattharathanit (2012) on identity maintenance in Lanna (cf. Bradley 2007, Premsrirat 2007). Research on linguistic politeness is being done, mostly in relation to national languages, and with reference to the languages' elaborated systems of social deixis, for example in their systems of personal pronouns, and the pragmatic alternatives that effectively create open class systems for person reference (see Cooke 1968, Haas 1969, Luong 1990). The more complex documented systems of person reference belong to the major literate languages of the area, including Thai, Cambodian, Vietnamese, and Burmese (Cooke 1968). There has been recent work in this domain on languages including Lao (Enfield 2007: ch. 5; 2015: ch. 5). On Vietnamese, see Srichampa (2008) on politeness strategies, and Sidnell and Shohet (2013) on avoidance strategies (see also Luong 1988). Linking social life to central concerns of historical linguistics and typology, there has been recent work on sociolinguistic conditions for borrowing (Alves 2009). For similar work see Thurgood (2010) comparing two varieties of Cham with the Tibeto Burman language Anong. And a new line of work in MSEA is in conversation analysis: Enfield (2013) presents several case studies of

Lao language in conversation; Hå (2010, 2013) presents studies of Vietnamese conversation with a focus on the role of prosody, for example in repair and backchannelling (see also Umaporn 2007 on backchannelling in Mon); and a series of in progress studies by Jack Sidnell and colleagues examine strategies of repair in Vietnamese. Constraints of feasibility and the remit of the series mean that the coverage of linguistic topics in this book is necessarily incomplete. I look forward to the day when a full length volume on the sociocultural anthropology of language in MSEA can complement the present book's focus on historical and structural linguistic matters.

ACKNOWLEDGEMENTS

I thank all those who have given me comments on material that has found its way into this book, improving it immeasurably: especially Peter Bellwood, Marc Brunelle, Tony Diller, Bob Dixon, Bill Foley, Gwen Hyslop, Nerida Jarkey, Weijian Meng, Pittayawat Pittayaporn, Mark Post, Martha Ratliff, Jack Sidnell, Paul Sidwell, David Solnit, and Martin Stuart Fox. I thank my research assistants for indispensable help: Weijian Meng, Gus Wheeler, Marlena Lutz Hughes, and Naomie Nguyen. The maps and figures were created by Gus Wheeler. The Chinese language sources were accessed by Weijian Meng. I am grateful to Angela Terrill, of Punctilious Editing Sweden, for her genuinely punctilious work in compiling the three indexes. I am grateful to the anonymous reviewers for their generous advice, to the editorial board of the Cambridge Language Surveys series, and to Kate Brett and Helen Barton at Cambridge University Press for their guidance and support. This work was supported in part by the Australian Research Council (Discovery Project DP170104607).

NOTE ON THE NATURE AND ORGANIZATION OF THE BOOK

This book incorporates most of the material included in *Mainland Southeast Asian Languages: A Concise Typological Survey* (Cambridge University Press 2019), which focused solely on typological features of the spoken languages of core MSEA. At nearly twice the length of that book, the present book adds two main things: (1) it expands the data coverage and analysis in the descriptive/typological chapters (3–7), and (2) it places the linguistic diversity of the area in historical context, both in terms of the social (pre)history of the mainland Southeast Asia area (in Chapter 1) and in the historical comparative linguistics of the area's languages (in Chapter 2).

Readers interested in the historical and social context of the languages of mainland Southeast Asia will get most out of Chapter 1, which gives an overview of the prehistory and history of the area. Chapter 2 focuses on historical comparative research, outlining current and evolving views on the relatedness of modern MSEA languages, with many references for those who wish to follow up on the details. The rest of the book covers grammatical topics, moving through broad categories of structural organization. A general outline of the typology of the area's languages is provided in Chapter 3, and subsequent chapters delve into details under four rubrics: Phonological systems (Chapter 4), Word formation (Chapter 5), Reference and nominal syntax (Chapter 6), and Predication and clausal syntax (Chapter 7). The organization of chapters 6 and 7 is based loosely on a functional distinction between reference and predication, the two basic elements of any proposition. Of equally crucial functional importance are devices for textual cohesion and for managing interpersonal relations in language use. These are covered under the appropriate functional domain in which they occur (for example, politeness based choice occurring in pronoun systems, discussed in Chapter 6) or are included in Chapter 4 (e.g., sentence final particles for the expression of speech act and other interpersonally grounded pragmatic distinctions).

This book cites language examples from a wide range of sources. We have endeavoured to represent data exactly as given in the original sources, though we note an occasional

exception made out of convenience: we have sometimes used ‘*a*’ (italic *a*) for ‘*a*’ (roman *a*) and vice versa, as long as this has not introduced ambiguity internal to the example cited (i.e., by collapsing the two characters). For this reason (and also as a matter of general principle), readers are advised to consult the primary sources before citing the examples.

ABBREVIATIONS

I	first person	3SGA	third person singular
2	second person	A	A argument
3	third person	3SS	third person singular
1DU	first person dual		subject
1P	first person	A	A argument
1PL	first person plural	ACC	accusative
1S	first person singular	ACHV	achievement
1SG	first person singular	ACOP	attributive copula
1SS	first person singular subject	ADJT ADVRS	adjunct adversative
2DU	second person dual	AFF	affirmative
2HON	second person honor	AGR	agreement
2P	second person	ALL	allative
2PL	second person plural	ALP	attributive linking
2S	second person singular		particle
2SG	second person singular	AMB	ambient noun with
3A	third person A argument		weather predication
3D	third person dual	AMSL	ambifunctional selector
3DU	third person dual	ANIM	animate
3P	third person plural	ANTICAUS	anticausative
3PL	third person plural	AO	agent orientating
3PLA	third person plural A argument	ASP ASS	aspect associative
3PLS	third person plural S argument	ASSOC AT	associative locative preposition
3POSS	third person possessor	ATT	attainment
3S	third person S argument	AUG	augmented
3SG	third person singular	AUX	auxiliary

B	bare	EXPR	expressive
BRO	brother	EXIST	existential
CAUS	causative	EXST	existential
CAUSE	causative	EXT	external
CL	classifier	F	female
CLASS	classifier	FA	familiar
CLF	classifier	FUT	future
CMPL	completive	FW	functional word
CMPR	comparative	GEN	genitive
COLL	collective	GRP	group marker
COM	comitative verb particle	GSL	group selector
COMP	complementizer	HIRESP	high respect
COND	conditional	IDEO	ideophone
CONT	continuous	IMM	imminent
COP	copula	IMP	imperative
CT	class term	IMPEF	imperfective
DAT	dative	IMPERF	imperfective
DECL	declarative	IMPF	imperfective
DEM	demonstrative	IMPFV	imperfective
DEP	dependent	IMPORT	importance
DET	determiner	INCL	inclusive
DID	verbal prefix (<i>d</i>) <i>id</i>	INDEF	indefinite
DIM	diminutive	INSTR	instrumental
DIR	directional	INTER	interrogative
DISC	discourse particle	INTJ	interjection
DIST	distal	INTNS	intensive
DO	direct object	INTR	intransitive
DST	distal	INTRG	interrogative
DU	dual	IO	indirect object
DUR	durative	IRR	irrealis
E	elder	LOC	locative
EBR	elder brother	LORESP	low respect
EMP	emphatic	LP	linking particle
ERG	ergative	M	male
EUPH	euphemism	MC	modifier classifier
EVID	evidential	MID	middle voice
EX	exclusive	MOD	modality
EXCL	exclusive	N	noun
EXP	experiential	NEG	negation

NEG1	first position negator	QM	question marker
NEG2	second position negator	QPLR	polar question marker
NFIN	non finite	QUE	question marker
NFUT	non future	QW	quantity word
NMLZ	nominalizer	R	realis
NMZ	nominalizer	RCNT	recent
NOM	nominalizer	RCP	reciprocal
NON	non	RCPL	reciprocal
NVOLT	non volitional	RDP	reduplication
OBJ	object	RDUP	reduplication
OBL	oblique	REC	reciprocal
OBLIG	obligatory	RECIP	reciprocal
OCOMP	object of comparison	RECP	reciprocal
P	polite	RED	reduplication
PART	particle	REF	reflexive
PASS	passive	REFL	reflexive
PAST	past tense	REL	relativizer
PATIENT	patient marker	RESP	respect
PCL	particle	RT	relational tense
PERF	perfective	S	S argument
PFV	perfective	SAT	quantifier/intensifier
PL	plural		(s)at
PN	proper name	SC	speaker conclusion
POL	polite	SEQ	sequential
POSS	possessive	SFP	sentence final particle
PP	pragmatic particle	SFP1	first position sentence
PREF	prefix		final particle
PRF	perfect	SFP2	second position
PROG	progressive		sentence final particle
PROH	prohibitive	SG	singular
PROX	proximal	SIB	sibling
PRT	particle	SP	species
PRX	proximal	SPMY	socially conditioned
PTC	particle		pronoun, male, younger
PTCL	particle	SPNY	socially conditioned
PV	verbal particle		pronoun, neutral,
PVF	final verbal particle		younger
PVP	postverbal particle	STAT	static aspect
Q	question marker	SUB	subject

TELIC	telic	YBR	younger brother
TGTHR	together	YN	yes no
TLNK	topic linker		
TOP	topic		<i>Language Families</i>
TPC	topic		
UNIMP	unimpeded	AA	Austroasiatic
USL	unit selector	AN	Austronesian
V	verb	HM	Hmong Mien
V.S.	verb suffix	SN	Sinitic
VBLZ	verbalizer	TB	Tibeto Burman
Y	younger	TK	Tai Kadai

Context

1.1 The Mainland Southeast Asia Region

Mainland Southeast Asia can be broadly defined as the area occupied by present day Cambodia, Laos, Peninsular Malaysia, Thailand, Myanmar, and Vietnam, along with areas of China south of the Yangtze River. Also sometimes included are the seven states of Northeast India.

There are different interpretations of what is and is not included in MSEA, but a core area is generally assumed (Comrie 2007: 45). MSEA is always taken to include former Indochina – Vietnam, Laos, and Cambodia – together with Thailand, and, usually, Peninsular Malaysia and part or all of Myanmar. In this book, our scope is Greater MSEA, thus including regions of southern China and northeast India, but we will naturally tend towards a focus on Core MSEA (see Map 1.1).

MSEA is a tropical and sub tropical area with rugged and well forested hills and river systems running from higher altitudes in the northwest to the plains and deltas of the south. Among the biggest rivers are the Mekong, the Brahmaputra, the Red River in north Vietnam, the Salween and Irrawaddy rivers in Myanmar, the Pearl and Yangtze rivers in China, and the Chao Phraya in central Thailand. The lower reaches of these river systems are fertile alluvial plains, which have attracted people partly because of the mobility the environment affords, but also because of their suitability for paddy rice farming. Paddy farming, in which rice plants are kept continually flooded as they grow, requires management of water via systems of dykes and channels (O'Connor 1995, Hartmann 1998; note, though, that shallow flooded plains do occur naturally, as around the Tonle Sap in Cambodia). This method is significantly more productive than upland dry field methods, and can support larger populations (Bellwood 1992: 90). It also reduces biodiversity. Geography has played an important role in the historical demography of the area, as outlined in section 1.4 on the history of migrations, livelihoods, communications and politics over the last several millennia.



Map 1.1 *Mainland Southeast Asia*

Present-day Cambodia, Laos, Peninsular Malaysia, Thailand, Myanmar, and Vietnam, along with China south of the Yangtze River, and Northeast India.

Sources: Esri, Airbus DS, USGS, NGA, NASA, CGIAR, N. Robinson, NCEAS, NLS, OS, NMA, Geodatastyrelsen, Rijkswaterstaat, GSA, Geoland, FEMA, Intermap, and the GIS user community

1.2 Mainland Southeast Asian Languages

The degree of linguistic diversity in MSEA (i.e., the number of languages per square km) is high (Enfield 2011b), and it is highest in upland areas. Lower language density in lowland areas is due in part to the effects of geography on the nature of social networks (see Nettle 1999) and in part to politics and the exercise of social power. Historical demographic processes of the kinds discussed in Chapter 2 caused formerly diverse lowland communities in MSEA to become homogenized through a combination of two processes. One process was ethnolinguistic shift. Some groups stayed where they were but stopped passing on their languages and identities to their children, instead adopting the languages and identities of new dominant groups. This process has been taking place for at least 2,000 years in the area and can be observed all over MSEA today. Another process was out migration, typically to more isolated hill areas (Scott 2009). Geographical isolation is a force that still promotes language diversity in the region, where former diversity of lowland areas is decreasing. Many MSEA languages are heavily endangered (Matisoff 1991a, Enfield 2006a, Bradley 2007, Premsrirat 2007). This is exacerbated by effects of the concentration of political power of modern nation states in the lowlands. In recent decades, processes of language standardization in MSEA nations (Simpson 2007) are heavily reducing language diversity.

The languages of MSEA are from five major language families:¹ Sino Tibetan, Tai Kadai, Hmong Mien, Austroasiatic, and Austronesian.² More than 550 distinct languages are spoken in greater MSEA. If we exclude the China and India data, thus representing the core MSEA area, the number of languages is about half this amount; see Table 1.1.³

The high linguistic diversity in northeast India and southern/southwestern China adds dramatically to the number of languages included in the estimate for this area. It also reverses the relative proportion of Sino Tibetan and Austroasiatic languages.

¹ Languages of other language families not noted in Table 1.1 include widely used languages of colonial origin English and French along with other languages used in the MSEA area by (descendants of) migrants and travellers. These include languages of South Asia and the Middle East, and languages from Japanese to Yoruba that may be encountered in major urban centres of MSEA such as Bangkok, Phnom Penh, or Hanoi. They also include sign languages of the Deaf (see section 2.9). In Myanmar, the Indo-Aryan language Rohingya (arguably in a dialect relationship with Bengali) is spoken by around a million people in Rakhine State.

² The Andamanese languages are located just outside MSEA as defined here. For work on these lesser-known languages, see the bibliography in Comrie and Zamponi (2019).

³ These figures were assembled by Weijian Meng. Language coordinates are from: Glottolog 4.0, <https://glottolog.org> (accessed 3 July 2019). Administrative borders are from Natural Earth, www.naturalearthdata.com (accessed 26 Aug 2019), river centre lines are from Natural Earth www.naturalearthdata.com (accessed 5 Sep 2019) and the Harvard WorldMap, https://worldmap.harvard.edu/data/geonode:chiangyangtze_river_tnc (accessed 26 Aug 2019). Core MSEA was defined for this count as Cambodia, Laos, Myanmar, Thailand, and Vietnam; Greater MSEA includes this together with Peninsular Malaysia, areas of India east of 90 degrees (i.e., the states of Arunachal Pradesh, Nagaland, Manipur, Mizoram, Assam, Meghalaya, and Tripura) and China south of the Yangtze river (specifically, the provinces of Zhejiang, Jiangxi, Hunan, Guizhou, Yunnan, Guangxi, Guangdong, Fujian, and Hainan).

Table 1.1 *Languages of MSEA: (a) numbers of languages in Core MSEA, by family; (b) numbers of languages in Greater MSEA, by family*

	Core MSEA		Greater MSEA	
Austroasiatic	122	(44%)	138	(24%)
Sino-Tibetan	74	(26%)	288	(49%)
Tai-Kadai	51	(18%)	93	(16%)
Austronesian	25	(9%)	26	(4%)
Hmong-Mien	8	(3%)	38	(7%)
Total	280		583	

Core MSEA is defined as Cambodia, Laos, Myanmar, Thailand, and Vietnam. Greater MSEA is defined as Core MSEA plus states of India to the east of 90°E, China south of the Yangtze river, Peninsular Malaysia.

There is good agreement among specialists of MSEA languages as to the basic language family affiliation of known languages. There are unresolved issues about lower level subgroupings and there are unresolved hypotheses about possible macro groupings. But for every known language, scholars agree about which of the five main language families it fits into. This is unusual firstly because it means that each language's basic affiliation is apparently uncontroversial, and secondly because no language isolates have yet been identified (cf. Blench 2011: 125–6). To put this in context, there are an estimated 136 language isolates in the world, of which nearly two thirds come from the Americas, and only 7% come from anywhere in Asia (and none from MSEA; Campbell 2010). For more on the historical linguistic background of MSEA, see Chapter 2.

This book provides information on the history and classification of MSEA languages, and the main linguistic properties of the languages. The scope is necessarily limited, given the number and diversity of languages under discussion. For more detailed coverage of specific sub areas or subsets of languages in the area, see Jenny and Sidwell (2015) on Austroasiatic languages, Diller *et al.* (2008) on Tai Kadai languages, Thurgood and LaPolla (2017) on Sino Tibetan languages, Grant and Sidwell (2005) on Chamic and other mainland Austronesian Languages, and Ratliff (2010) on Hmong Mien languages, along with Simpson (2007), Goddard (2005), Comrie (1990, 2007), and Vittrant and Watkins (2019) on MSEA more broadly, and the many references in all of those sources, as well as in this book.

1.3 Nomenclature, System Ontology, and Language Data Selection

In any discussion of languages, language families, and ethnic or cultural groups, we inevitably encounter problems of nomenclature. Countless terms in hundreds of

languages have been used to label human groups in present day and historical MSEA. Not one of them refers to a homogenous entity. Like any other linguistic category, a label for a human group is a useful but imperfect tool that glosses over differences between members of the category.

This glossing over of difference can be used for good or ill. The utility of labels for human groups is not only practical but political. This is obvious in the case of terms like Thai and Vietnamese, which can refer to identities associated with nation states, or terms like Hmong and Karen whose members are politically organized. But it is no less true in the case of names for small ethnolinguistic groups such as Kri and Chong. And the referents of labels for historical groups, such as Pyū and Hoabinhian, can be similarly strategic and potentially misleading in their implication of category internal uniformity (see below for discussion). As we shall see, there is little evidence that these terms refer to people who shared a unitary linguistic, cultural, or ethnic profile.

All ethnolinguistic terms are fluid and contestable. They all involve some form of essentialization and/or naturalization by insiders and outsiders, including academics and other researchers. Such essentialization or naturalization can be problematic if it is based on false or inaccurate understanding, or if it is put to certain prejudicial uses. But it is ultimately necessary for any discourse about language, for two reasons. First, we cannot talk about any natural phenomena without generalizing in some way. The issue is not whether we generalize. It is whether we generalize in relevant, accurate, fair, and productive ways. Second, while there is always variation within any language ecology, languages exist. They have system properties. They enable communication, just as they create real barriers to communication. They serve conventional and collective functions within definable human groups, and as such they bind groups of people together, defined as those who are socialized within a shared linguistic system, who have an understanding of that system, and who can use it.

Of course, the names of language systems do not equal the systems themselves. There are important issues concerning ethnolinguistic nomenclature and its uses. One problem is that most languages or human groups can be referred to by more than one label. An example in MSEA is the Northern Tai language spoken in certain villages of the middle Nam Noi valley in upland Central Laos. This language is referred to as Saek [sæ:k] in the linguistic literature (e.g., Hudak 2008: 53–8, 2010), [z̥ɛ:k] by the villagers themselves, [zø:?] by Kri speakers in neighbouring villages a few kilometres upstream, and [jɔ:j] by Brou speakers in neighbouring villages a few kilometres downstream. This sort of situation is globally the norm. English, for example, is also called *Engels*, *Anglais*, *Paynekit*, and *Yīng Wén*, among many other names, depending on the language spoken. It is important to keep track of these different labels.

A second issue is that language names can change over time, as can the connotations of those names, and in turn people's preferences and practices in language naming.

A language name that is established and accepted now may become marked or pejorative later, and speakers of the language may come to prefer that the name not be used. For example, Lao speakers today widely refer to the Hmong language as *mong4*. Another word for the language, and the ethnic group more broadly, is *mèèw4* (cf. Chinese *Miao*). A few decades ago, the word *mèèw4* was an unmarked term in Lao, but today it is considered by many Hmong, and some Lao, to be pejorative and offensive.

A third problem concerning the labelling of linguistic entities is the theoretical question alluded to already of whether there is in fact any real entity to be labelled. A classical puzzle in linguistics is to determine, for two language varieties, whether the varieties are two separate languages or two dialects of the same language. One way to answer the question is with reference to the criterion of mutual intelligibility. If the two varieties are mutually intelligible, despite certain readily defined differences, then they are dialects of a single language. If they are not mutually intelligible, then they are separate languages. Another way to answer the question is with reference to sociopolitical criteria. If the labeller's goal is to demarcate two communities as having separate identities, then their varieties may be labelled as different languages, even if they are mutually intelligible. This is the case, for example, for the two language varieties of MSEA known as Thai and Lao. On linguistic grounds, their high degree of mutual intelligibility would lead us to conclude that they are dialects or varieties of a single language. Conversely, if our goal of labelling were to convey that the two groups are part of one sociocultural entity, then we might want to label these significantly different language varieties in the same way. This explains why Mandarin, Cantonese, and Shanghainese are often referred to as 'dialects of Chinese'. These three language varieties are mutually unintelligible and would on technical linguistic grounds be regarded as separate languages. But on sociopolitical grounds, they may be grouped together on the basis that they are spoken by communities who are regarded as part of a single Sinitic cultural sphere.

When we consult secondary sources, it is important to be mindful of the possibly quite distinct reasons a certain language name may have been used. There are different kinds of categorization, and it is easy to mistake one kind of categorization for another. Many first time travellers to Southeast Asia will assume that Thai and Lao are not mutually intelligible, simply because they are referred to as different languages and are written using different scripts. But their categorization as different languages is made on primarily political grounds. Particular caution is needed with linguistic categorization and nomenclature offered by non specialists, such as in the writings of early colonial surveyors and explorers, or modern day tourists and amateur observers. On the problem of defining language and culture groups in Myanmar, Watkins (2007: 277) remarks: 'The taxonomic fervour of colonial ethnographic surveyors and their lack of accuracy in

discerning genuine divisions among continua of cultures and languages has been frequently remarked on with regard to South Asia, and a warning about the over simplification of ethno linguistic categorization is equally well warranted in the context of Burma/Myanmar'. The same is true for MSEA more broadly.

This point has been appreciated in postcolonial anthropology since at least Leach (1954). Leach's classic study of the social interrelations among highland Burmese groups established that the classical assumption of ethnography 'treating culture groups as social isolates' does not hold up:

In the Kachin Hills Area as a whole we find a considerable number of named groups culturally distinct or partly distinct. In places these groups are segregated into fairly well defined areas, in other places they are all jumbled up. A study of Kachin social organization cannot therefore proceed in the classical manner which treated culture groups as social isolates.

(Leach 1954: 60)

Leach showed that 'the social system is not uniform'. Instead, people of different culture groups – people who speak different languages and identify as ethnically distinct – are 'part of a single social system', where the possibility of shifting cultural identity is natural and commonplace (Leach 1954: 60). There may be periods of equilibrium, but 'any such equilibrium as may appear to exist may in fact be of a very transient and unstable kind' (Leach 1954: 61).

We can draw an analogy between the stability of ethnic groups and the stability of landscape features around small scale alluvial river systems. When I first travelled to the upland riverine area of Laos in which Kri speakers live, I naïvely assumed that the streams, pools, crossings, and riverbanks I encountered were permanent features. I was puzzled that stairways and bridges were only crudely constructed with a minimum of effort, and thus with no longevity. In subsequent trips I learned that river morphology can develop rapidly. Changing flows quickly transform streams into stagnant ponds, washing old walking trails away, opening up new lines of access, and so on. The same fluid nature is found in social systems, ethnic identities, and inter ethnic relationships. What upon first impression may seem like an established structure can be quickly reorganized. People will adapt to the changes with equal flexibility.

This is the essence of Leach's point about the dynamic social system that Shan people and Kachin people shared in the Kachin Hills. Their interdependent social subsystems allowed such flexibility that it was possible to speak of 'Kachins becoming Shans or of Shans becoming Kachins' (Leach 1954: 61). The system 'has no stability through time', Leach suggested. 'What can be observed now is just a momentary configuration of a totality existing in a state of flux' (Leach 1954: 63).

This leads us to a more fundamental question about labelling languages. If a language is an imagined entity and/or merely a momentary configuration of something that is in flux, do we have the right to label languages and thereby treat them as things at all? There are real issues of language ontology, genuine questions about what kind of a thing a language is (Harris 1980). But these issues should not prevent us from describing such systems as if they were stable structures. For one thing, the fact that the systems are distributed, dynamic, and reconfigurable does not mean that the systems do not exist. They do exist, as evidenced by the fact certain groups of people collectively understand, and can productively use, certain sets of words, phrases, idioms, and not others. Another reason is that the dynamic, distributed processes at hand are not always obvious or even detectable within our lifetimes, so from the point of view of the individuals who live within these systems, as members of social groups and as agents of cultural transmission, the systems in question may be effectively fixed and permanent. A third reason is that without labelling them and treating them as real, we would be unable to talk about, analyse, or understand the linguistic phenomena we wish to focus on in this work.

Ontological questions about the reality of social systems like languages and cultures are dealt with by taking the nuanced conceptual approach that social reality demands.⁴ As Searle (2010) argues, a piece of social reality can be both subjective and objective in nature, because it is *ontologically subjective* (its existence depends on human perspective and experience) while being *epistemically objective* (the truth of claims about it can be settled as a matter of fact). In any case, for the purposes of this book we need to leave these questions aside and get on with the job. This may be an act of essentialism. If so, it is a strategic one. It is necessary for our purposes.

The naming of language *families* runs up against two further problems in addition to those just discussed for the labelling of languages and sociocultural groups.

The first problem with language family labels is that individual language names are often used to index entire groups of languages. In MSEA linguistics, language family labels including Mon Khmer, Hmong Mien, and Tibeto Burman do this, following a common recipe. They pair two individual language names to denote a group of many languages. The languages that are foregrounded in the family labels happen to be spoken by the largest populations and happen to be culturally and sociopolitically

⁴ Discussing anthropological research in our area, O'Connor (1995: 968–9) notes: ‘Once ethnologists studied Southeast Asia as a region (Heine-Geldern 1956; Sharp 1962; Burling 1965; Provencher 1975; Keyes 1977), but their empirical generalizations have given way to an anthropology of discrete cases. While archaeologists still piece together regional types (Bellwood 1985; Higham 1989), ethnographers dote on the people or village they study ... [M]any ethnographies naively treat their subjects as if they existed and were knowable apart from larger conditions and logical models. In this fieldwork-driven empiricism where anthropology collapses into ethnography (Stocking 1992: 357, 362–72), local description, synchronic constructs, and inductive reasoning have come to displace the regional comparison, historical perspective, and deductive logic that rigor and balance require.’

dominant. For this reason (in part), Blench and Post (2014: 92–3) regard Sino Tibetan as a ‘highly inappropriate name for the phylum’. They say the same for Tibeto Burman. They argue that the named languages have ‘no special classificatory significance’ and that the political or cultural prominence of these languages is not a good reason to foreground or privilege them. Furthermore, privileging them in this way comes at the expense of sociopolitically less powerful groups of people, or of languages that do have special classificatory significance. Blench and Post (2014: 93) endorse an alternative term Trans Himalayan, which they say ‘would capture the geographical locus of much of the phylum without emphasizing individual subgroups’.

But this strategy of using a geographical designation to label a language family raises a second potential problem, as LaPolla writes:

If we look at the various proposals for subgrouping in Tibeto Burman, almost all include geographic designations such as ‘Western Himalayan’. These designations assume that either the languages involved have always been at that location or that all the languages developed from a single ancestor which migrated to that location at some time in the past. Yet we know that there have been waves of migration, particularly into and/or through certain regions (see LaPolla 2001), and so there is a suspicion that the languages were not originally closely related, but have come to seem similar because of long term contact.

(LaPolla 2012: 120)

Issues of nomenclature are important, particularly for our practices in future research in which we have a responsibility for referring to language varieties in clear, informed, and appropriate ways. Labels for ethnic groups are not straightforwardly discovered or created. Linguistics has to make do with secondary sources, and often the materials we consult are produced in very different contexts from our own, by people with different motivations, standards, and knowledge from our own.

However, with all of this in mind, it would be impossible to revise or update all of the terminology in the literature, or to otherwise adjudicate on the implications, connotations, or appropriateness of every language or culture label cited in this book. I therefore make no attempt to change the nomenclature derived from secondary sources. I will note if or when clarification of reference to a language or a language family is needed.

The language selection drawn on in this book is, by necessity, a selection of convenience. Of the many hundreds of languages in the MSEA area, only a small fraction have been described in detail, and so there is a bias towards languages that happen to be well described in published sources. A problem with this is that the best described languages may not necessarily be the most important for research purposes, depending on the question (though they may be of broader public interest).

If we are going to understand human language, then we need representative data. Our null hypothesis must be that each language is of equal value as a data source (with the caveat that many languages do not represent independent data points because of their historical relationship with each other; see Galton's remarks in Tyler 1889). The starting assumption has to be that a language like, say, Kri, spoken by 500 people in an isolated upland pocket of central Laos, will be no less important as a source of evidence for a science of language than Vietnamese, which is spoken by up to a hundred million people. That said, there are reasons why we might want to concentrate our efforts in research on a big language like Vietnamese. Such work might be of interest to larger numbers of people (including non specialists). And the available information on Vietnamese is likely to be more reliable, detailed, and extensive. But to the extent possible, in this book I have endeavoured to shift the balance to lesser known languages.

My chosen emphasis on lesser known languages raises two points I would like to emphasize. The first is that even with a significant recent increase in the amount and quality of research reports on lesser known languages, we are still in need of good quality information about the many languages of MSEA: field research with, and by, speakers of the languages is needed. The second point is that when you think of MSEA languages, you should think first of minority, non official languages like Semelai, Lahu, Saek, Moken, Kri, or Mien, and not those that are usually thought of first or cited most often: Thai, Lao, Khmer, Vietnamese, or Burmese. As we shall see through the rest of this book, the minority languages are more representative of the area, in terms of history, sociolinguistics, and typology.

1.4 **History of MSEA**

MSEA has seen a long and complex history of human movement, contact, and diversification, leading to the linguistic diversity that we see today. This section provides an outline of the historical processes that often indirectly help to answer questions of what MSEA languages are like and why.

1.4.1 Prehistory

The land mass that we now identify as MSEA is a recent formation in geological time. The mainland area as we know it was formed by processes of separation, rotation, and collision of tectonic plates over many millions of years. By around 40 million years ago, the MSEA area had begun to take recognizable shape, most notably due to the collision with Asia of the Indian subcontinent, moving up from the southwest, and the coming together of the Indochina, Shan Tai and south China cratonic areas (Bunopas and Vella

1983, Hall 1996, Meyerhoff *et al.* 1996, Longley 1997, Murphy 1998, Domeier 2018). These processes gave rise to the area's mountain chains, most notably the Annamite chain running from the southeast of MSEA to the Himalayas, as well as other significant ranges in the region. These geological features account for the shape and orientation of the main watersheds of MSEA, which in turn have shaped the natural resources that have afforded certain livelihoods and patterns of migration in MSEA prehistory, as described below. In turn, those patterns of livelihood and migration have ultimately shaped the evolution and diversification of human language in MSEA.

While the land mass of MSEA has existed in roughly its present shape for at least ten million years, this of course pre-dates the evolution of modern humans and thus of any human language(s) by many orders. We can assume based on current evidence that the number of languages spoken in the MSEA area up until around 70,000 years ago was zero. In the subsequent seventy millennia, the development and diversification of human societies have brought about today's situation in which more than five hundred distinct languages are spoken. And hundreds if not thousands of distinct languages have existed in MSEA during that period, but we have no information about them. We focus in this book on just those languages that can be observed today.

Anatomically modern human remains found in a limestone cave in northern Laos are dated to around 70,000 years ago (Demeter *et al.* 2017, Shackelford *et al.* 2018). Remains from a Pleistocene cave in Sumatra have similarly been dated at between 73,000 and 63,000 years ago (Westaway *et al.* 2017). At that time, Sumatra was joined by dry land to today's mainland due to low sea levels during Marine Isotope Stage 4 (MIS 4; Hope 2005 and Martinson *et al.* 1987; cited in White 2011: 14; see Map 1.2).⁵ These discoveries indicate that the first arrivals of modern humans in MSEA are associated with the human dispersion out of Africa that took place during MIS 4 and MIS 3. We have no idea what language these people spoke. As we shall see below, almost all modern MSEA peoples, including those who would be regarded as indigenous peoples today, have entered the area within the last 2,000 years.⁶

Between the Last Glacial Maximum (around 26,500 years ago) and the beginning of the Holocene (the modern era, beginning around 12,000 years ago) – that is, during MIS 2 sea levels changed dramatically from a human perspective (see Figure 1.1). At the extreme of the Last Glacial Maximum, global sea levels were 120 m lower than they are today (see

⁵ Marine Isotope Stages (MIS) are alternating warm and cold phases in the Earth's history. Core samples of sediment from the ocean floor reveal oxygen isotope variation which is correlated with greater and lesser global ice volumes, and hence lower and higher sea levels, respectively (Emiliani 1955, Aitken and Stokes 1997). The stages are numbered going back from the present: odd numbers are warming stages, even numbers are cold.

⁶ A possible exception: the 'Negrito' people of peninsular Malaysia are argued to share genetic material with those earliest arrivals in MSEA (see Wangkumhang *et al.* 2013, Hill *et al.* 2006, Thangaraj *et al.* 2006, Macaulay *et al.* 2005).



Map 1.2 'Sundaland'

Sundaland is shaded white: the continental shelf of dry land exposed when sea levels were 120 m lower than they are today during the Last Glacial Maximum, approximately 26,500 years ago. Current-day coastlines are also shown for reference. The oval shows the approximate range of Hoabinhian peoples, who left behind evidence of a lithic technological tradition from as early as 40,000 years ago up until as late as 500 BCE.

Map 1.2; Chappell and Shackleton 1986, Tooley and Shennan 1987). This implied radically different possibilities for human movement and livelihoods to those we see today. At that time, a person could walk on dry land in a straight line from the site of present day Ho Chi Minh City to Kuala Lumpur, and then in another straight line to Bali and again up to Brunei (Voris 2000, Oppenheimer 2011, White 2011). Today, these places are separated by sea.

1.4.2 Hoabinhian Neolithic Tradition

There is archaeological evidence of human activity in MSEA well before the adoption of agriculture, and far beyond the limited time horizon of historical linguistics. Evidence of a stone tool industry (or lithic technological tradition; White 2011: 27) is attested

across a wide swathe of the mainland from as far back as 43,500 years ago (Ji *et al.* 2016, Forestier *et al.* 2017) up until about 2,500 years ago (White 2011). The label Hoabinhian is widely used as a catch all term for people who lived over an area thousands of kilometres across, and over a period of many thousands of years (Gorman 1969, 1970, 1971). See Map 1.2.

As Bellwood (1992: 86) notes, the term Hoabinhian should not be understood to denote a unified group, as there was ‘a great deal of local and regional variation which still remains to be documented’:

The use of such a term for all early Holocene assemblages of mainland Southeast Asia from beyond the Tropic of Cancer almost to the equator does not imply technological homogeneity, or that the makers were necessarily closely related in linguistic or biological terms.

(Bellwood 1992: 86)

Did Hoabinhian populations play a role in the development of agriculture in MSEA? There is evidence in Hoabinhian sites of hunting and food gathering activities – for example, shell deposits and animal bones of all sizes, from rats and squirrels to rhinoceros and elephants – but little to suggest animal domestication (Gorman 1969, Bellwood 1992: 87–8). These groups represented ‘a stable and broad spectrum hunter gatherer adaptation which may have lasted in this remote region to the first millennium CE, albeit with earlier sporadic contact with lowland agricultural groups’ (Bellwood 1992: 88).

1.4.3 Earliest Agriculture

From around five thousand years ago, the archaeological record in MSEA complexifies dramatically, with evidence of diverse tools, pottery, burial practices, and animal domestication (Higham 2002, White 2011). Soon after, from around four thousand years ago, evidence shows that there was a rapid and widespread introduction of agriculture in the area. Agriculture is one of the key drivers of social transformation in prehistory, and its development and spread are directly implicated in the historical diversification of languages.

Was agriculture innovated by *in situ* populations, or was it introduced by outsiders? Given the complexities of agricultural practice, and the major investment of labour needed, some commentators find it unlikely that ‘successful early Holocene foragers who had sufficient food and experienced no pressure on their resources would have wanted to change’ (Bellwood 1992: 88). ‘Modern hunter gatherers in Southeast Asia generally resist the total adoption of agriculture unless shrinking land resources leave them with few other options, and many hunters and gatherers of the Holocene probably did not switch their economies simply because concepts of food production entered their

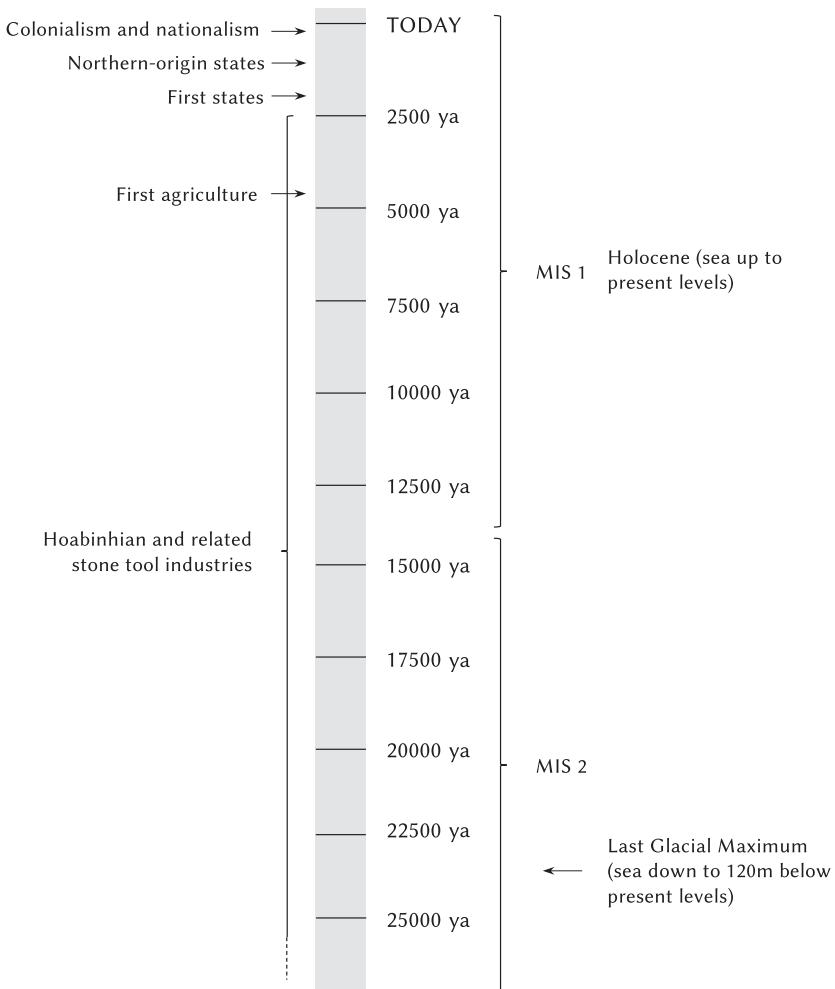


Figure 1.1 *Chronology, I: Mainland Southeast Asia during the last two Marine Isotope Stages (MIS 1, 2)*

lives' (Bellwood 1992: 90). The implication is that agriculturalists migrated into the areas, from the north, bringing agriculture with them, and eventually replacing or assimilating previous inhabitants. Here is a clear statement of Bellwood's influential view:⁷

⁷ The question is not settled. See White (2011) and references therein for discussions of the possibility that agriculture in MSEA may have involved a greater degree of indigenous innovation than Bellwood here suggests.

In most of Southeast Asia (excluding southern China and possibly coastal northern Vietnam) there is no evidence to suggest that any primary development of agriculture occurred, and it seems basically to have been introduced by people already acquainted with the cultivation of rice, millet and other subtropical crops like yams, taro and sugarcane. They also kept domesticated pigs, chickens, dogs, and perhaps cattle. All evidence to hand suggests that this expansion into Southeast Asia commenced mainly from the coastal regions of southern China. However, it is very important to emphasize that many tropical fruits and tubers native to Southeast Asia were brought into cultivation systems as they expanded southwards, and existing foraging populations may well have contributed useful knowledge of such plants to agricultural groups.

(Bellwood 1992: 91)

The most important element of the spread of agriculture into core MSEA especially is the suitability of much of the land for wet rice agriculture. Incoming migrations of Tai speakers over the last millennium or so was facilitated by the availability of flat, swampy, alluvial land along the rivers of MSEA. This favoured constant migrations from southern China moving southwest into MSEA. ‘Under conditions of low population density people might have preferred to seek new natural swamps along rivers which flooded during the wet season, rather than to create their own by laborious processes of damming, bunding and water transport’ (Bellwood 1992: 93). Bellwood further notes that rice and taro ‘may both have been grown in natural wet fields of a simple and seasonal kind from the very beginning, and dryland shifting cultivation in most areas may be a secondary development’ Bellwood (1992: 93). Precisely this kind of mixed livelihoods system can be observed today in rural MSEA, especially in upland areas (White 2011).

While a fair amount is known from bioarchaeological evidence about more recent human activity in the pre agricultural period (Oxenham and Tayles 2006), the time horizon of comparative linguistics has been limited to the last few thousand years (for recent reviews relevant to MSEA, see the papers in Enfield 2011a; for new exploratory methods of language comparison being used to propose speculative dating of language splits going back 7,000 years, see Sagart *et al.* 2019). The beginnings of agriculture in MSEA are just behind that horizon.

1.4.4 Bronze and Iron Age States

After the beginnings of the spread of agriculture in MSEA, the next major development was the introduction of bronze around 3,000 years ago, followed by iron some 500 or so

years after that (Higham 1996). These developments drew on Indic, Chinese, and Austronesian influences. Tools included weapons such as spearheads and fishhooks, jewellery such as bracelets and beads, and tools including axes, bells, and bowls, and even ornate bronze drums found from Guangxi all the way down into the Malay peninsula. The transition to iron led to significant changes in MSEA societies and economies. The developments ‘may have caused a shift from a gently ranked lineage type of social organization, evidently with a fairly elaborate network of exchange of prestige goods, to one with truly stratified classes and a ruling echelon controlling considerable amounts of wealth, labour and perhaps military force and tribute’ (Bellwood 1992: 120). Settlement sizes expanded, with increasingly centralized organization.

These technological and associated social developments up until about 1,000 years ago implied new forms of organized and centralized government. Taylor (1992), Hall (1992), and O’Connor (1995) provide detailed overviews of these complex developments, which we draw upon in the rest of this section (see also Coedès 1968, Marr and Milner 1986). During the early centuries CE, an array of loosely organized states emerged, overseen by distinct kingdoms or polities, including the Pyū (in present day Myanmar), the Mon (in present day Myanmar and Thailand), the Khmer (in present day Cambodia, Thailand, and southern Vietnam), and the Cham (on the coast of Central Vietnam). None of these were monocultures or even clearly delineated kingdoms. They are as imagined, and as internally diverse, as any other political/ethnic historical category in MSEA.

Funan and Chenla

In the southeastern corner of Indochina, the sea oriented state of Funan centred in the lowland coastal area of the Mekong Delta had begun to emerge under Indic cultural influence by around 2,000 years ago. The archaeological evidence for this state ‘is manifested in walled cities, brick temples, extensive canals, a writing system and monumental statuary’ (Higham 2002: 243). Funan’s location had the strategic advantage of allowing participation in the southern, maritime silk route. ‘Control of this jugular node also involved the interior, the most likely origin for the wood, plumage and beasts which were sent as tribute to the Chinese court’ (Higham 2002: 243). Written records make reference to ‘kings, a legal system, taxation, warfare, and subjugation of rivals’ (Higham 2002: 243). By around 600 CE, Funan had declined as a centre of political power. This coincided with the development of interior agrarian states referred to under the rubric of Chenla. Several regional states showed greater population concentration and further development in social complexity from the seventh century CE onwards. Also during the first millennium CE, the coast of present day Vietnam saw

the development of sophisticated and powerful Cham states, which had relations with both Chenla in the inland and China to the north. ‘To judge from the records of Chinese wars and booty taken, the rulers not only disposed of substantial, well equipped armies which included war elephants and a fleet, but also amassed much wealth in the form of embellished religious statues and *linga*’ (Higham 2002: 278).

Champa

Polities known collectively as Champa were located along coastal areas of current day central and southern Vietnam from the late second century to the fifteenth century CE. Champa was dominated by speakers of Austronesian language varieties, accounting for the modern day distribution of Chamic languages in MSEA. The latter half of that period saw a concentration in territory towards the south and an increasing absorption within Vietnamese authority (though Chamic kings continued to reign in name in some areas until the nineteenth century).

Cham polities experienced war and other forms of interaction with neighbouring empires, including both China and Vietnam. There is no overarching narrative of Champa history. The realm consisted of distributed islands of authority. Champa was not a unified kingdom but ‘an archipelagoically defined cultural political space’, a ‘network, or series of networks, of ethnic, religious, political, and commercial relationships connecting the Cham territories with the Malay world of peninsular and insular Southeast Asia’ and also connecting them to upland minority peoples, who spoke Austroasiatic languages (Taylor 1992: 153, 156). But people of Champa remained centred on the coast. These dispersed Malayo Polynesians approximated a form of maritime nomadism in virtual islands on the rim of the MSEA mainland.

The earliest maritime trade between India and China took long routes that hugged the coastline, around the Gulf of Thailand. Funan – located around the Mekong Delta and on the coast of modern Cambodia – was a way station on that route. But in the beginning of the fourth century an all sea route came into use (Wolters 1967: 71–85, cited in Taylor 1992: 158–9). Boats then made a beeline for the coast northeast of the Mekong Delta from the straits of Malacca, bypassing Funan and reaching the coast closer to Champa terrain. This development played a role in precipitating a decline of coastal entrepôts and the rise of inland centres of power, or *mandalas*, whose economies depended on rice growing. By the sixth century, political power in MSEA had shifted from the China–India trade route locus of coastal Funan to the inland directed paddy rice agricultural orientation of Chenla.

Viet Dynasties

Meanwhile, further north, political developments were dominated by interactions with China. By the first century CE, Han Chinese armies had conquered the Red

River (in current day north Vietnam) (Taylor 1992: 137). For the following millennium, local leaders administered the northern part of Vietnam on behalf of successive Chinese dynasties, though at various times (e.g., during the eighth and ninth centuries) with a degree of autonomy. During the Tang era (618–907 CE), ‘the site of modern Hanoi became the political centre of Vietnamese lands’ (Taylor 1992: 137). In the tenth century, with the fall of the Tang dynasty, Vietnamese rulers ‘emerged as custodians of the political and strategic wisdom accumulated during the previous centuries of participation in China’s imperial system’ (Taylor 1992: 139). By early in the second millennium CE, self styled Vietnamese emperors had developed and strengthened their centre of power, through economic and military interactions with mountain minority people and with the Cham to the south. These early Vietnamese dynasties introduced and affirmed Mahāyānā Buddhism, incorporated indigenous spirit beliefs, and established tributary relations with China. Naval power played a role too. Subsequent centuries, notably the long Dai Viet period beginning in the mid eleventh century and ending at the turn of the nineteenth century, featured strongly Sinicised political and administrative institutions, including a complex bureaucracy with competitive entry, along with power relationships that combined indigenous elements (see Taylor 1992: 137–51).

Some summary key points from the history of early states in the southeast of MSEA can be identified (drawn from Higham 2002: 297):

Early First Millennium BCE

- independent village communities engaged in the exchange of prestige goods, and with lineage ranking
- development of differentially large centres which represented chiefdoms

Last Few Centuries BCE

- increased maritime exchange in strategic coastal tracts
- agricultural intensification and control over riverine routes to the interior plains or mountain passes
- coastal chiefs of lower Mekong accumulate and distribute novel prestige goods and commence local manufacture

First Millennium CE

- concentration of populations in large centres
- agricultural intensification by draining marshland
- warfare for territorial expansion and the accumulation of slaves
- conquered land placed under the relatives of this expanding state
- overlords adopted Indian script, titles, and Hindu religion

- interior chiefdoms developed links with coastal trading networks, which were advantageous for forming defensive alliances against larger predatory states
- friction with the north, in conjunction with changing maritime trade routes, saw the decline of the delta state of Funan
- power then passed into the hands of competing overlords in the interior
- alignment with Hindu religion was a selective advantage, with temple construction being a strong sign of this
- other polities also sought independence: ‘It was a period of chronic warfare resolved, albeit partially, with the establishment of the great state centred at Angkor in the early ninth century.’

(Higham 2002: 297)

Angkor

The polity of Angkor, centred in present day Cambodia, showed an unprecedented degree of centralization, rooted in Indic religious imagery and practice (Coedès 1968, Audric 1972, Chandler 1992, Higham 2001). Prior to the establishment of Angkor in the early ninth century, ‘there was a multiplicity of polities in the lower Mekong basin, relatively small and transitory realms representing the personal charisma and power of particular individuals rather than institutionalized political systems’ (Taylor 1992: 158). Khmer life had already been going on in the interior, but Angkor provided a new centre. Angkor was well situated for paddy rice fields, surrounded by expansive plains. This provided the rice agriculture that supported larger populations, who could be taxed and conscripted to increase the wealth and power of rulers.

While the construction of grand architectural creations of the Angkor temple complexes attracts our attention as symbols of power, other achievements are equally remarkable (Hall 1992: 230). The kingdom constructed public works – dams and channels – that irrigated some five million hectares (Hall 1992: 230). In a system of ‘theocratic hydraulics’ (van Liere 1980), temples acquired land through donations that would in turn bring religious merit to villagers. ‘What was transferred by donors was not “ownership” of land, but the right to income from land’ (Hall 1992: 232). In this and other ways, Khmer elites exercised a strong form of centralized political and economic control over their subjects (Hall 1992: 229–40). ‘Control over labour and production rather than ownership of land was critical for the development of early Southeast Asian states’ (Hall 1992: 232).

The religious hierarchical society that developed at Angkor was decidedly Indic in its cultural and religious orientation – witness the imagery of the famous temple complexes. At its height, Angkor maintained diplomatic relations with courts of China, and other kingdoms in Southeast Asia. Relations were particularly fraught with the Cham to

the east, and conflicts broke out in the eleventh and twelfth centuries. Over time Angkor expanded to become the dominant power in MSEA, ruling Champa, the Malay kingdoms, and territories to the north and northwest, as far as present day Vientiane and Sukhothai.

Tai Polities

Towards the end of this period of Khmer domination, not only did Champa remain a power to the east, firmly resisting Angkorean control, but Tai speakers gained in strength and numbers to the north, migrating from present day southern China. ‘In the thirteenth century, the outer edge of authority enjoyed by Angkor and Pagan recoiled from the ambitions of Tai speaking chieftains’ (Taylor 1992: 168). During the last centuries of the first millennium CE, Tai speakers steadily dispersed south and southwest from southern China. By the twelfth century, Tai leaders had established centres of power in the valleys of the upper Mekong and the northern tributaries of the Chaophraya, where they provided mercenary recruits for lowland rulers.

Through the course of the fourteenth and fifteenth centuries, in central Thailand, the Khmer/Angkorean polity of Sukhothai was transformed into the Tai kingdom of Ayutthaya. This outcome derived from long term intensive interaction between Southwestern Tai speaking and Khmer speaking peoples, and it provided the context for a sustained process of ethnolinguistic shift on the part of Austroasiatic populations towards Tai languages and identities. Ayutthaya had the kind of ethnic plurality and diversity that we associate with modern mainland Southeast Asia. The founding of Ayutthaya ‘is generally understood as a bringing together of the Angkorean style administrative skills of the Mons and Khmers of Lopburi, the manpower and the martial skills of the Tais of Suphanburi, and the wealth and commercial skills of the local Chinese merchant communities’ (Taylor 1992: 170).

While these changes were happening in the area of current day Central Thailand, the middle Mekong was a site of another Tai power that would be centred on Luang Prabang and Vientiane. In these ways, Tai speaking peoples laid the foundations for their power in the modern nations of Thailand and Laos. Tai groups from the north brought a formidable mix of know how, ambition, and dynamic organization to create powerful kingdoms dominating MSEA. As Taylor (1992: 169) put it, Tai leaders ensured a lasting legacy in MSEA by producing ‘vigorous responses to the opportunities of their age’.

Decline of Angkor

Angkor saw a slow decline, and by the mid fifteenth century its people had ultimately abandoned Hindu beliefs for the Theravāda Buddhism of the polities of the Chaophraya basin. ‘When the formation of Thai polities in the Chaophraya and Mekong basins posed

a higher level of threat to Angkor in the thirteenth century, the Khmers adapted with fundamental economic and cultural changes; they did not respond as if they had irrevocably invested themselves in any particular political or cultural heritage' (Taylor 1992: 162). The Austroasiatic speaking Khmers underwent thoroughgoing cultural adaptation and change, becoming fully intermingled within the culture and control of new Tai authorities (implying, similarly, that the Tai speakers of the Chaophraya basin were similarly reoriented by intensive adaption to Khmer; see Huffman 1973, citing Maspéro 1952: 583, on the linguistic effects on Thai and Khmer of this long term contact situation). This was not so much in response to direct military pressure from the Tai as to a more complex set of factors that led people to change their ways of life.

Modern narratives of Thai and Khmer history strive to distinguish the origins and identities of these nations. The Khmer, for instance, appropriate Angkor as central to their history. But this is based upon 'relatively recent reconstructions of the past': 'Between the days of Angkor and the twentieth century were generations of Khmers for whom Angkor, if not unknown, was nonetheless without the significance now attached to it' (Taylor 1992: 163). The truth is that the people who speak Khmer and Thai today share deeply intertwined histories. This is attested by heavy admixture in all aspects of their being, including language, cultural practice, and human genetics.

Pyū

Further to the west, a similar overall pattern of historical development took place, with early states rising and falling in the first millennium CE or soon after, then being overtaken by powers from further north, who would go on to dominate the area until the modern period.

People associated with the term Pyū moved into the upper Irrawaddy Valley from current day Yunnan in around the second century BCE and would rule for more than a thousand years (Win 2015: 1, Aung Thwin 2005: 13, Moore 2009: 108):

[B]etween the second century BC and the late ninth century AD, much of the country known currently as Myanmar was dominated, or at least occupied, by people conventionally known as the Pyū. Its heartland was the plains areas carved out by the confluence of the Irrawaddy and Chindwin rivers, a Y shaped configuration located in the Dry Zone of Upper Burma that has been inhabited since the Paleolithic Age.

(Aung Thwin 2005: 13)

The Pyū conducted trade and cultural relations both with South Asia (which as epigraphy reveals had a strong influence in terms of religion and culture) and with China. It

is largely assumed that people of the Pyū polity spoke a Tibeto Burman language and practised Buddhism of the Sarvāstivādin school.

While a ‘Pyū culture’ is taken to have existed in the upstream part of the Irrawaddy valley, the truth is as we find for other so called kingdoms, polities, and states across MSEA: the Pyū may be more accurately thought of as a loose network of villages, cities, and kingdoms with great internal diversity. The villages of the Pyū ‘were distinct or competitive enough to label them separate ethno cultural spheres’, and so we cannot assume ‘ethnic and linguistic homogeneity’ (Moore 2009: 115). Pyū groups were united only in so far as they were ‘forerunners of the Tibeto Burman movement into the southern parts of the Irrawaddy valley’ (Blagden 1919: 61). As with other high level ethnolinguistic and geopolitical terms, Pyū is a term of convenience (see Moore 2009: 104, citing Luce 1985: 70 1 and Blagden 1919: 61). According to Moore (2009: 107), the idea of Pyū culture is ‘hazardously general’. The evidence for Pyū unity is drawn from fragmentary and distributed evidence of social and political organization: an assemblage of fingermarked bricks, brick walls, gates, ramparts, urns, coins, beads, pottery, iron artefacts, and inscriptions.

Dvāravatī and Mon

To the south and east of the area occupied by the Pyū, and over a partially overlapping time period, was a polity known as Dvāravatī. The term Dvāravatī must be used with the same caveats as all of the ethnic/political terms we have so far encountered: it ‘can be employed to describe a time period, an art style, a material culture and/or a polity or polities in what is today central and northeast Thailand’ (Murphy 2016: 367). The Dvāravatī period spans from as early as the fourth century CE to as late as the eleventh century CE, though there is controversy about the exact dates (see Murphy 2016: 381, also Barram and Glover 2008, Murphy 2013). What is clear is that Dvāravatī played an important role in the spread of Buddhism eastwards from around the sixth to the eleventh century CE.

Largely overlapping in time and adjacent in space to the Dvāravatī, the Mon entered present day lower Myanmar as early as the sixth century CE (Tun 2015: 18, Moore 2009). Mons spoke Austroasiatic language varieties. They were Theravāda Buddhists whose prosperity derived from sea oriented trade, like other first millennium polities on the coasts of MSEA. A paradigmatic Mon polity has been widely assumed in scholarship on the history of this area. It carries with it the familiar assumptions in MSEA historiography of a relative homogeneity in ethnolinguistic make up of each kingdom or state. As with the Pyū case, the idea of a unified Mon identity is a product of ‘reified ethnicity’ inherent in politicised discourse and MSEA history making (Aung Thwin 2005: 13, Moore 2009). This relates to the question of demarcating the concept of Dvāravatī: ‘Dvāravatī may have been no more than a specific Mon kingdom with its heartland in central Thailand and

a Buddhist culture quite distinct from that of any of its neighbours, or it may be part of a wider culture that included all the Mon lands of Burma and Thailand and even extended into pre Angkor Cambodia' (Hall 1992: 294).

Pagan and the Burmans

Pyū and Mon peoples had lived in the Irrawaddy Basin for several centuries by the mid ninth century, when Burmans first entered the area. The religious milieu of the Sarvāstivādin Buddhist Pyū and the Theravāda Buddhist Mon provided a focused cultural background for the development of the Kingdom of Pagan. Incoming Burmans Tibeto Burman speakers from the north then established the Theravāda Buddhist polity of Pagan from the mid eleventh century as a centre of power in northern Myanmar. Meanwhile, the Mon in the south of Burma controlled maritime contact with Sri Lanka and India, absorbing influences that were both economic and cultural/religious in nature. These connections with India helped to maintain both continuing borrowing from Indic languages and religious practices. Pagan and the Mon polities had intense cultural interactions including both conflict and cooperation. Taylor (1992: 166) discusses the situation in the late eleventh to early twelfth century period:

The Mons appear to have dominated the religious and intellectual life of Pagan at this time, and their language was widely used among ruling class people. The language of the Pyū continued to be a cultural force as well, seemingly as the repository for the legacy of the Pyū cities of earlier times. Pali became the language of scripture and liturgy as the Theravāda monkhood flourished under royal patronage. The Burmans themselves learned to write their language in a Mon script. Their initial contribution was, of course, military, for it was their battlefield prowess that had made of Pagan an assembly of such diverse cultural elements; but the story of Pagan eventually is concerned with how the Burmans assimilated these elements and went on to establish a Burman cultural tradition.

(Taylor 1992: 166)

Four languages Burman, Mon, Pāli and Pyū were used together in Pagan. 'A fitting symbol for Kyanzitha [Pagan ruler between 1084 and 1113] is the Myazedi pillar, erected near the end of his reign, with identical inscriptions in [all] four languages' (Taylor 1992: 166):

The cities of the Pyus, the Mons, and the Burmans shared a single lowland geo strategic site, were vulnerable to mountain based powers, were in close land contact with the borderlands of the Indian world (the Arakanese and Bengali coasts and the Assam basin), and were in regular

maritime contact with the eastern coast of the Indian subcontinent and with Sri Lanka. The rulers of Pagan who aspired to unite these cities under their authority from the eleventh century claimed a superior measure of merit as defined by the religious ideas that united the different ethnocultural (and socio economic) patterns of that time and place.

(Taylor 1992: 164)

Pagan, like Angkor, was an inland agrarian polity which was able to unite people and enforce its authority over coastal areas. On a similar arc and timeline to Angkor, Pagan remained a powerful kingdom from the mid ninth century until it was fatally weakened by Mongol invasions of the mid thirteenth century.

Hmong Mien Migrations

Hmong Mien groups have long resided in a broad area of southern China. Their earlier origin is unknown. In the period between the mid seventeenth to the mid nineteenth centuries, during the Qing Dynasty, there was an extraordinary increase in population in China, from one hundred million to more than three hundred million (Culas and Michaud 2004: 64). This precipitated a southward and westward expansion by Han peoples, enabled in part by the adoption of new, more flexible forms of agriculture, notably maize from the Americas. Upland minority groups experienced major incursions on their territory:

Together with excessive taxation, this invasion constituted one of the most important causes of conflict with highland minorities trying to preserve their integrity vis à vis the Han centralized administration and regional rulers and warlords. Some minorities responded by simply migrating further south or into the higher mountain ranges of Guizhou, Sichuan, and Yunnan. Others took up arms. A small number went as far as leaving the Chinese Empire altogether to find shelter in the unpopulated ranges which were later to be incorporated into the modern states of the Indochinese Peninsula.

(Culas and Michaud 2004: 64)

Suppression of a series of anti Qing rebellions led significant numbers of Hmong Mien peoples especially those speaking languages of the Hmong branch to flee from Chinese territory to their current distribution across the north of MSEA (see Jenks 1994).

1.4.5 Summary: Fall of First Millennium States, Rise of Second Millennium States

A broad trend in the history of MSEA over the last two thousand years is captured by O'Connor (1995: 971) in his comparison between the overall situation in the first and second millennia CE:

In the early era Pyū, Mon, Khmer, and Cham rule the mainland, while to their north are Burmese, Tai, and Vietnamese. Ten centuries later the northerners dominate everywhere. In irrigable niches the southerners are gone (Pyū) or in decline (Cham), and their remaining strength is in the unirrigable lower deltas where Mon and Khmer flood farmers live in states under northern overlords.

(O'Connor 1995: 971)

Figure 1.2 shows the timeline comparison between dominant states in the first millennium versus the second millennium CE.

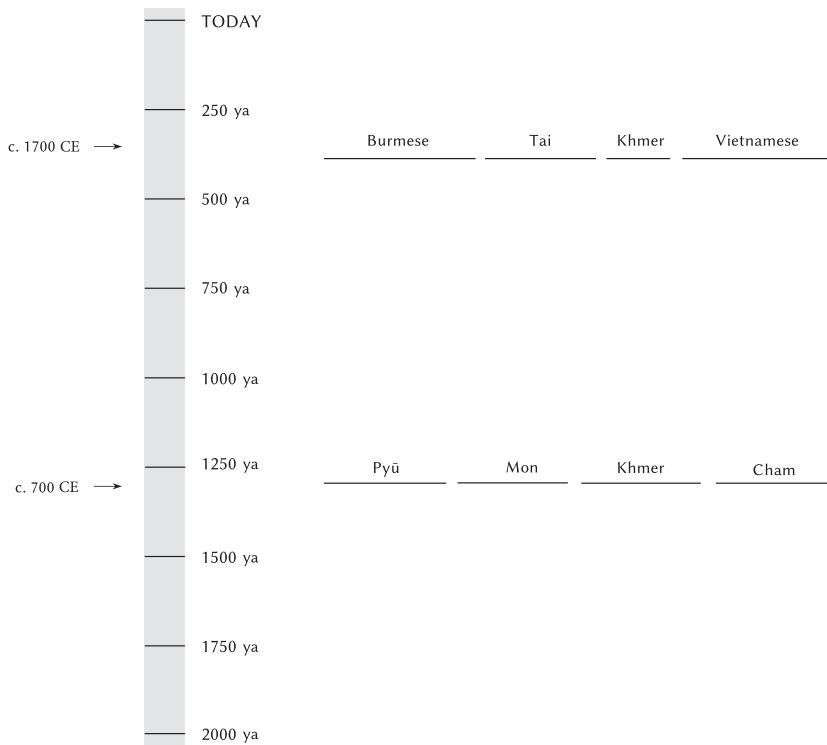


Figure 1.2 ‘Schematic depiction of succession of dominant states’

From O'Connor 1995: 971. The figure compares the dominant polities of MSEA, arrayed from west to east, at 700 CE and at 1700 CE. It is important to note that these ‘states’ were not states as we know them, and they were ethnolinguistically highly diverse. That said, there were undoubtedly asymmetrical relations between the languages spoken within states, as administrative/prestigious languages must have been understood by broader swathes of the population than other languages.

O'Connor's analysis foregrounds the region wide importance of agricultural innovation in this process:

In the first millennium CE mainland Southeast Asia's first great states arise, but then in the span of a few centuries these Indianized realms collapse and their Pyū, Mon, Khmer, and Cham peoples decline. In their place Burmese, Tai, and Vietnamese states arise and go on to rule the mainland as their peoples come to dominate the second millennium. Case by case these shifts appear to be ethnic and political successions wherein the strong displace the weak, but seen together regionally the similarities suggest an agricultural change whereby an irrigated wet rice specialization from upland valleys displaced gardening and farming complexes native to the lowlands.

(O'Connor 1995: 968)

These developments directly affected, and were directly affected by, the motivations and experiences of the individuals who were making their livings along the rivers and hillsides of MSEA:

Once states appeared, adaptive conditions changed yet again at least for farmers. At that moment, mobility allowed farmers to escape the impositions of states and their wars. I call this tertiary dispersion. The other two revolutions agriculture and complex society were secure but the state's domination of its peasantry was not, and so we find a strategy of 'collecting people ... and establishing villages'.

(O'Connor 2003: 281 2, cited in Scott 2009: 4)

Though these periods of major transition in the lowlands may have been driven by agricultural innovation, the extension of agricultural settlement probably also reflects the increasingly widespread use of iron (cf. Bellwood 1992: 120). The resulting displacement of ethnic groups using earlier technologies suggests that the distribution of 'hill tribes' in mainland Southeast Asia the highly diverse swathe of minority ethnolinguistic groups along the higher altitudes running from eastern Cambodia up along the Annamite Cordillera to northeast India and beyond is likely to be relatively recent. This line of thought led to Scott's Zomian hypothesis.

1.4.6 Away from States: Zomia

The upland areas in which many MSEA minorities live are conjoined in a single, elongated area, crossing political borders and encompassing 'virtually all the lands at altitudes above roughly three hundred meters all the way from the Central Highlands of

Vietnam to northeastern India' (Scott 2009: ix). This area has been referred to as *Zomia*, a term coined by van Schendel (2002) in making the point that arbitrary research areas can be constructed and reified by 'academic politics' (cf. Michaud 2010). Van Schendel's proposal of a Zomia area is a conceptual exercise, useful because it counteracts the politically sanctioned alternative narratives of centralized states, and so helps to combat modern nation oriented bias in our understanding of MSEA. The term has gained some recognition, particularly due to Scott (2009) (though ironically not without danger of creating the kind of reification it was warning against; Jonsson 2011, 2014).

According to Scott, it is not that the inhabitants of Zomia simply share the fate of having been marginalized by states. Instead, he argues, they share a cultural distaste for being governed: they have chosen to remain isolated from central government control. Scott's thesis puts agency in the hands of the marginalized peoples. There is some controversy as to whether Scott's analysis is correct, or indeed whether the Zomia concept is useful. But it does mount a much needed challenge to many of our most recalcitrant biases and assumptions about the relative status and role of the many and varied ethnolinguistic groups encountered across MSEA. After all, the vast majority of ethnolinguistic groups across MSEA are Zomian, and so Scott's ideas are a vital corrective to an incorrigible tendency to background, overlook, and misconstrue those groups. It is worth quoting him at length:

A wealthy and peaceful state center might attract a growing population that found its advantages rewarding. This, of course, fits the standard civilizational narrative of rude barbarians mesmerized by the prosperity made possible by the king's peace and justice – a narrative shared by most of the world's salvational religions, not to mention Thomas Hobbes.

This narrative ignores two capital facts. First, as we have noted, it appears that much, if not most, of the population of the early states was unfree; they were subjects under duress. The second fact, most inconvenient for the standard narrative of civilization, is that it was very common for state subjects to run away. Living within the state meant, virtually by definition, taxes, conscription, corvée labor, and, for most, a condition of servitude; these conditions were at the core of the state's strategic and military advantages. When these burdens became overwhelming, subjects moved with alacrity to the periphery or to another state.

(Scott 2009: 7)

Scott's idea is that continual 'state making' in MSEA created 'a zone of refuge or "shatter zone," where the human shards of state formation and rivalry accumulated willy nilly' (Scott 2009: 7). By this account, Zomians are not merely fleeing, but choosing a preferred way of life:

Their subsistence routines, their social organization, their physical dispersal, and many elements of their culture, far from being the archaic traits of a people left behind, are purposefully crafted both to thwart incorporation into nearby states and to minimize the likelihood that statelike concentrations of power will arise among them. State evasion and state prevention permeate their practices and, often, their ideology as well. They are, in other words, a “state effect.” They are “barbarians by design.” They continue to conduct a brisk and mutually advantageous trade with lowland centers while steering clear of being politically captured.

(Scott 2009: 8)

By far the majority of languages spoken in MSEA today are spoken by so called Zomian peoples. Accordingly, in this book I try to reflect this by focusing as far as possible on these languages wherever data permit.

1.4.7 Dynamics of Ethnolinguistic Diversity in MSEA

Whether or not Scott is correct in his analysis of upland peoples’ historical relation to states, what we do know is that present day upland minorities are where they are because incoming peoples with advantages of technology, know how, and sheer numbers have historically taken possession of lowlands, for agriculture in particular. In the expansive plains of northeast Thailand – the epicentre of core MSEA – it is clear that multiple migrations implied successive processes of displacement, dispossession, and/or assimilation.

Earliest inhabitants, such as the Hoabinhian foragers, were displaced or assimilated by groups associated with the initial introduction and adoption of agriculture, and with early lowland states (Pyū, Mon, Khmer, Cham). These incomers were speakers of Tibeto Burman, Austroasiatic, and Austronesian languages. In turn, those groups were further displaced or assimilated by incoming high intensity paddy rice agriculturists, most notably the Tai speakers. As Bellwood (1992: 109) remarks, ‘it is quite apparent that a once continuous distribution of Austroasiatic languages over most of mainland Southeast Asia, and even across into the Nicobars and possibly northern Sumatra, has been broken up by the historical expansions of the Chinese, Tai, Vietnamese, Burman and Austronesian (Malay and Cham) peoples’. More recently, there has been wave upon wave of migration into and within MSEA, including rural migrations (e.g., Hmong Mien speakers as refugees from the north over the last two centuries; Jenks 1994), urban migrations (e.g., Chinese and South Asian business and labour in the urban centres), and movements engineered both by colonial and nationalist powers (e.g., Vietnamese moving to Laos as administrators during the French Indochina

period 1887–1954) and by the dominant ethnic majority governments after postcolonial national independence (e.g., movements of Burmans to the minority dominant north of Myanmar after independence in 1948). Figure 1.3 charts these developments.

These historical ethnolinguistic developments imply repeated processes in which one group displaces, dispossesses, and/or assimilates another. But the process was surely more gradual and shaded than a two sides account might suggest. Consider the

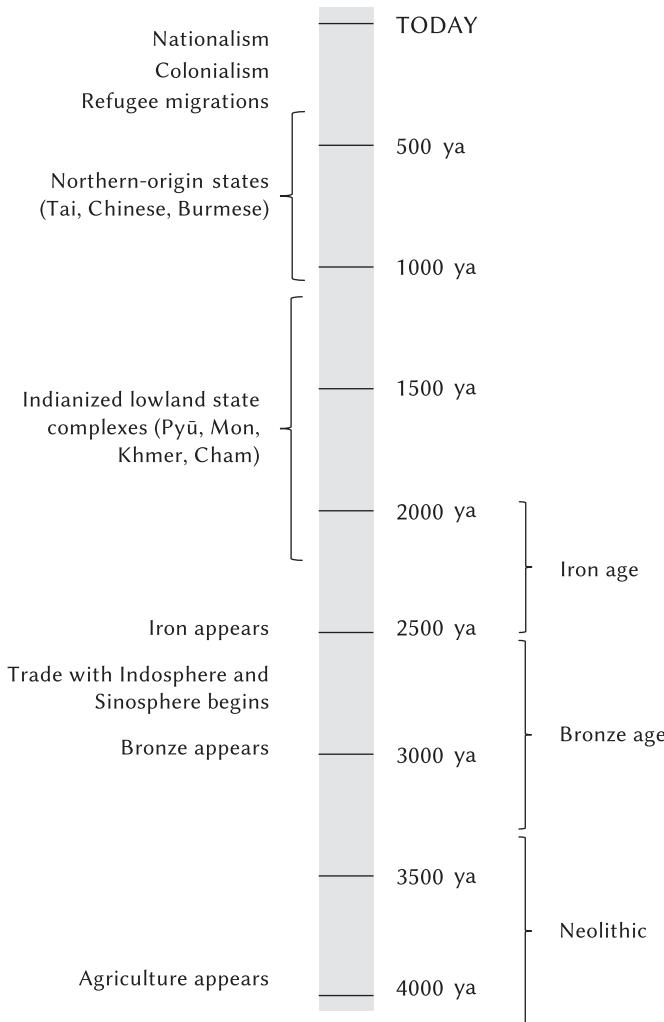


Figure 1.3 *Chronology: Mainland Southeast Asia over the last 4,000 years*

historical transition from Pyū to Burman in the Irrawaddy Basin between the first and second millennia CE. When the Pyū kingdom suffered devastating raids by the Nan Chao kingdom from China, in 832 and 835 CE, thousands of people were imprisoned and/or taken away as labour. On one view, this created a vacuum which Burmese speaking Burmans (either waiting in the wings, further north, or entering together with Nan Chao forces) could fill, precipitating Burmese ascendancy. But it is equally possible that there was ‘an indigenous, historical continuity between Pyū and the Burmese speakers irrespective of the Nan Chao raid’ (Aung Thwin 2005: 37). This is consistent with the idea that the ‘Burmese’ were in fact part of an ethnically diverse world, in a way that is quite familiar from our observation of human groups in MSEA today, and from other historical accounts, such as the situation of ethnolinguistic pluralism that Leach described for ‘Shan’ areas in upland Burma in the 1950s:

Nearly all the Shan states within the Kachin Hills Area include non Shan elements of population. In a good many cases the non Shan elements are considerably more numerous than the Shan elements. The political capital of a Shan state is in all cases a Shan township located in the vicinity of irrigated rice lands, but the feudal dependencies of such a state may include not only other communities of wet rice growing Shans, but also a variety of hill villages with a non Shan population.

(Leach 1954: 35)

But the on the ground nuances of graded ethnolinguistic reality do not preclude sudden, intense, or violent disruptions to life. The complex historical processes that have created the current distribution of human groups and livelihoods in MSEA have involved sometimes radical transformations, with upheavals in lifestyles, practices, languages, locations, and modes of interaction. These transformations were not always without conflict. Bellwood (1992: 109) states that the prehistoric sites of northeast Thailand ‘were inhabited by speakers of Austroasiatic languages which eventually fell victim to the assimilatory tendencies of the historical Thai kingdoms after the thirteenth century CE’. His phrasing suggests harsh realities of dominance and hegemony between indigenous groups throughout the history of the area. We are of course aware of the massively disruptive and often violent exercising of European economic and political power in MSEA over the last two centuries. But life for many centuries before that was also prone to incursion, land acquisition, replacement, and conflict. Not all contact events between ethnic groups are peaceful or fair, particularly when they are brought about by migration in search of valuable resources and improved livelihoods. Such migration has been the main driver of social and thus linguistic transformation and change in MSEA for at least four thousand years.

Historical relations between ethnolinguistic groups in MSEA may have ranged from asymmetrical, hegemonic, and violent relations at one end, to symmetrical, productive, and peaceful relations at the other. Which type of relations are to be inferred in any given case may depend in part on which of two models of the diffusion of language and culture one foregrounds: (1) demic diffusion or (2) cultural diffusion. Those who argue that modern ethnolinguistic diversification in MSEA was associated with demic diffusion⁸ (Bellwood 1992; Blust 1994; Higham 2002; Edmondson and Gregerson 2007) emphasize the incoming migration of groups of people who rely on agriculture. With new livelihoods, which could support large populations, incoming groups physically replaced less populous and less powerful existing forager populations (Ammerman and Cavalli Sforza 1971, Cavalli Sforza *et al.* 1993, Nichols 1992, Kutanan *et al.* 2017). Genetic evidence is offered to support this account. The genetic diversity among speakers of the later arriving Tai languages in Thailand and Laos is lower than the genetic diversity of speakers of Austroasiatic languages in the same region, consistent with the hypothesis that human migration was the main carrier of Tai languages into the area (Kutanan *et al.* 2017).

A cultural diffusion account is not incompatible with demic diffusion. When people move to new areas, they bring not only their genes. They also bring their cultural practices and their languages, which can be rapidly adopted by others within their lifetimes. Resident populations may remain in place but may take on new practices and ways of speaking. O'Connor (1995: 987) argues that 'there is no direct evidence that an actual influx of immigrants ever displaced earlier peoples' in MSEA. He argues instead that an agricultural paradigm is what diffused, alongside migrations that were piecemeal and not necessarily replacive. Certain know how and practices spread from group to group, offering more productive livelihoods, and bringing with them a 'society shaping complex' (see Jonsson 2011, 2014 for discussion). Other critiques of a simple demic diffusion account of human history in MSEA include White (2011), who argues that hunter gatherer communities have played a central role in shaping modern MSEA ethnographic diversity, and Fix (2011), who examines the genetics of ethnolinguistic diversification. Fix presents an alternative to the standard account of demic diffusion in the Malay Peninsula, with a model he calls trickle effect colonization.

⁸ Demic diffusion is the spread of genes. It is usually associated with the outcomes of migration. In world history, this has often involved the movement of groups who have adopted agriculture, and who are therefore more populous and viable than those (e.g., hunter-gatherers) who are resident in the area being entered. Demic diffusion may be associated with population displacement or replacement, but this need not necessarily be the case. There may be genetic admixture between an incoming population and a resident population, such that some proportion of the genes of the resident population survives. Thanks to Mark Stoneking and Dan Dediu for clarification of these points.

It is clear that the historical process of peopling and ethnolinguistic diversification in MSEA was driven both by the spread of people and by the spread of ideas and practices. The exact nature and balance of these forces through history is unknown. But the modern distribution of ethnolinguistic groups is clear. In lowland areas, populations are relatively large and dense, culturally and linguistically homogeneous, and closely affiliated with state political power. In upland areas, populations are relatively small and sparse, culturally and linguistically diverse, and have limited access to infrastructure, education, or political power. The dominant lowland populations of MSEA are clearly distinct from each other in terms of political identity: the Thai vs. the Lao vs. the Khmer vs. the Burmans vs. the Vietnamese. These groups have certain things in common, one being their demographic and political dominance. Similarly, the upland minority populations that straddle these nations have something in common: they are politically and geographically marginalized.

When we focus in more closely on this broad stroke analysis, we find that within ‘subordinate upland populations’ there are inter group asymmetries and dominance relations in microcosm. Taken synchronically, few ethnolinguistic groups are macro dominant, and few are truly subaltern. Most ethnolinguistic groups of MSEA are politically interstitial. In certain frames they dominate over other groups, while in other frames they are dominated.

The dynamic developmental periods of the rise and fall of sophisticated complexes of state like entities throughout MSEA in the first millennium CE and beyond interfaced directly with the Indosphere to the west and the Sinosphere to the north, both of which would ultimately have significant impacts on the linguistic situation in the area (Matisoff 1991c: 485ff.). This laid the foundation for increasing centralization of political power whose outcomes would eventually combine with the forces and effects of nineteenth century colonization (Tarling 1992a, Ileto 1992, Stockwell 1992) and twentieth century nationalism (Kratoska and Batson 1992, Turnbull 1992, Tarling 2004) to yield the exceedingly complex and multi layered sociopolitical situations we observe today.

1.5 On Modern Dominant Languages

The processes of in migration and ethnolinguistic diversification described in the last section likely brought peak linguistic diversity to the MSEA area by around 250 years ago. The beginnings of colonialism, nationalism, and globalization have since brought in new, more radical, and much swifter transformative processes of social change (see Tarling 1992b for background). Given our interest in the linguistic diversity of MSEA, these most recent developments are relevant in so far as they have initiated intensive and

pervasive processes of language disruption, shift, and loss across the area, affecting the languages of minority peoples in particular.

Within the last century, a tiny minority of the languages of MSEA have gone in the opposite direction: Burmese, Thai, Lao, Vietnamese, Khmer, and Mandarin Chinese have attained national language status and have thereby become as strong and secure as it is possible for a language to become. These languages are protected and promoted by governments, in education, media, and law. They have become instruments of state political power. And unlike the dozens of other languages spoken within the borders of their respective nations, they are afforded state protection.

The stories of each national language differ in significant and important ways. But the broad outlines are the same. We can illustrate with the example of Burmese, the national language of Myanmar (Watkins 2007). With colonization of Burma by the British – in pieces, formally beginning in 1836 – the language of the colonizers came to be used in education, linking an elite class with the English language, and hence the opportunities that this elite class would enjoy. A nationalist movement emerged in the late colonial period, led by these western educated Burmese, who ‘increasingly came to challenge colonial rule and call for independence for the country’ (Watkins 2007: 270). This nationalist movement helped to achieve Burma’s full formal separation from Britain in 1948.

Three properties of the movement are worth noting (Church 2003: 114–15, cited by Watkins 2007: 270). The first is that it was led and dominated by speakers of the numerically dominant language, Burmese. They elevated the (written) Burmese language and Burmese speaking literature and culture as national symbols. Second, the movement was heavily critical of non Burmese ‘foreigners’, including those within Burma (e.g., urban Chinese and Indian influence on the country’s economy, non Buddhist communities such as the Rohingya, giving rise to ‘anti foreigner’ sentiment). Third, it elevated Burman culture, and especially the Buddhist religion, to the exclusion of other cultures and religions, including Christianity, Islam, and forms of animism. These three aspects of the status of the Burmese language in Burmese nationalism served not only to help focus and galvanize efforts to achieve political independence from colonial rule, they also helped to quash definitively the hopes of many indigenous groups of having independence over their own affairs.

In the wake of Burmese independence in 1948, the Burmese language became a focus of intense interest in political symbolism. The (unnamed) author of a 1948 Burmese language dictionary reported ‘feeling compelled to create the dictionary for fear that the Burmese language and with it the culture of the Burmans might be swamped by foreign dominance’ (Watkins 2007: 272). The main perceived threat was English. But another threat to Burmese as a national language was the instability of the nation, given that the country is made up of so many large and distinct ethnolinguistic groups.

‘One of the provisions of the new constitution therefore declared that the official language of the Union of Burma should be Burmese’ (Watkins 2007: 272). The post colonial phase was marked by a strong drive towards ethnolinguistic homogenization. The colonial language English was suppressed, as were the many dozens of non Burmese languages spoken natively by ethnic minorities throughout the country. The resulting situation is as today: the Burmese language dominates in national affairs and in official discourse. Burmese is supported by a powerful state apparatus that controls language in education and media.

Minority languages of Burma are mostly unwritten and are not supported by the state. But some of the languages are powerful, both in terms of sheer numbers Karen varieties, for example, are spoken by some seven million people in Myanmar alone and in terms of political organization. Karen peoples have long pursued independence within Myanmar. The Karen National Union and related entities such as the Karen National Progressive Party not only include people who are ethnically and linguistically Karen, but also members of numerous other ethnic groups. By ‘trying to mould a unified Karen identity’ from a diverse ethnic mix, the Karen National Progressive Party is creating ‘essentially a microcosm of the situation in Burma/Myanmar generally, using “policies in education to strengthen the designated national language and to de emphasize languages which pose a threat to national goals”’ (Watkins 2007, citing Sproat 2005 and Dudley 2000).

Two aspects of this story are worth highlighting. The first is that non Karen minority groups may be subsumed under a Karen identity, thus becoming in a sense doubly subordinated: both to Karen identity and to Burman identity. The second point is that Karen peoples speak mutually unintelligible language varieties:

Significantly, no Karenic lingua franca has emerged that can be used throughout the wider Karen area, creating an obstacle to any language assisted furtherance of nationalism. An ironic result of this is that the Karen National Union (KNU), the force which has been fighting the Burmese army for half a century, uses Burmese for its formal meetings rather than any of the Karen languages.

(Watkins 2007: 284)

This is evidence that the Burmese nationalists’ drive to spread Burmese as part of its nationalist agenda ‘has not triggered the acceptance of a Burman/Bamar centred national identity and a willingness to be part of a state with such an identity’. It also demonstrates that ‘nationalist organizations may function effectively using a range of languages including a “foreign” language, even when the “foreign” language is that of a political and military opponent’ (Watkins 2007: 284).

This sketch of the postcolonial rise to dominance of the Burmese language in Myanmar has broad parallels with the rise of Mandarin Chinese, Khmer, Lao, Thai, and Vietnamese in their respective countries, though the details of course differ significantly. For reviews and references on modern national language histories in MSEA, see Chen (2007) on China, Heder (2007) on Cambodia, Simpson and Thammasathien (2007) on Thailand, Enfield (1999) on Laos, and Lê and O'Harrow (2007) on Vietnam.

MSEA languages have prevailed over European languages in the postcolonial era. But the promotion of a small handful of languages including Burmese, Khmer, Lao, Thai, and Vietnamese as national languages has demoted, ignored, and sidelined the vast majority of languages spoken in MSEA, and thus almost all of the languages that are the focus of this book. This is one reason I try not to foreground major languages of MSEA here, given that they are no more important than minority languages to the broad questions of linguistic typology, and that they are vastly over represented in discourse about and research into languages of the area to date.

1.6 Resources and Developments in MSEA Linguistics

1.6.1 Conferences and Publications

The South East Asia Linguistic Society (SEALS) was founded by Martha Ratliff and Eric Schiller at Wayne State University, Detroit, in 1990. SEALS was due to hold its thirtieth annual meeting in 2020. Prior to 2009, proceedings of SEALS meetings were published in edited volumes. Since then they have appeared in the open access *Journal of the Southeast Asian Linguistics Society*.⁹

The SEALang Projects website (www.sealang.net; see especially the *Southeast Asian Linguistics Archive* www.sealang.net/sala) makes accessible a range of primary and secondary sources on MSEA languages. Other regular publishing venues for research on MSEA languages include the journals *Mon Khmer Studies* (an open access journal, see www.mksjournal.org) and *Linguistics of the Tibeto Burman Area* (see <http://sealang.net/sala/ltna/htm/index.htm>).

Some recent interdisciplinary explorations of the history of ethnolinguistic diversification have focused on languages of MSEA and neighbouring places (e.g., Sagart *et al.* 2005, Enfield 2011a). The last ten years have seen the publication of multiple landmark overviews of MSEA language families, including Tai Kadai (Diller *et al.* 2008), Sino Tibetan (Thurgood and LaPolla 2003, 2017, cf. Matisoff 2003a), Austroasiatic (Jenny and Sidwell 2015, cf. Shorto 2006), and the Austronesian languages of MSEA (Thurgood 1999, Grant and Sidwell 2005, Larish 2005, Blust 2013b: 70–5). I draw on

⁹ Accessible here: <https://evols.library.hawaii.edu/handle/10524/52368>

these works in my thumbnail sketches of language history and subgrouping in the next chapter.

1.6.2 New Descriptive Work

A key measure of progress in an area is the production of reference materials based on new empirical research.¹⁰ Full sized descriptions of MSEA languages published since the turn of the century include grammars of Semelai (Kruspe 2004), Jahai (Burenhult 2005), Garo (Burling 2004), Deuri (Jacquesson 2004), Mongsen Ao (Coupe 2007), Lao (Enfield 2007), Anong (Sun and Liu 2009), Hainan Cham (Thurgood *et al.* 2014), Turung (Morey 2010), the Tai languages of Assam (Morey 2005), Lisu (Yu 2007), Thai (Higbie and Thinsan 2003, Iwasaki and Ingkaphirom Horie 2005), and Cambodian (Haiman 2011). Numerous grammars have been completed as doctoral dissertations. In the area of northeast India, for example, see grammars of Galo (Post 2007), Atong (van Breugel 2014), and Karbi (Konnerth 2014). Sketches or partial descriptions have appeared of languages including Pacoh (Alves 2006), Kri (Enfield and Diffloth 2009), and Arem (Ferlus 2014), and detailed descriptions have appeared of specific domains of grammar such as phonetics/phonology; see, for example, Watkins (2002) on Wa and Coupe (2003) on Ao. Major dictionaries of minority languages are less abundant; notable examples are Suwanarat and Reilly (1986) on Thai Sign Language, Purnell (2012) on Iu Mien, Watkins (2013b) on Wa, and Svantesson *et al.* (2013) on Kammu Yùan. An important preoccupation of descriptive linguistics globally is the documentation of endangered languages; for excellent examples of new empirical work with this orientation in the MSEA context, see Morey (2005, 2010); see also Premsrirat (1998b, 2008).

A significant amount of new data and analysis from MSEA languages has become available on most if not all domains of interest to linguists, and on most if not all language families and sub areas of MSEA. As just one example, here I mention the Aslian languages of Peninsular Malaysia (Matisoff 2003). Since the turn of the century, we have seen the publication of typological overviews of the Aslian languages as a group (Matisoff 2003b), new reference grammars (Kruspe 2004, Burenhult 2005), other descriptive materials (Burenhult and Wegener 2009, Wnuk and Majid 2014, Wnuk 2016), new interdisciplinary research on the history and diversification of ethnolinguistic subgroups

¹⁰ Only a selection of those recent materials that have been published in English are mentioned here, though a substantial descriptive literature on MSEA languages is being published in other languages, including Chinese, French, Indonesian, Thai, and Vietnamese (for some examples, see: Bo 2002, Bon 2014, Buakaw 2012, Chen 2005, Gai 2002, Giaphong 2004, Kosaka 2000, D. Li 2003, 2004, Y. Li 2003, Lidz 2010, Mao and Li 2002, 2007, Mayuree 2006, Ploykaew 2001, Samarina 2011, Seng Mai 2012, Shee 2008, Shintani 2008, Srisakorn 2008, Wayesha 2010). Further chapters of this book draw on works written in Chinese (with thanks to Weijian Meng for locating and translating these materials).

(Burenhult *et al.* 2011, Dunn *et al.* 2011, Bulbeck 2011, Fix 2011, Oppenheimer 2011, Dunn *et al.* 2013), and field research on the psychological implications of semantic systems that are indigenous to Aslian languages and world views (Burenhult and Majid 2011, Majid and Burenhult 2014, Wnuk and Majid 2014). Not only is this breadth and depth of new work improving our basic understanding of MSEA languages and their socio historical contexts, it is also helping to balance our perspective on the MSEA area, with effects on our image of what a Southeast Asian language is typically like. The availability of new descriptive materials means that we can progress in the field by testing existing proposals and by continuously expanding the scope of our work (see Pittayaporn 2009 for a good illustration of this point).

1.6.3 New Methods

As new methods in linguistic research are developed and applied in linguistics globally, so they are developed and applied in mainland Southeast Asia. In phonetics and phonology, for example, new instrumental and computational technologies are rapidly transforming the realms of possibility in data collection and analysis, both by making new kinds of measurement possible, and by making the equipment smaller and more portable for fieldwork; see Edmondson and Esling (2006: 172–5) for the use of laryngoscopy to study the phonetics of breathy vocal register in Jianchuan Bai (spoken in Yunnan) and Brunelle (2009a) for the use of electroglottography to study register in Cham dialects in Vietnam (see also Brunelle *et al.* 2010 on Northern Vietnamese). Newly developed statistical techniques are being applied with interesting results: in historical linguistics, probability based bio informatic techniques are being used for exploring cladistic representations of language relatedness (see, for example, Burenhult *et al.* 2011); and in areal typology, statistical modelling is being used to test dependencies among phonological features, language history, and language contact (Brunelle and Kirby 2015). In lexical and grammatical work, new field methods are being applied in the exploration of semantic fields, in a range of functional and conceptual domains (see for example, Burenhult 2006, Wnuk and Majid 2014, Enfield 2015). There is an increasing interest in combining methods in order to further our knowledge of the area’s languages, for example in the interdisciplinary collaborations of historical work (Sagart *et al.* 2005, van Driem 2007, Enfield 2011a). And computational power is being exploited in building larger and better databases of, or including, MSEA languages (Dryer and Haspelmath 2013, Donohue *et al.* 2013).

Language Histories and Classifications

This chapter reviews current scholarship in the historical comparative linguistics of five major language families spoken in the MSEA area: Austroasiatic, Austronesian, Hmong Mien, Sino Tibetan, and Tai Kadai.¹ More extensive data and analysis may be found in recent book length treatments of each of the language families: see Jenny and Sidwell (2015) on Austroasiatic languages, Diller *et al.* (2008) on Tai Kadai languages, Thurgood and LaPolla (2017) on Sino Tibetan languages, Grant and Sidwell (2005) on mainland Austronesian Languages (Chamic and Moklenic), and Ratliff (2010) on Hmong Mien languages.

Research in historical comparative linguistics continues apace in MSEA. Advances are being made in classification and reconstruction in all the major language families. Old hypotheses are being tested with new data and techniques, and new hypotheses are being put forward. The appearance of new data, in particular, has made an important difference, enabling, for example, Pittayaporn (2009) to propose new reconstructions of the phonologies of Proto Tai and Proto Southwestern Tai respectively, Sidwell (2009b) to offer an improved account of vowels in Proto Mon Khmer, and Matisoff (2015b, 2019, cf. Matisoff 2003a) to re-examine the place of Jingpho within Tibeto Burman. In research on historical Hmong Mien, Ratliff (2010) has recently provided an assessment of previous work and offers substantial new reconstructions, with consideration of their implications. Historical Austroasiatic has seen substantial developments, including a suspension of the assumption of a highest level split between Munda and Mon Khmer. It is no longer widely assumed that ‘Mon Khmer’ languages represent descendants of a single ancestor language below Proto Austroasiatic (although the term Mon Khmer is still useful with the meaning ‘non Munda Austroasiatic languages’; for

¹ Languages of the Indo-European language family have played an important role in the area over the last few hundred years, associated with political and economic relations with European powers, including the formal colonization of most countries of the area. Today, English dominates everywhere, though French is supported through efforts such as those of the Alliance Française in countries of former Indochina.

a range of perspectives on this, see discussion in Sidwell and Blench 2011, Diffloth 2011, Sagart 2011, and Van Driem 2011). Similarly, in Sino Tibetan linguistics, old assumptions are being questioned. For example, recent reconsiderations of the position of Chinese in the family have assigned it to a lower level subgroup rather than the standard placement as a high major branch; more subgroups of Sino Tibetan are identified, and the time depth of reconstructed Proto Sinitic is pushed back to before Old Chinese (Blench and Post 2014, van Driem 2013, Sagart *et al.* 2019). In later sections of this chapter, I discuss some of these developments, with a focus on higher level claims and findings.

Let me register some caveats. First, any claim about historical relations among language varieties can never be more than a hypothesis. But second, hypotheses can be more or less strong. They are to be tested and revised or discarded. We evaluate and compare hypotheses with reference to evidence and argumentation. In MSEA linguistics, we have little access to the past situations and events that we seek to describe. There are few written records of MSEA languages, with the exception of some major languages. The vast majority of MSEA languages have never had a written form at all. Historical linguists' traditional approach to this problem is to compare the phonological form of words in modern spoken language varieties and abduct forms which hypothesized proto languages might have had – that is, languages that are common ancestors of present day languages of the same family. The reflexes of proto forms in modern languages show evidence of past sound changes that occurred when ancestor varieties split and underwent innovations, for example due to communities migrating to another area. Reconstructions of proto forms and shared innovations allow us to define language family groupings and sub groupings. The innovations and changes in question take place over time periods of well over a thousand years in the case of MSEA. Massive change can occur in a language over those periods of time, due to internal innovations as well as borrowings and other effects due to contact between languages. A result of such massive change is that many words from a proto language may be lost or replaced, leaving a heavily reduced number of data points to use in evaluating the historical comparative status of the language. Historical comparative linguistics has to make do with fragmentary data, and this adds to the tenuous and tentative nature of our claims about the historical relatedness among languages. So, the claims and findings reviewed in this chapter are not settled science. There remain countless controversies, disputes, and unresolved questions, especially concerning lower level subgroupings within language families. That said, there are many broad points that experts do agree on. We will focus largely on these areas of broad agreement. Our understanding of historical relations among the languages of MSEA is developing quickly and will continue to do so.

2.1 Current Developments

There are three main reasons for the high speed of development in historical linguistic research in MSEA today: unprecedented access to linguistic data, unprecedented access to other sources of relevant information, and new methods of historical linguistic analysis.

2.1.1 Unprecedented Access to Linguistic Data

Over recent years we have gained unprecedented access to relevant linguistic data. A greater volume of empirical material is being published today than ever before, in word lists, dictionaries, and grammatical descriptions. New findings are possible in part because more researchers are doing field work today than ever. This is partly due to freer and easier access to areas in which previously undescribed or little described languages are spoken. Improved access is a result of improved transport and infrastructure in MSEA, along with more open sociopolitical conditions. Due to government controls associated with conflict and political instability which have dominated the region since the 1960s, though easing since the early 1990s researchers have long been unable to gain permission to freely travel and conduct research in many parts of MSEA. These restrictions are beginning to ease. New data and analyses are being published not only in English but also in the national languages of MSEA. New findings may change our understanding of how MSEA languages are interrelated, and what events may have led to the cases, and kinds, of structural divergence that we observe. A new word list, for example, could allow us to distinguish between competing hypotheses of a linguistic subgrouping, competing theories of direction and scale of migration, or competing views as to the role of contact between communities in language change.

2.1.2 Unprecedented Access to Other Sources of Relevant Information

The second factor in the uptick of historical comparative language research in MSEA today is a similarly unprecedented degree of access to new *non* linguistic data. In genetics, there has been an explosion of new research findings of relevance to questions about historical relations between modern populations. We have extensive new information about the distribution of specific genes and genetic markers in people of the MSEA area (Oota *et al.* 2001, Matsumura and Hudson 2005, Hill *et al.* 2006, Oppenheimer 2011, Dancause *et al.* 2009, Macaulay *et al.* 2005, Stoneking and Delphin 2010, Wangkumhang *et al.* 2013). This information can be used in inferring or abducting patterns of past migration and interaction among human groups. The

resulting hypotheses can be checked for correlations with hypotheses made on the basis of linguistic data.²

The availability of new findings in human genetics does not guarantee that recalcitrant research questions will readily be answered. It depends on the case at hand. Fix (2011) examines the evidence from DNA measures of speakers of Austronesian languages in peninsular Malaysia, and argues that ‘waves of migration’ – as proposed by demic diffusion models – ‘are not required to explain the origins of genetic diversity among the Malaysian Orang Asli’ (Fix 2011: 289).

Advances in the analysis of ancient DNA are leading to empirical breakthroughs with the potential to settle long standing questions. McColl *et al.* (2018) were able to carry out full genome DNA sequencing from the skeletal remains of twenty six individuals who lived in or near MSEA during periods up to eight thousand years ago, well before the adoption of agriculture in the area. The authors used these new findings to test competing hypotheses about the peopling of Southeast Asia since the occupation of the area by people of the Hoabinhian complex, whose lithic technological tradition was sustained for a period of at least 40,000 years, across the last three Marine Isotope Stages. One hypothesis is that ‘Hoabinhian hunter gatherers adopted agriculture without substantial external gene flow’ (McColl *et al.* 2018: 88; see Hanihara 2006, Pietruszewsky 2005). Another is that agriculturalists from present day China replaced the indigenous Hoabinhian inhabitants around four thousand years ago (McColl *et al.* 2018: 88; see Matsumura and Oxenham 2013, Jinam *et al.* 2017). McColl and colleagues surmise from these new data that ‘both Hòabìnhan hunter gatherers and East Asian farmers contributed to current Southeast Asian diversity, with further migrations affecting island SEA and Vietnam’ (McColl *et al.* 2018: 88). They conclude that present day Southeast Asian populations derive ancestry from numerous ancient populations. These include the mainland Hoabinhians, who lived in the area for more than thirty thousand years, a demographic expansion from present day China around four thousand years ago during the Neolithic transition to farming, an expansion of Austroasiatic speaking farmers some two thousand years ago, and two further expansions from the northeast, those of Tai Kadai speakers and Austronesian speakers, both within the last two thousand years. McColl *et al.* (2018) tie their account of genetic distribution directly to the expansion of language families:

The evidence described here favors a complex model including a demographic transition in which the original Hòabìnhan admixed

² Since Boas (1911) it has been acknowledged that genes, language, and culture do not necessarily travel together. That said, they do often travel together, and so hypotheses based on such correlations are normally assumed.

with multiple incoming waves of East Asian migration associated with the Austroasiatic, Kradai, and Austronesian language speakers.

(McColl *et al.* 2018: 91)

New research in genetics includes work on plant and animal species, bringing new information to bear on questions of human history. The distribution of genetic markers in domesticated plants and animals of MSEA and neighbouring regions allows for new or refined hypotheses about past patterns of human livelihood, inter group contact, and directions of migration (see for example Fuller 2011 on rice, Larson *et al.* 2010 on pigs). Again, these may be checked for correlations with inferences from linguistic data.

Alongside the new genetics data, we are also seeing an explosion of empirical findings in other research areas of relevance to human history. New findings in bioarchaeology are yielding insights into past dietary habits, of obvious relevance to hypotheses about agricultural practices. Halcrow and Tayles (2011) outline developments in bioarchaeological work that shed light on kinship and migration in the past. The evidence points to a deterioration of health in the Iron Age (Halcrow *et al.* 2016, King *et al.* 2017, Clark *et al.* 2017), associated with the advent of agriculture.

Several new kinds of data and analysis are relevant. Isotopic analysis measures the presence of stable isotopes such as strontium and oxygen in bones and teeth. Differences in these isotope measures are associated with different geological regions, and so can provide evidence that individuals have migrated to the location where their bones or teeth are recovered (Bentley 2006). Biological distance analysis uses morphometric measures to find variations in phenotypic skeletal size and shape, allowing inferences about population history, including migration and gene flow (Stojanowski and Schillaci 2006). Yet further measures of skeletal and dental remains provide evidence of the state of health, diet, physical activity, and population demography of individuals whose remains have been recovered in archaeological sites. All of this evidence may be relevant for comparing and evaluating hypotheses about human history from other data sources of the kinds mentioned already, from linguistics to genetics.

Over recent decades, there have been major advances in paleogeography, paleoclimatology, and paleoecology, providing evidence of past habitat changes, such as sea levels, forest cover, and climate. Again, these considerations are of direct relevance to questions of human livelihoods and migrations at different places and in different times, and thus to questions of the spread and diversification of languages in MSEA. And finally, new methods in the study of artefacts and other evidence from archaeological digs can be used in inferring a range of things, from people's evident knowledge and practices of pottery techniques, to their know how in hunting and weapons technology, to aspects of their social structure.

2.1.3 New Methods of Historical Linguistic Analysis

A third reason why the field of historical linguistics is developing quickly is that innovative new methods are disrupting current scholarship in historical linguistics and language classification. By around the turn of the twenty first century, biologists had exploited unprecedented degrees of computational power, combined with probabilistic statistical methods, to devise new bioinformatic techniques for evaluating relatedness among biological units such as species. These techniques are now being applied to language data (see for example Greenhill and Gray 2009; see Gray and Jordan 2000 on Austronesian, Dunn *et al.* 2011 on Aslian, and Sagart *et al.* 2019 on Sino Tibetan).

*

In the following sections, we survey research on the histories and classifications of languages in the five main languages families of MSEA.

2.2 Austroasiatic

Austroasiatic languages are spoken across the entire swathe of MSEA territory, from coastal Vietnam on the southeastern tip of the area, to the south in peninsular Malaysia, up and across the breadth of Zomia and far beyond MSEA to the west, across to coastal Myanmar, and deep into pockets of central India (see Map 2.1). An estimated 126 Austroasiatic languages are spoken in greater MSEA.

The distribution shown in Map 2.1 is discontinuous, with the languages found in distinct pockets spread across a large area. Around 130 Austroasiatic languages are spoken today, in up to thirteen established branches. These branches vary widely in size, from Khmer containing just one language (Cambodian) to Bahnaric containing some thirty languages. Speaker populations for each branch also vary widely in size and sociopolitical status. Some languages are spoken in a handful of villages, while others have millions of speakers. Vietnamese is spoken by up to 100 million people, Khmer by around 16 million, and Santali (spoken in eastern India) by more than 7 million. Map 2.2 shows Austroasiatic languages in core MSEA.

Two Austroasiatic languages – Vietnamese and Cambodian – have a special status relative to the rest. They are the primary languages of nation states, associated with political dominance by majority ethnic groups within their own nations. Vietnamese is spoken both in Vietnam and in large expatriate communities in Laos and Cambodia, as well as in North America, Australia, and Europe. Cambodian, or Khmer, is the national language of Cambodia, and is also spoken by a significant ethnic minority population in provinces of Thailand along the northern border of Cambodia and in the Mekong Delta of Vietnam.

Scholarship on the Austroasiatic family of languages began in the mid nineteenth century when language researchers observed possible connections based on word forms



Map 2.1 *Approximate distribution of languages of the Austroasiatic family*
Sub-branches are represented in different shades.

between geographically separated languages, from Munda languages spoken in India, through Mon spoken in Myanmar and Thailand, to Vietnamese and other languages spoken in Vietnam. Sidwell (2015) and Sidwell and Rau (2015) provide detailed overviews of the history of research on the Austroasiatic language family, building on earlier work by Adams (1989), Parkin (1991), Diffloth and Zide (1992), Peiros (1998), van Driem (2001), Sidwell (2009a), and Nagaraja (2011). Table 2.1 shows lexical correspondences noted a century ago by Przluski (1924) of the numerals ‘one’ to ‘four’ in languages that are spoken some 3,000 kilometres apart: Santali (a Munda language spoken in India), Mon (a Monic language spoken in southeast Myanmar and Central Thailand), and Muong (a Vietic language spoken in Vietnam):

Table 2.1 *Lexical correspondences in Santali, Mon, and Muong*

	Santali	Mon	Muong
‘one’	mit	mwai	môt
‘two’	bar	bā	hal
‘three’	pä	pi	pa
‘four’	pōn	pan	pōn

Noted by Przluski (1924: 386); this table is from Sidwell 2015: 153.



Map 2.2 *Approximate distribution of Austroasiatic languages in Core MSEA*
Sub-branches are represented in different shades.

Scholars debated whether correspondences such as these, and many others that were brought to light, were due to contact and diffusion or shared inheritance. This long lasting debate was largely settled soon after the turn of the twentieth century,³ when Grierson's edited volumes of the *Linguistic Survey of India* (1903–1928) brought a trove of new linguistic data to bear on the issue (Grierson 1904). Using data from Grierson's survey, Wilhelm Schmidt published a series of works between 1901 and 1906 which established

³ Despite some lasting opposition; e.g., Sebeok 1942.

the idea of shared inheritance from a common Austroasiatic ancestor. Schmidt's work 'put comparative Austroasiatic studies on a strong footing' (Sidwell 2015: 151).

2.2.1 Austroasiatic Subgrouping

An example of the criteria for high level subgrouping of Austroasiatic languages is the development of Proto Austroasiatic stops. A recent reconstruction of Proto Austroasiatic posits a three way distinction in plosives: voiceless unaspirated, voiced, and implosive (see Table 2.2).

In most Austroasiatic languages, processes of merging have seen this three way distinction collapse into a two way distinction. There are different patterns of merger. Examples in Table 2.3 show that 'implosive bilabial *b̥ survives in Old Mon (Monic), Bahnar (Bahnaric), and Katu (Katuic), whereas it has variously merged with the equivalent plain voiced and plain voiceless stops elsewhere' (Sidwell and Rau 2015: 239).

A similar pattern can be observed with the implosive apical *d̥. Old Mon, Bahnar, and Katu preserve a reflex of the implosive *d̥ while languages in other branches have merged it either with *t or *d. This is shown in Table 2.4.

Table 2.2 *Proto-Austroasiatic inventory of syllable-initial consonants*

*	p	t	c	k	?
b	d	J		g	
b̥	d̥	(f?)			
m	n	jn		ŋ	
w	r, l	j			
	s				h

From Sidwell and Rau 2015: 238.

Table 2.3 *Different patterns of merger of the Proto-Austroasiatic three-way initial bilabial plosive split, showing retention of implosive bilabial *b̥- survives in Old Mon (Monic), Bahnar (Bahnaric), and Katu (Katuic)*

	Old Mon	Khmer	Bahnar	Katu	Khuu			
					Cuang	Palaung	Khasi	Sora
'drunk'	bəlbul	pūl ¹	bəŋul ²	bal			bu: ?a:c	buʔul
'two'	bar	pì:	ba:r	ba:r	ba:r			bagu
'shoots'	tbaŋ	tùmpèəŋ	təbaŋ	ɿabay	(tapáŋ) ³	bəŋ		tabəŋ
'obtain'				bo:n	bian	buun		

¹ 'vegetable poison', ² 'toadstool', ³ Kammu Yuan
After Sidwell and Rau 2015: 239.

Table 2.4 *Different patterns of merger of the Proto-Austroasiatic three-way initial apical plosive split, showing retention of implosive apical *d-* in Old Mon (Monic), Bahnar (Bahnaric), and Katu (Katuic)

	Old Mon	Khmer	Bahnar	Katu	Khmu Cuang	Palaung	Khasi	Kharia
‘bamboo’	<i>kdiŋj</i> ¹		<i>d̥iŋ</i>		(<i>dɔŋ</i>) ⁶	<i>diŋ</i>	<i>tindɔŋ</i>	
‘water’	<i>daik</i>	<i>tik</i>	<i>da:k</i>	<i>dɔk</i>				<i>da?</i>
‘span’	<i>eda?</i> ²		<i>səda:</i>	<i>cada:</i>	<i>sn̥da?</i>		<i>(chəthá?)?</i> ³	<i>sda</i>
‘cover/hide’	<i>kðøp</i>	<i>tùp</i>	<i>kðøp</i>	<i>pad̥øp</i>	<i>dap</i>	<i>døp</i>		<i>da?b</i>
‘ripe’		<i>tum</i>	<i>du:m</i>	<i>dam</i>	<i>hnduum</i> ⁴	<i>kəndum</i> ⁵		

¹ Middle Mon ‘areca-nut container’; ² Middle Mon; ³ Lamet; ⁴ ‘to ripen’; ⁵ ‘bud’; ⁶ Mlabri

After Sidwell and Rau 2015: 239

These changes in the structure of the proto stop series provide evidence of splits in the history of Austroasiatic languages. But they are an exception to the otherwise remarkable stability of proto initials in the history of these languages. Table 2.5 puts the proto implosive stops into the context of the other proto initials, which show great consistency in their reflexes across languages of the family.

2.2.2 Munda Languages

Munda languages have been treated as exceptional in relation to the other Austroasiatic branches. While Schmidt (1906) grouped the Munda languages together with other Austroasiatic languages, later work – especially Pinnow (1954, 1959, 1960, 1963) would establish a view that Austroasiatic ‘consists of two coordinate families, Munda and Mon Khmer’ (Sidwell 2015: 167).⁴ The idea of a highest level binary branching between Munda and Mon Khmer became the received view for decades. But this model was never conclusively demonstrated. Some have tried to group Munda together with sub Mon Khmer branches, or to place it as a sister of those branches in a rake like structure for the family (see section 2.2.3). Today, largely by applying the principle of parsimony in the face of insufficient evidence, it is no longer generally assumed that Munda is one of just two highest level Austroasiatic coordinates, with Mon Khmer as the other (see Sidwell 2015: 167, Benjamin 2012: 136 fn 2). The term Mon Khmer is still widely used, but it should be understood to refer simply to non Munda languages of the Austroasiatic family rather than a group of languages descended from a single proto language.

There are several reasons why Munda languages are readily thought of as distinct from all of the other Austroasiatic languages. One is that they are spoken entirely outside of the MSEA area, with their speakers residing exclusively in India, as far west as

⁴ Mon-Khmer has had other labels; e.g., Khmer-Nicobar in Pinnow’s work.

Table 2.5 Summary of Austroasiatic proto-branch level initial-consonant correspondences

pAA	*p-	*p ^{b/-} /*p-	*p-	*p-	*p-	*p-	*p-	*p-
pMunda	*b-	*b-	*b-	*b-	*b-	*b-	*b-	*b-
pKhasian	*b-	*b-	*b-	*b-	*b-	*b-	*b-	*b-
pPalaungic	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-
pKmuic	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-
pVietic	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-
pKatuic	*s-	*s-	*s-	*s-	*s-	*s-	*s-	*s-/h-
pKhmer	*s-	*c-/*s-	*c-	*c-	*c-	*c-	*c-	*s-
pPearic	*t-	*t-/*t-	*t-	*t-	*t-	*t-	*t-	*t-
OldMon	*d-	*d-	*d-	*d-	*d-	*d-	*d-	*d-
pAslian	*r-	*r-	*r-	*r-	*r-	*r-	*r-	*r-
pNicobarese								

Reproduced from Sidwell and Rau 2015: 240 Shaded cells indicate that the modern reflex departs from the majority Proto-Austroasiatic pattern

Maharashtra State in central India. Another is that while lexical evidence clearly establishes the Munda languages as Austroasiatic, they differ markedly from other Austroasiatic languages in terms of a large set of typological features. This typological differentiation is presumed to be due to intensive contact with South Asian languages. For example, while Munda languages have falling (initial) phrase accent, inflectional morphology, trochaic word canon,⁵ and simple CVC syllables, other Austroasiatic languages have rising (final) phrase accent, no inflectional morphology, iambic word canon and complex syllables (Donegan and Stampe 2002; see section 4.9 below on prosodic hierarchy and phonological words for further discussion). These differences are understood to be the result of contact induced systemic changes in the languages rather than genealogically acquired differences of any classificatory significance.

2.2.3 Non Munda ('Mon Khmer') Subgrouping

Going beyond the question of Munda versus the rest, researchers through the twentieth century – for example, Maspéro 1912, Sebeok 1942, Pinnow 1959, and Diffloth and Zide 1992 – worked to establish internal sub groupings within Austroasiatic. Breakthroughs came in the 1960s and 1970s, notably through the work of Thomas, Headley, Huffman, and Smith (see especially Thomas and Headley 1970, Thomas 1966, Smith 1981). These researchers relied heavily on lexicostatistical methods (Swadesh 1955, Hymes 1960) which have been strongly critiqued and discredited (Hoijer 1956,⁶ Bergsland and Vogt 1962).⁷ But the hypotheses that they put forward have been supported by subsequent evidence and analysis.

Thomas and Headley (1970: 405) argued that the Austroasiatic phylum contained ‘at least four families: Munda, Mon Khmer, Malacca, Nicobarese’. The key advance from their work was their subgrouping of languages under Mon Khmer. They discerned subgroupings of languages that had previously been lumped together, with evidence for distinguishing Pearic, Katuic, and Bahnaric groups. They divided Mon Khmer into nine branches of equal status and proposed a three way sub branching of Bahnaric. Figure 2.1 shows the scheme diagrammatically.

The subgroupings shown in Figure 2.1 have largely been supported by subsequent data and analysis and are now generally accepted by experts in the field. Thomas and

⁵ Ring and Anderson (2018) have recently presented empirical challenges to this claim, with evidence that there are iambic structures in Munda.

⁶ See also Zhang and Gong (2016).

⁷ These authors were aware of the issues: ‘Lexicostatistics is not a precision tool. Careful phonological reconstruction is necessary if one desires detailed information about language relationships. Lexicostatistics is useful, however, for giving a quick general picture of language groupings’ (Thomas and Headley 1970: 411).

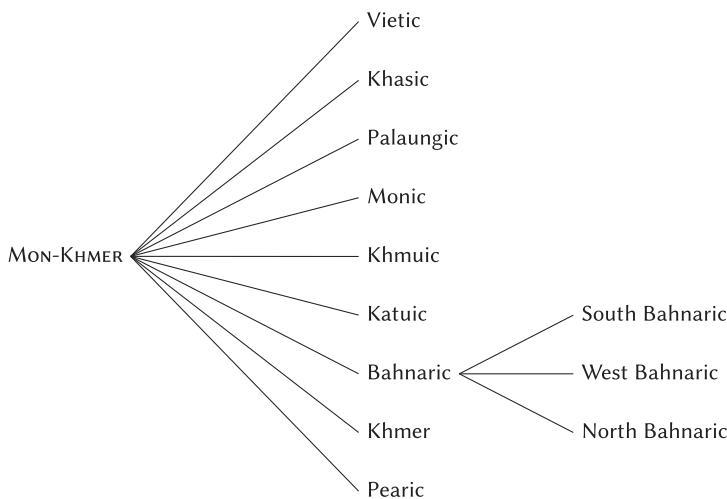


Figure 2.1 *Mon-Khmer classification according to Thomas and Headley*
After Thomas and Headley 1970: 405 7.

Headley's proposals 'successfully framed discussions of Austroasiatic classification ever since' (Sidwell 2015: 160).

The question that has not been resolved is whether these established groups can be meaningfully subgrouped at intermediate levels. One attempt to join subgroups in a deeply nesting tree representation of Austroasiatic was made by Diffloth and Zide (1992). This featured, firstly, a binary top level division of Austroasiatic into Munda and Mon Khmer, justified by some distinctive Munda and Mon Khmer features. Second, Mon Khmer was divided into Northern, Eastern, and Southern divisions. In a revised proposal, Diffloth (2005: 79) elevated the Khasi Khmuic branch (corresponding to Northern Mon Khmer in previous work) to a primary branch of Austroasiatic, alongside Munda and Khmero Vietic/Nico Monic. This is shown in Figure 2.2.

These proposals are often cited but they have not been widely accepted in the field. There has since been a return to viewing the Austroasiatic family as quite flat in structure. Recent work, drawing on both historical comparative and computational phylogenetic methods (e.g., Sidwell and Blench 2011) suggests the classification scheme for Austroasiatic shown in Figure 2.3 (after Sidwell 2015: 179).

One reason why proposals of medial subgroupings – that is, groupings of established groups – have not been broadly accepted in Austroasiatic linguistics has to do with problems of access to appropriate data. With adequate and high quality data, it may turn

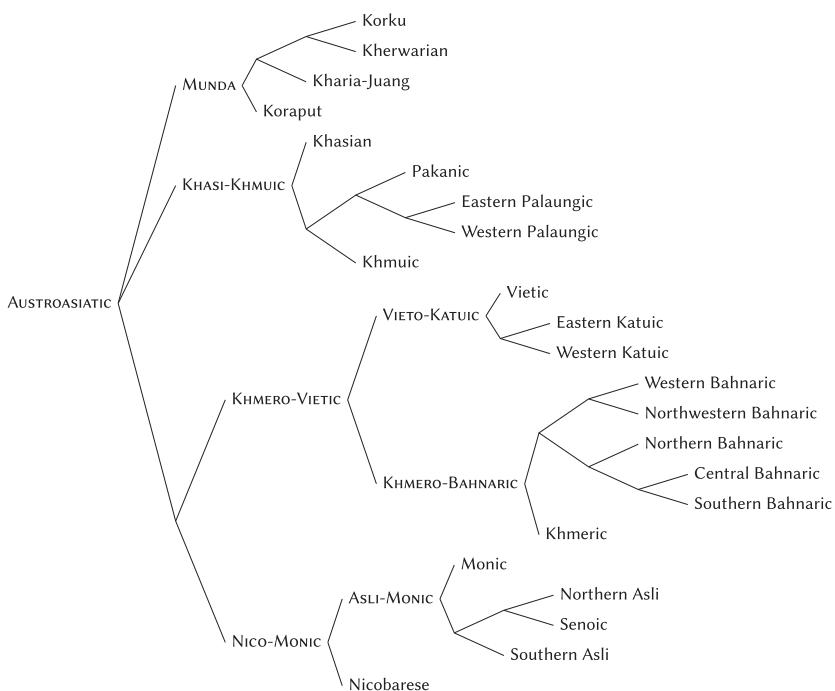


Figure 2.2 *Austroasiatic* (after Diffloth 2005: 79)

This version omits the speculative datings for each split offered in Diffloth's original diagram.

out that the rake like structure that we see being presented today is in fact an accurate representation of the natural history of Austroasiatic languages. Or clear evidence for mid level subgroupings may emerge. Meanwhile, the rake like structure in Figure 2.3 is appropriate given the lack of conclusive evidence for hierarchical grouping of established lower level subgroups. For now, Austroasiatic can be presented as a list of thirteen branches, as shown in Table 2.6.

2.3 Tai-Kadai

The Tai Kadai family of languages also known as Kra Dai, Kadai, and Zhuang Dong (Edmondson and Solnit 1988: 3–5) consists of languages spoken by at least ninety million people in a wide distribution throughout MSEA, in southern China, central and northern Myanmar, northern Vietnam, northern Cambodia, and throughout Thailand and Laos. An estimated ninety Tai Kadai languages are spoken in greater MSEA. See Map 2.3.

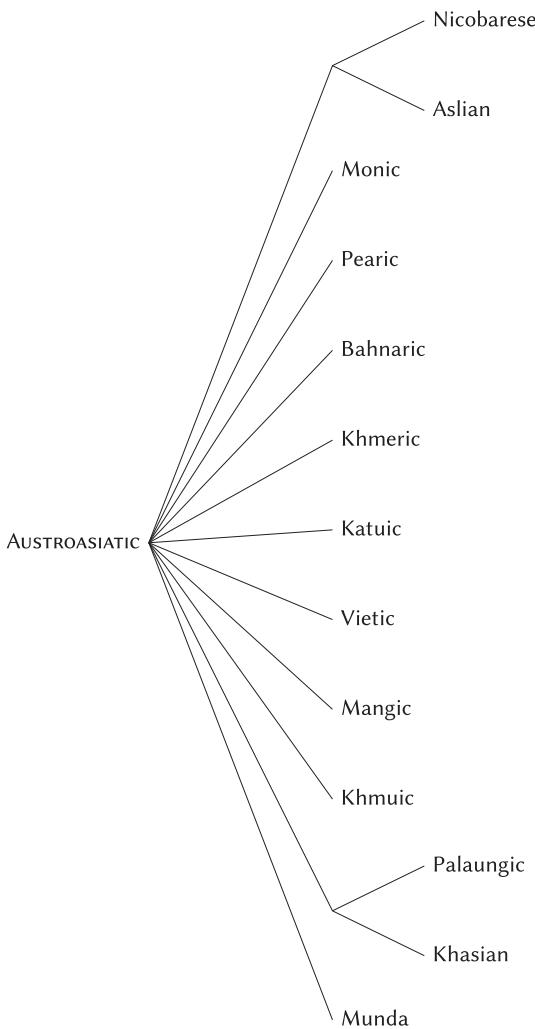


Figure 2.3 *Austroasiatic* (after Sidwell 2015: 179)

Note that there is no Mon-Khmer branch in the figure. The language groups previously considered to constitute a Mon-Khmer unit (i.e., everything but Munda) now all directly branch from the top level (with Khasian-Palaungic and Aslian-Nicobarese here proposed as subgroupings).

A number of the biggest languages of MSEA belong to the Tai branch of the Tai Kadai family. From the Southwestern Tai branch, Lao and Thai – arguably in a dialect relationship with each other – are the national languages of Thailand and Laos, with some eighty million speakers combined. Lao is spoken by as many as twenty million people in Thailand, where it is referred to either as Lao or Isan (from the name for the northeast region of Thailand

Table 2.6 *Thirteen branches of Austroasiatic*

Branch	Main regions where spoken
Aslian	Malay Peninsula
Bahnaric	central Indo-China
Katuic	central Indo-China
Khasian	Meghalaya State of India
Khmer	Cambodia and neighbouring areas
Khmuiic	northern Laos
Mangic/Pakanic	southern China and Vietnam
Monic	southern Myanmar and central Thailand
Munda	eastern and central India
Nicobarese	Nicobar islands of India
Palaungic	Shan State of Myanmar
Pearic	Cambodia and Thailand
Vietic	Vietnam and central Laos

After Sidwell 2015: 148 9.

where Lao is locally spoken). Also in the Southwestern Tai branch, Shan is spoken by more than three million people in central and eastern Myanmar, as well as in neighbouring areas of China and Laos. And from Northern and Central branches of Tai, varieties of Zhuang are spoken by some sixteen million people in Guangxi and parts of neighbouring provinces of China.⁸ The larger Tai Kadai family includes not just the Tai languages that spread from southern China into core MSEA over the last millennium or so, but also many minority languages still spoken in China and across the northeast of core MSEA (see Map 2.3).

2.3.1 Subgrouping of Tai Kadai Languages

Edmondson and Solnit (1997: 7) propose a number of common traits for Tai Kadai languages. These include the following:

- morphemes are largely monosyllabic
- every (stressed) syllable has a tone
- there is relatively little tone sandhi (change of tone in context)
- modifiers are postposed and clause structure is verb medial
- clusters of stop (oral and nasal) plus liquid must be assumed for the proto language
- a common core vocabulary

Some of these criteria are not specific to Tai Kadai. The last criterion is of course the most central to the majority of research on the classification and historical background of the modern languages. Table 2.7 shows some basic vocabulary items that have

⁸ The term Zhuang is used as an ethnic label for Tai-speaking peoples of Guangxi autonomous region in China. This can create confusion, as the Northern Zhuang and Southern Zhuang varieties are in fact separate languages from distinct branches of Tai.



Map 2.3 *Approximate distribution of Tai-Kadai languages*
Sub-branches are represented in different shades.

Table 2.7 *Some basic vocabulary in Tai-Kadai languages from distant branches*

	Gelao1	Gelao2	Hlai	Kam	Zhuang	Thai
‘pig’	<i>mau</i> ˥	<i>mau</i> ˥	<i>pou</i> ˧	<i>yu</i> ⁵	<i>mou</i> ¹	<i>mu</i> ¹
‘bird’	<i>m̥nay</i> ˧	<i>maŋnoŋ</i> ˧	<i>tat</i> ˧	<i>nok</i> ⁸	<i>nok</i> ⁸	<i>nok</i> ⁸
‘blood’	<i>pla</i> ˧	<i>pla</i> ˧	<i>la:ŋ</i> ˧	<i>pha:t</i> ⁹	<i>luu:i</i> ⁹	<i>luət</i> ⁹
‘intestine’	<i>sai</i> ˧	<i>taŋse</i> ˧	<i>ra:i</i> ˧	<i>sa:i</i> ³	<i>sa:i</i> ³	<i>sai</i> ³
‘rain’	<i>n̥l̥mən</i> ˧	<i>mae</i> ˧	<i>fun</i> ˧	<i>pjan</i> ¹	<i>fuuŋ</i> ¹	<i>fon</i> ¹
‘teeth’	<i>m̥pje</i> ˧	<i>pjan</i> ˧	<i>fan</i> ˧	<i>pjan</i> ¹	<i>fen</i> ¹	<i>fan</i> ¹
‘dream’		<i>laŋpjan</i> ˧	<i>fan</i> ˧	<i>pjan</i> ¹	<i>phan</i> ¹	<i>fan</i> ¹

After Edmondson and Solnit 1997: 7.

reflexes in all modern branches of Tai Kadai (see also Ostapirat 2000: 3–8, 2005: 110 for a list of fifty pan Tai Kadai etyma).

The roots meaning ‘teeth’ and ‘dream’ in Table 2.7 are attested from Gelao to Thai languages that are only distantly related in the Tai Kadai tree, and that are spoken some two thousand kilometres apart. These roots ‘must have been very similar all the way back to the proto language’ (Edmondson and Solnit 1997: 8).

Early statements of the historical grouping of Tai Kadai languages were based on pioneering studies including Grierson (1904), Maspéro (1911), Burnay and Cœdès (1927), and Wulff (1934). In an attempt to link Tai Kadai languages to Indonesian, Benedict (1942: 587) defined Tai (which he labelled as ‘Thai’) to include ‘Ahom, Khamti, and Shan, in the west; Siamese and Lao, in the south; White Tai and Black Tai, in the east; Nung and Tho, in the northeast; and Dioi, in the north’. Benedict proposed that the Tai Kadai family has three main sub branches: Kra (Gelao, Lachi, Laha, Paha, Buyang, and Pubiao), Hlai (Baoding, Yuanmen, Heitu), and Kam Tai.

Kra languages can be grouped on the basis of a set of distinct Kra etyma. A selection of these is presented in Table 2.8 (from Ostapirat 2000: 9–13).

Hlai (or Gelao Hlai) languages can similarly be grouped on the basis of distinct etyma. Table 2.9 shows examples from the lexicon of wet rice paddy rice farming technology (Edmondson and Solnit 1997: 8).

Hlai and Kra languages are spoken in the northeastern part of the family’s distribution. By contrast, Kam Tai languages show a more spread out distribution, as many of these languages are understood to have descended from varieties spoken by people who spread southwest across the MSEA area over a period of centuries from the first millennium CE, in search of good land to cultivate paddy rice (see discussion in Chapter 1).

The three way branching of major components of the Tai Kadai family is widely assumed in modern descriptions of the family. It can be seen as the highest level split shown in Figure 2.4, a recent proposal for the grouping of Tai Kadai languages (from Owen 2012: 13, adapted from Diller 2008: 7).

While the main groupings provided in Figure 2.4 are ‘well established’, they should still be taken as provisional (Ostapirat 2005: 108). ‘Lexical and phonological evidence exists which suggests the possibility of grouping together Kam Sui and Kra on the one hand, and Tai and Hlai on the other’ (Ostapirat 2005: 108).

The validity of a Kam Tai grouping is generally accepted among experts (see Liang and Zhang 1996 for a reconstruction of Proto Kam Tai). But relations between languages below the level of Kam Tai in Figure 2.4 remain under discussion. For example, there is disagreement as to whether Tai and Kam Sui are sisters under a single sub group node, or whether they are of equal rank as direct branchings from the highest Tai Kadai level (Edmondson and Solnit 1997: 2–3; cf. Chamberlain 1997). This in part explains the different labels that are used in research

Table 2.8 Lexical correspondences supporting the grouping of Kra languages (*Gelao, Lachi, Laha, Paha Buyang, and Pubiao*) as distinct from languages of other branches of Tai-Kadui

	Gelao	Lachi	Laha	Paha	Buyang	Pubiao	Hhai	Sui	Tai
‘meat/flesh’	ʔɒ C1	ʔəu C1	ʔəu C1	ʔəu C1	ʔəu C1	χjau C1	gom C	naan C2	nua C2
‘fat’	nan A2	nfija A2	minal B2 -t	nam A2	ren A2	niñin A2	gwei C	pii A2	phi A2
‘hawk’	li C2	lifi C2	klaaj C2	ðaaŋ C2	laaŋ C2	laaŋ C2	ŋeau A	ŋjaau A2	jiau B2
‘water’	pau C1	hi C1	ʔuŋ C1	ʔɔɔŋ C1	ʔɔɔŋ C1	ʔɔɔŋ C1	nom C	nam C1/2	naam C2
‘forget’	te D2	tfiŋ D2	dap D2	dap D1	ʔdəŋ D1	ʔdəŋ D1	luuum B	laam A2	luuum A2
‘plant (v)’	tan C1	tjă C1	tam C1	tan C1	tap C1	tap C1	gwaa A	mba A1/2	pluuuk D1
‘four’	pu A1	pu A1	paa B1 -t	paa A1	pee A1	ts'hau C	ts'hau C	ɛi B1	si B1

After Ostapirat 2000: 9–13

Table 2.9 Rice farming vocabulary, establishing a split between Gelaо-Hlai and the rest of Tai-Kadai

Thai	Zhuang	Bouyei	Dai	Kam	Mulam	Sui	Lakkja	Gelaо	Hlai
'rice plant'	khaу ³	hau ⁴	hau ⁴	xau ³	qau ⁴	2au ⁴	kou ³	mon ¹⁴	mu n ³
'rice seedling'	kla ³	kja ³	tea ³	ka ³	ka ³	ka ³	kje ³	—	fan ¹
'rice straw'	faan ²	fuu ѡ ²	fuu ѡ ²	fuu ѡ ²	pa ѡ ¹	ma ѡ ¹	wa ѡ ²	—	—
'plow'	thai ¹	kwai ¹	eai ¹	thai ¹	khai ¹	khai ¹	khai ¹	ndai ¹	ηη ³
'carrying pole'	khaan ²	ha n ²	ha n ²	ka n ²	la n ²	qa n ¹	la n ¹	—	lei ⁶
'rice mortar'	— — —	yum ¹	zum ¹	kem ¹	kem ¹	kem ¹	kjuen ³	tshai ¹ fia ⁷	— — —
'water buffalo'	khuwa i ²	va i ²	xva i ²	va i ²	kwe ²	hw ²	sui ³ nou ²	ntai ¹ tu ¹	tui ³

After Edmondson and Sohnit 1997: 8

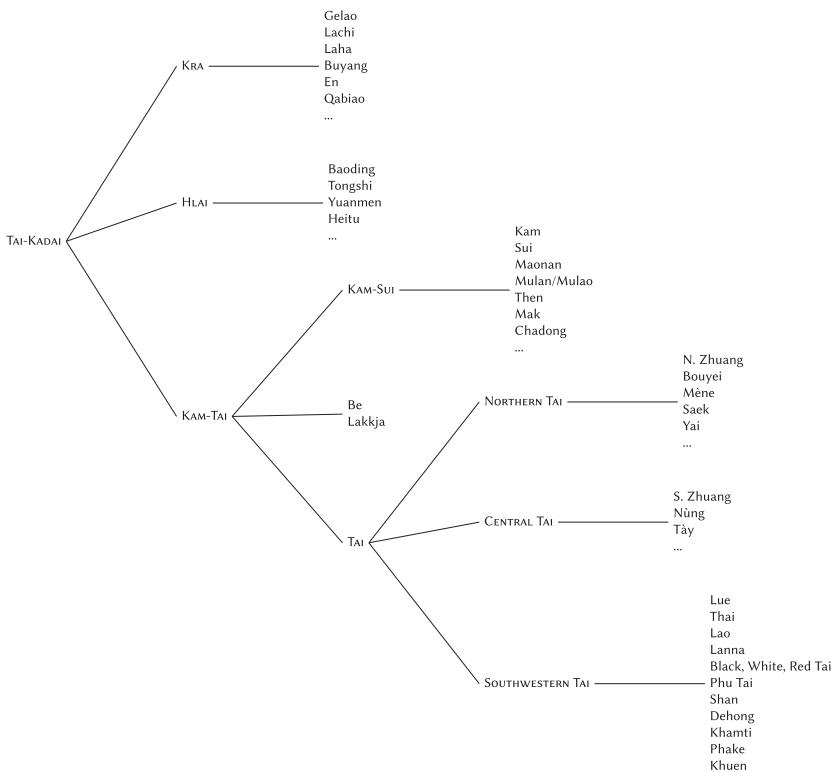


Figure 2.4 *Tai-Kadai languages*

Adapted from Owen 2012: 13, Diller 2008: 7.

on this family of languages. In his reconstruction of the phoneme inventory of Proto Hlai, based on data from twelve Hlai languages of Hainan Island, Norquest (2016: 1) makes reference to the ‘Kra Dai phylum (also commonly known as Tai Kadai)’. The label ‘Kadai’ (or ‘Kra Dai’) implies that there is a lower level Kam Tai branch which groups together Kam Sui and Tai languages, among others. The label ‘Tai Kadai’ implies that Tai and Kam Sui are separate sub families. Different proposals concerning Tai Kadai internal classification have different implications for the shape of proposed phylogenetic trees, ranging from more rake like to more hierarchically nested (Edmondson and Solnit 1997: 3, Norquest 2016: 2).

Other examples of ongoing discussions in comparative Tai Kadai concern the classification of certain Tai Kadai languages spoken in China. One example is Lakkja (or Lakkia), spoken in Guangxi Province. Thurgood *et al.* (1988: 180) argue that Lakkja is

a coordinate of Kam Tai, while both Solnit (1988) and Thongkum (1992) take the view that Lakkja is a daughter of Kam Tai (i.e., that Lakkja is descended from Proto Kam Tai): ‘The Lakkja speak a language that is indisputably Kam Tai’ (Solnit 1988: 220) and ‘There is little doubt that Lakkja is a language within the Kam Tai branch of the Tai Kadai family’ (Thongkum 1992: 88). But Solnit and Thongkum disagree as to the placement of Lakkja within Kam Tai. Solnit (1988: 229) argues that it is closer to Kam Sui. Thongkum (1992: 79) argues that it is closer to Tai. The reason for this disagreement is that the evidence is ambiguous, as Solnit (1988) shows in his review of the data on this question. In some ways, Lakkja appears to side with Kam Sui innovations, in others with Tai innovations. In some ways Lakkja appears to share retentions with Kam Sui, in others with Tai. And in addition, Lakkja shows unique innovations.

2.3.2 The Tai Branch

The best studied and probably best understood branch of Tai Kadai is the Tai family. This branch has three major subgroups: Southwestern Tai, Central Tai, and Northern Tai (see F. K. Li 1977). Southwestern Tai contains the most widely spoken and best known Tai languages, Thai and Lao. Together these account for the largest speaker populations. Accordingly, Southwestern Tai is ‘by far the best studied’ of the three branches (Pittayaporn 2009: 119; see F. K. Li 1977; cf. Brown 1965, Sarawit 1973, Jonsson 1991).

Pittayaporn (2009) proposed revisions to the hitherto accepted model of Southwestern Tai language origins, hypothesizing that Proto Southwestern Tai had the following features (by contrast with the hypotheses of F. K. Li 1977):

- a distinct series of uvular consonants
- a mid back unrounded vowel *ɹ
- contrastive vowel length
- absence of certain clusters that had been previously posited (specifically, ‘clusters *p^hr/l, *k^hl, and *mr reconstructed by F. K. Li (1977) and Jonsson (1991); Pittayaporn 2009: 135).

Perhaps the most prominent issue in historical comparative Tai linguistics is the historical phonology of tonal categories. Proto Tai had four tones, which are given the labels A, B, C, and D:⁹

⁹ The question of how the four-tone system came about in the first place is still open, but it is thought that final consonants in an ancestor language were lost, allowing incidental pitch and phonation differences to phonologize, in a process similar to that described for historical Sinitic, and for Vietnamese (Haudricourt 1954: 69ff.; see section 4.6 in Chapter 4 for further information).

Tones A, B, and C occurred in syllables ending in a continuant, namely a vowel or a nasal, and tone D only in syllables ending in a stop consonant. It is apparent that tone D occurred only in a special type of syllable in which the distinctions of tones may be said to be neutralized. We set up a special tone class D for this type of syllable, because it is impossible to identify it with any of the other tones which have been set up for the other type of syllables.

(F. K. Li 1977: 25)

This four tone system is an assumed starting point for the development of systems of tone in all Tai languages. Languages of the Tai family ‘all agree in the basic framework of their phonological structure’ (i.e., the basic form of syllables) but will be readily distinguished from each other in terms of the ‘system of tonal contrasts’ (Gedney 1972: 191):

One may consider that he has crossed a dialect boundary if he finds an increase or decrease in the number of tones in the system, or if he finds that a list of morphemes which in the previously studied dialect agreed in tone is now distributed among two or more different tones, or conversely, that a previously noted tonal distinction is now lost.

(Gedney 1972: 191)

From the starting point of the four tone system in Proto Tai – with tones A, B, C, and D – the subsequent development of tones in the Tai languages was influenced ‘by the phonetic quality of the initial consonant of the syllable’. ‘The opposition of voicing and voicelessness of the initial consonant influences practically all the tones in all dialects’ (F. K. Li 1977: 25). This is why the reconstruction of Proto Tai initials is closely linked to the history of tones in the languages. Gedney (1972: 195) unpacks this idea, yielding a template for comparative historical analysis of tone in any Tai language. Gedney presented ‘a short cut to discovering the structure of the tonal system of a Tai dialect’, a method of synchronic description and analysis that draws on diachronic Tai linguistics.

Gedney (1972: 197) first presents a six cell table with Proto Tai tones in syllables with ‘live’ or continuant finals – that is, vowels, glides, and nasals. The table features an early split between tones A, B, and C, into A1, B1, C1, A2, B2, and C2, as a result of erstwhile voicing distinctions. This is shown in Table 2.10.

Table 2.10 *Early Tai tone categories for ‘live’ syllables,¹ resulting from a historical split of three proto tones (A, B, and C)*

Initials at time of split	Proto-Tai tones		
	A	B	C
	Voiceless 1	3	5
Voiced 2	4	6	

¹ i.e., syllables with continuant finals.

The split was conditioned by the status of syllable-initials as voiced versus voiceless.

Gedney (1972: 197) constructed a checklist of widely attested modern Tai words associated with each cell. The checklist could be used by language analysts in making a first pass in describing any modern Tai tonal system. Here, Gedney supplies the forms with their Thai/Siamese consonant and vowel values:

Column A

Box 1 (A1)	<i>pii</i> ‘year’ <i>naa</i> ‘thick’
Box 2 (A2)	<i>phii</i> ‘fat’ <i>naa</i> ‘ricefield’

Column B

Box 3 (B1)	<i>pii</i> ‘flute, oboe’ <i>naa</i> ‘arrow of crossbow’
Box 4 (B2)	<i>phii</i> ‘older sibling’ <i>rua</i> ‘to leak’

Column C

Box 5 (C1)	<i>naa</i> ‘face’ <i>maa</i> ‘to increase’
Box 6 (C2)	<i>naa</i> ‘mother’s younger sibling’ <i>maa</i> ‘horse’ <i>rua</i> ‘fence’

Gedney then adds four further boxes, corresponding to ‘dead’ syllables, that is, syllables with stop finals. These boxes were further divided into those with long vowels versus those with short vowels (from Gedney 1972: 199). This is shown in Table 2.11.

Table 2.11 *Early Tai tone categories for ‘dead’ syllables,* with long versus short vowels*

		Proto-Tai Tone D	
		short	long
Voiceless	7	9	
	8	10	

* i.e. syllables with stop/checkered finals.

The split of proto tones, conditioned by the status of syllable-initials as voiced versus voiceless, is also shown.

Again, Gedney supplied modern Tai/Thai forms associated with each box:

Box 7	(D1 short)	<i>lak</i> ‘stake’ <i>mat</i> ‘flea’
Box 8	(D2 short)	<i>lak</i> ‘to steal’ <i>mat</i> ‘to tie up’
Box 9	(D1 long)	<i>maak</i> ‘fruit’
Box 10	(D2 long)	<i>maak</i> ‘classifier for tools’

As Gedney went on to show, it is possible to discern further historical splits based on the tone systems of modern Tai languages, defined by further distinctions in the type of syllable initial segment that occurred in the proto language. He showed that the development of tones in languages descended from Proto Tai depended not only on the status of a syllable’s initial stop as voiced versus voiceless. He established four types of syllable onset that mattered for tonal development in modern languages:¹⁰

- ‘voiceless friction sounds’ (*s, hm, ph, etc.)
- ‘voiceless unaspirated stops’ (*p, etc.)
- ‘glottal’ sounds (*?, ?b, etc.)¹¹
- voiced sounds (*b, m, l, z, etc.)

This led him to devise a twenty cell table displaying ‘a maximum of possible tonal distinctions resulting from the various types of tonal splits that have been described’ (Gedney 1972: 202). This is shown in Table 2.12.

¹⁰ These distinct onsets are attested in modern languages of the Kam-Sui group.

¹¹ The digraph ‘?b’ represents a sound that is not distinguishable from an implosive b.

Table 2.12 *Gedney's complete system of twenty pan-Tai tone boxes*

	Proto-Tai Tones					
	A	B	C	D-short	D-long	
Initials at time of tonal splits	voiceless friction sounds, <i>*s, hm, ph</i> , etc.	1	5	9	13	
	voiceless unaspirated stops, <i>*p</i> , etc.	2	6	10	14	
	glottal, <i>*?</i> , <i>ʔb</i> , etc.	3	7	11	15	
	voiced, <i>*b, m, l, z</i> , etc.	4	8	12	16	
}			smooth syllables		checked syllables	

Five columns were for early Tai tone/syllable-type categories, and four rows for distinct types of syllable-initial segment.

Gedney supplied a checklist of words associated with the twenty historical Tai tone boxes. This checklist has not only provided researchers of Tai languages with analytical insights into the phonological history of the family, but it has also provided an immensely practical tool for field workers to begin the task of ascertaining the relationship between an observed modern tone system and the Proto Tai system. Here is the full checklist (Gedney 1972: 202–4):

Column A

- | | |
|-------|--|
| Box 1 | <i>huu</i> ‘ear’
<i>khaa</i> ‘leg’
<i>hua</i> ‘head’ |
| Box 2 | <i>pii</i> ‘year’
<i>taa</i> ‘eye’
<i>kin</i> ‘to eat’ |
| Box 3 | <i>bin</i> ‘to fly’
<i>deej</i> ‘red’
<i>daaw</i> ‘star’ |
| Box 4 | <i>mii</i> ‘hand’
<i>khwaay</i> ‘water buffalo’
<i>naa</i> ‘ricefield’ |

Column B

- Box 5 *khay* ‘egg’
 phaa ‘to split’
 khaw ‘knee’
- Box 6 *paa* ‘forest’
 kay ‘chicken’
 kee ‘old’
- Box 7 *baa* ‘shoulder’
 baaw ‘young man’
 daa ‘to scold’
- Box 8 *phi* ‘older sibling’
 phɔɔ ‘father’
 ray ‘dry field’

Column C

- Box 9 *khaw* ‘rice’
 sia ‘shirt’
 khaa ‘to kill’
 khay ‘fever’
 haa ‘five’
- Box 10 *paa* ‘aunt (older sister of either parent)’
 klaa ‘rice seedlings’
 tom ‘to boil’
- Box 11 *baa* ‘crazy’
 baan ‘village’
 ?aa ‘to open (the mouth)’
- Box 12 *nam* ‘water’
 nɔɔŋ ‘younger sibling’
 may ‘wood’
 maa ‘horse’

Column D-short

- Box 13 *mat* ‘flea’
 suk ‘cooked, ripe’
 phak ‘vegetable’
- Box 14 *kop* ‘frog’
 tap ‘liver’
 cep ‘to hurt’
- Box 15 *bet* ‘fishhook’
 dip ‘raw, unripe’
 ?ok ‘the chest’
- Box 16 *nok* ‘bird’
 mat ‘to tie up’
 lak ‘to steal’

Column D-long

- Box 17 *khaat* ‘broken, torn’
 yiak ‘the gums’
 haap ‘to carry on a shoulder pole’
- Box 18 *pɔɔt* ‘the lungs’
 piik ‘wing’
 tɔɔk ‘to pound’
- Box 19 *deet* ‘sunshine’
 ʔaap ‘to bathe’
 dɔɔk ‘flower’
- Box 20 *mitt* ‘knife’
 luuk ‘(one’s) child’
 l̥at ‘blood’
 nɔɔk ‘outside’

This scheme allows us to compare modern languages to see how the original Proto Tai tone categories are realized. The important thing in this comparison is not the phonetic value of the tone; it is whether the tone of a cell is distinguished from that of certain other cells. Table 2.13 illustrates with a direct comparison of a fragment of Gedney’s tone boxes (only boxes 3, 2, and 1).

The table shows that different modern languages – one from each of the main branches of Tai that F. K. Li (1977) proposed – reveal different patterns of splits and mergers (see also

Table 2.13 *Modern reflexes in Tai languages of syllables in boxes 1–3 under the A category in Proto-Tai*

Original Proto-Tai initial	Meaning	Thai (SW Tai)	Longzhou (Central Tai)	Bo Ai Cun (Northern Tai)
Glottal or pre-glotalized stop (Gedney’s Box 3)	‘to fly’ ‘good’ ‘month’ ‘to take’	bin 33 dii 33 dian 33 pau 33	bin 33 dai 33 biin 33 pau 33	min 31 nii 31 niin 31 pau 31
Voiceless unaspirated stop (Gedney’s Box 2)	‘to go’ ‘year’ ‘mother’s father’ ‘to eat’	pai 33 pii 33 taa 33 kin 33	pai 33 pii 33 taa 33 kin 33	pai 24 pii 24 taa 24 kin 24
Aspirated voiceless stop or voiceless continuant (Gedney’s Box 1)	‘to sell’ ‘white’ ‘rain’ ‘dog’ ‘grandchild’	khaai 24 khaau 24 fon 24 maa 24 laan 24	khaai 33 khaau 33 phiin 33 maa 33 laan 33	kaai 24 haau 24 hin 24 maa 24 laan 24

After F. K. Li 1977: 29, which labels the languages ‘Siamese’, ‘Lungchow’, and ‘Po-Ai’.

Ostipirat 2000: 50–3). Thai (Southwestern Tai) merges boxes 3 and 2, as distinct from box 1; Bo Ai Cun (Northern Tai) merges boxes 2 and 1, as distinct from box 3; and Longzhou (Central Tai) merges all three boxes, realizing them with a single tone category.

The tone boxes approach is also used in the historical phonology of other branches of Tai Kadai, as well as in other language families, where applicable (e.g., in Karen; see Manson 2011: 4, Luce 1985). Ostapirat (2000: 50–106) supplies a large amount of data on tonal correspondences across modern languages of the Tai Kadai family (which he terms Kra Dai), showing that the ‘A B C tonal system’ indicates ‘excellent correspondences’ across the entire family.

2.4 The Chamic and Moklenic Branches of Austronesian

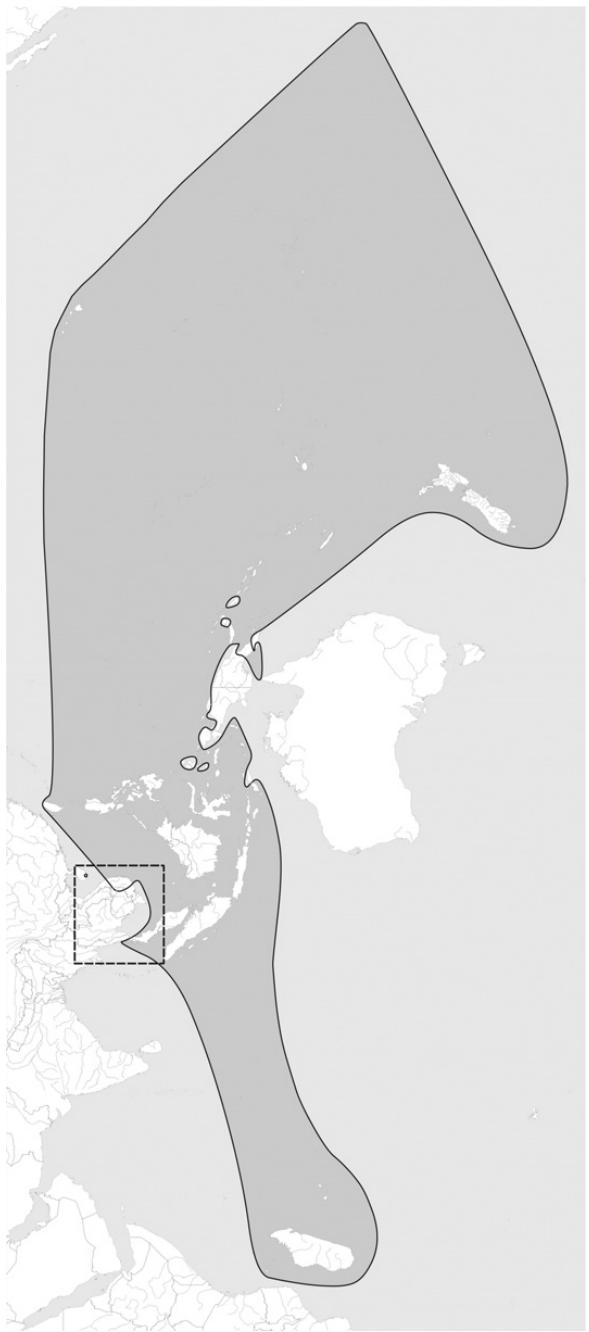
The Austronesian language family consists of some 1,200 languages, spoken by well over 300 million people spread across half the globe, from Madagascar in the west to Easter Island in the east, and from Taiwan in the north to New Zealand in the south (Adelaar 2005b: 1). Map 2.4 shows the approximate distribution of Austronesian languages.

The historical reconstruction of Proto Austronesian, the subgrouping of its descendants, and the reconstruction of mediating proto languages have occupied years of intensive scholarship, beginning with Dempwolff (1934–1938).

While the internal classification and subgrouping of Austronesian languages may be challenging or controversial, ‘the classification of particular languages as Austronesian is rarely problematic’ (Blust 2013a: 3). There are some ten primary subgroups of Austronesian. All but one of these are found in Taiwan: Blust (2013a: 30) names Atayalic, East Formosan, Puyama, Paiwan, Rukai, Tsouic, Bunun, Western Plains, and Northwest Formosan. The one family found outside Taiwan is Malayo Polynesian, which divides into numerous sub branches. These are associated with the great Austronesian migrations of the last few thousand years (Bellwood 1979, 2007, Bellwood *et al.* 2006). Figure 2.5 depicts the Austronesian family at a macro level.

Only a fraction of the languages of the Austronesian family are spoken in MSEA. All of them belong in just two of at least twenty three subgroups that make up the Western Malayo Polynesian branch of Austronesian.¹² One of these is Malayo Sumbawan, which includes the Chamic languages, the other is Moklenic. With Chamic and Moklenic taken together, there are twenty six Austronesian languages spoken in greater MSEA. See Map 2.5 for their approximate distribution.

¹² Western Malayo-Polynesian is a large grouping of languages that should not be taken to imply genealogical unity. There is no generally accepted evidence supporting the existence of a unique Proto-Western Malayo-Polynesian language.



Map 2.4 *Approximate distribution of Austronesian languages*

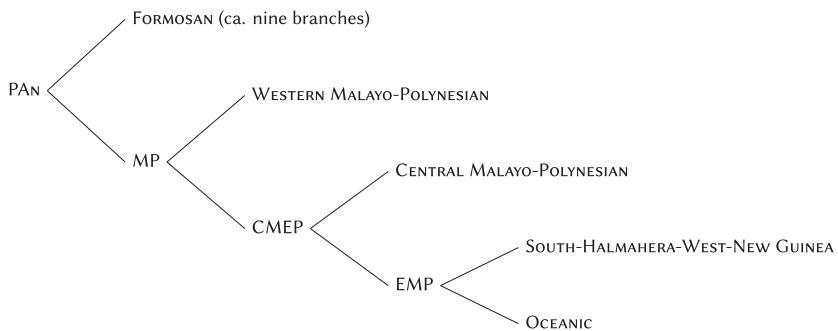


Figure 2.5 '*Tentative tree diagram of Austronesian languages*'

After Adelaar 2005b: 9; cf. Blust 1980, 1999.



Map 2.5 *Approximate distribution of languages of the Chamic and Moklenic branches of Austronesian*
The approximate distribution of languages of the Chamic branch lies to the east of the Gulf of Thailand, and those of the Moklenic branch to the west of the Thai/Malay Peninsula, in Burma/Thailand.

2.4.1 Chamic

Chamic languages spoken in MSEA include Phan Rang Cham (or Eastern Cham), Rade, Jarai, Haroi, Chru, Roglai (all spoken in Vietnam), Western Cham (spoken in Cambodia), and Tsat (spoken on Hainan Island). In early research on languages of this branch there was debate as to whether Chamic languages were Austronesian at all, owing to their typological resemblance to Austroasiatic languages (Schmidt 1906, Sebeok 1942). Today, Chamic languages are uncontroversially classified as Austronesian.

A close affinity between Chamic and Malayic languages was noted by Garrison (1975) and reinforced by Thurgood's (1999) reconstruction of Proto Chamic. But there are competing proposals regarding how this relation translates into local subgroupings. Blust (1994: 32) proposes that the closest linguistic relatives of Chamic are Acehnese and the Malayic languages. Blust proposed a Malayo Chamic subgroup, with two sub branches: Malayic and Achino Cham. This is shown in Figure 2.6.

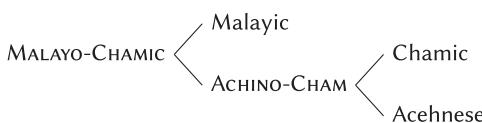


Figure 2.6 *Proposed Malayo-Chamic grouping*

After Blust 1994: 33.

Evidence in support of a Malayo Chamic grouping includes the following shared innovations (Blust 1994: 34, 40–1):¹³

Phonological

- PMP *q shifted to PMC *h
- PMP *R merged with PMP *r as PMC *r, but remained distinct from all other proto phonemes
- PMP *w disappeared in initial position
- as a consequence of the loss of initial *w, a historically secondary /h/ was introduced in the reflex of PMP *waRi ‘day’, initially in the construction *mata hari ‘sun’ (‘eye of the day’) (Blust 1994: 34)

Morphological

- disappearance of the PMP genitive marker *ni in the Malayo Chamic languages (Blust 1994: 40)

¹³ PMP Proto-Malayo-Polynesian; PMC Proto-Malayo-Chamic; PAN Proto-Austronesian.

Lexical

- replacement of Proto Austronesian numerals: *tuzuq ‘seven’ (not PAN *pitu), *dua lapan ‘eight’ (not PAN *walu), and *salapan/sambilan ‘nine’ (not PAN *siwa) (Blust 1994: 41)

An alternative subgrouping of Malay and Chamic is offered by Adelaar (2005b), in his Malayo Sumbawan subgroup proposal. This is shown in Figure 2.7.

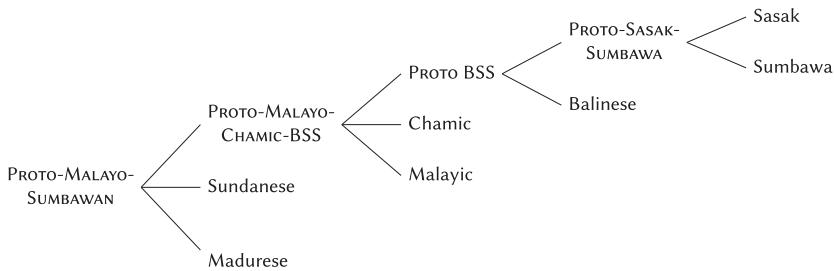


Figure 2.7 ‘*Malayo-Sumbawan*’

After Adelaar 2005a: 358.

Adelaar (2005a) presents evidence suggesting that BSS (Balinese, Sasak, Sumbawa) is closer to Malayic and Chamic than it is to other members of Austronesian. He argues that there is no Malayo Chamic unit, and indeed that, if anything, the evidence seems ‘more supportive of a link between Malayic and BSS alone than it does of a link that also includes Chamic’ (Adelaar 2005a: 385).¹⁴

While the higher level affiliation and subgrouping of Chamic languages remains open to discussion, the existence of a Chamic branch is well established.

2.4.2 Moklenic

The Moklenic group contains two languages: Moken and Moklen. These are closely related languages, though their speakers maintain both linguistic and cultural distinctions (Larish 2005: 513ff.; see Larish 1993). Moken speakers live on islands off the coast of Myanmar, while Moklen speakers live along rivers, beaches, and islands on the west coast of peninsular Thailand and Malaysia. The precise status of these languages in relation to their closest genealogical neighbours is still open. Larish (1999, 2005) posits Moklenic as a primary group within a higher level grouping together with Malay, Acehnese, and Chamic. By contrast, Thurgood (1999: 58–9) suggests that the likely

¹⁴ Blust (2010: 80–91) critiques this hypothesis and proposes a Greater Northern Borneo subgroup which includes Malayo-Chamic and some, but not all, of the languages which Adelaar (2005a) takes to be Malayo-Sumbawan.

phonological history of Moklenic languages sets them clearly apart from Malayic and Chamic (including Acehnese):

[Moken] does not share the PAn [Proto Austronesian] *q > *h change found in the whole of Malayo Chamic, a lack which sets it, not just outside of Chamic, but outside Malayo Chamic as well; instead, it has the change PAn *q > Moken /k/, a rather unusual reflex of PAn *q. Like Chamic, it does diphthongize both the high vowels in word final position, but in this its similarity to Chamic is only typological. The Chamic and Moken reflexes of the word final high vowels are split in entirely different ways in the two languages. Instead of PC [Proto Chamic] * εy (< PAn * i) and * ɔw (< PAn * u), Moken has uy, oi and ui, oi, respectively, at least suggesting that the PAn * i and * u have been merged in Moken. In any case, the Moken developments make it clear that Moken does not share in the Chamic diphthongization of word final high vowels nor in the PAn *q > PC /h/ change. In fact, thus far, although there are some typological similarities, there is no evidence that Moken shares any of the more marked inherited innovations characteristic of the Chamic languages.

(Thurgood 1999: 58 9)

Other differences include the following (Adelaar 2005b: 20):

- PAn *w became Moklenic b whereas *w was lost in Malayic
- PAn *w after following *a became u in Chamic
- PAn *R became Moklenic l but in Malayic and Chamic became r

2.4.3 Structural Change in Chamic and Moklenic

Thurgood (1999) presents evidence that Chamic languages underwent dramatic structural change due to prolonged contact with MSEA languages. Chamic languages have converged closely with the typological profile of MSEA languages and have diverged from that of Austronesian languages.

One important kind of structural change is the alteration of the Chamic basic word, from disyllabic to monosyllabic (Thurgood 1999: 60ff.). Greenberg (1970: 139) wrote that Phan Rang Cham and related languages Jarai, Rade, and Chru ‘tended toward loss or reduction of the vowel of the first syllable, thus producing extensive monosyllabism’ (cited in Thurgood 1999: 60). See section 4.8 below.

The second important kind of structural change is in the development of phonation type register systems, and in turn, lexical tone (Thurgood 1999: 178ff.). This

development is related to the changes in syllable structure just noted. Similar developments have been observed in Moklenic languages as well. Moken ‘may have adopted word final stress under Mon Khmer influence, and this single change could have served as a catalyst for a complete typological shift’ (Larish 1999: 381). Pittayaporn (2005) presents an alternative analysis of the developments in Moken away from the Austronesian structural type, arguing that contact with Austroasiatic languages may not have been the cause of the observed changes. We revisit these questions on contact induced structural change at the end of Chapter 4.

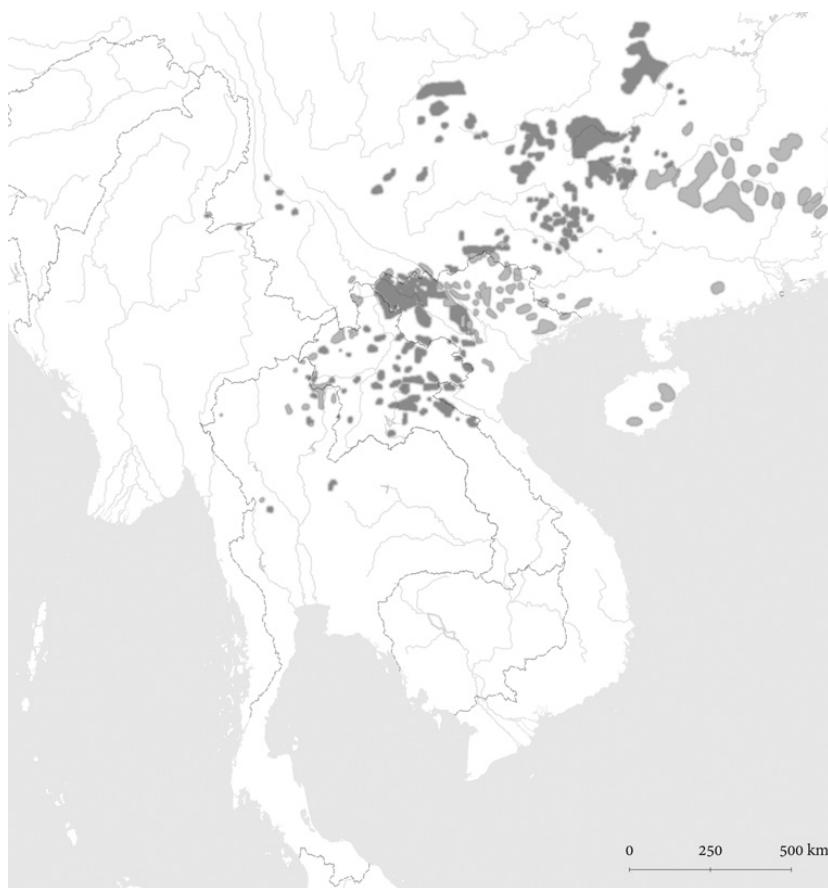
2.5 **Hmong-Mien**

Hmong Mien speakers in MSEA inhabit a range that is geographically small in comparison to the distribution of the other major language families we have reviewed in this chapter. An estimated thirty eight Hmong Mien languages are spoken in greater MSEA. See Map 2.6.

The total population of speakers of this family is fewer than seven million (Ratliff 2010: 3). This makes Hmong Mien by far the smallest of the major families represented in MSEA. The family is also the smallest in terms of the number of languages contained within it. Niederer (1998) compiles and reviews published work on Hmong Mien languages conducted prior to 1994, including scholarship in Chinese and French, with overviews of forty Hmong Mien languages. Ratliff (2010) provides a more recent overview.

The languages of the two branches of this family – Hmongic (sometimes called Miao) and Mienic (sometimes called Yao) – are spoken in many upland pockets throughout an area that includes north Vietnam, northern Laos, northern Thailand, and areas of the southwestern Chinese provinces of Guizhou, Hunan, and Yunnan (although some reside in Sichuan, Guangdong, Hubei, Jiangxi, and Hainan provinces).

The presence of Hmong Mien speakers south of the Chinese border is a recent phenomenon in MSEA history. Multiple southward migrations from Chinese territory took place from around the turn of the eighteenth century and continuing into the nineteenth and twentieth centuries due to pressure from Han Chinese population expansions inside China (Jenks 1994, Culas and Michaud 2004). Many thousands of Hmong Mien speakers live outside this range, a result of migrations over the last few decades, especially in the wake of the establishment of new governments in Laos and Vietnam in 1975, which put pressure on Hmong Mien populations.



Map 2.6 *Approximate distribution of Hmong-Mien languages in MSEA*
Dark grey Hmongic; light grey Mienic.

2.5.1 Comparative Historical Hmong Mien

Ratliff (2010) provides an authoritative introduction to the historical linguistics of the Hmong Mien family, within the context of her reconstruction of Proto Hmong Mien. In Chinese scholarship on Hmong Mien languages, it has been widely assumed that the languages belong to the Sino Tibetan language family. The same scholars take Sino Tibetan not only to encompass Chinese and Tibeto Burman languages but languages of the Tai Kadai family as well (Ratliff 2010: 1). However, the idea that languages of these three families – Hmong Mien, Sino Tibetan, and Tai Kadai – are part of a single family

‘is not widely shared by linguists outside China’ (Ratliff 2010: 1 2; see section 2.7 below). That said, numerous scholars outside China have proposed similar relationships. For example, Dower (1963), Haudricourt (1966), and Peiros (1998: 155 60) suggest that Hmong Mien languages and Austroasiatic languages share common ancestry. Benedict (1942, 1975) sought to group Hmong Mien languages with Austronesian and Tai Kadai in a macro family labelled Austro Tai. This, again, is not a generally accepted proposal. Ratliff offers caution:

Until a careful separation of layers of Chinese borrowings from native Hmong Mien vocabulary has been completed and the remaining core has been systematically compared to these other families, the question of wider relationship cannot be resolved.

(Ratliff 2010: 2 3)

One of the reasons that some linguists suggest these higher level relationships is the striking typological similarity between Hmong Mien languages and their neighbours. But because contact induced change can bring about the same similarities, the similarities alone cannot be taken as definitive evidence of common ancestry.

Pioneering work in historical/comparative Hmong Mien linguistics was carried out by Kun Chang, in a series of papers published between 1947 and 1976, focusing primarily on the tone systems of the languages (Chang 1947, 1953, 1966, 1972, 1976; cited in Ratliff 2010: 2 3). During the same period, Haudricourt (1954) published a comparative study of more than a hundred etyma in ten Hmong Mien languages. In 1970, Purnell published the first reconstruction of Proto Hmong Mien, as well as the two highest level descendant proto languages, Hmongic and Mienic, based on more than a thousand cognate sets (Purnell, 1970). This work has since served as a reference point in the field. Also in the 1970s, Wang (1979) presented a reconstruction of Proto Miao (Proto Hmongic). This was published fifteen years later (Wang 1994), followed by a reconstruction of Proto Hmong Mien (Wang and Mao 1995). (At around the same time, Thongkum (1993) published a reconstruction of Proto Mienic.) Ratliff states that Wang’s reconstruction ‘is unquestionably the most important contribution to Hmong Mien historical studies to date’ (Ratliff 2010: 4).

2.5.2 Hmongic versus Mienic

Within the Hmong Mien family, a highest level binary split between Hmongic and Mienic is generally accepted. Table 2.14 shows lexical evidence from four Hmongic varieties (Pa hng, Hmong North, East, and West) and two Mienic varieties (Iu men and Kim mun), showing a clear distinction between Hmongic and Mienic.

Table 2.14 Lexical evidence for a historical distinction between Hmongic and Mienic languages

	Pa-hng	HmongN	HmongE	HmongW	Iu-men	Kim-mun
‘white’	kwə (1)	qwə (1)	lu (1)	tʃeu (1)	pɛ (8)	pe (6) T!
‘hat’	ku (7)	ku (7)	—	kou (7)	lap (8)	gjap (8b)
‘pig’	mbe (5)	mpa (5)	pa (5)	mpua (5)	tug (4)	tur (4b)
‘comb’	vi (6)	zɑ (6)	ɣə (6)	zua (6)	tsa (7)	ta (7a)
‘spirit’	kwɪ (1)	qwei (1)	ɻəŋ (1)	tɻəŋ (1)	mjen (3)	man (3)
‘pot’	vĩ (4)	we (4)	vi (4)	za (4)	tshe:ŋ (1)	ten (1')
‘clothing’	?x (3)	?ə (3)	?u (6)	—	lu:i (1)	gui (1)
‘beard’	ŋi (6)	ŋi (6)	ŋaq (6)	—	sja:m (1)	tōɔm (1')
‘girl’	phe (7)	mpha (7)	phi (7)	ntshai (7)	sje (7)	sa (7)

After Niederer 2004: 132.

Table 2.15 Modern reflexes of Proto-Hmong-Mien tight (*NC-) versus loose (*N-C-) prenasalized obstruents

Proto-Hmong-Mien	Hmongic (White Hmong)	Mienic
*ntuət ‘navel’	ntau ⁷	dut ⁷ (Luoxiang Mien)
*mpeiH ‘to dream’	mpau ⁵	bei ⁵ (Mun)
*N-top ‘skin’	tau ⁵	dup ⁷ (Luoxiang Mien)
*N-peiX ‘soybean’	pau ³	bei ³ (Mun)

Data are from Ratliff 2010.

Ratliff reconstructs two kinds of prenasalized obstruent in Proto Hmong Mien: (1) tight (or fused) *NC and (2) loose *N C. These provide a basis for distinguishing Hmongic and Mienic languages. In Mienic, this distinction between the two kinds of prenasalized obstruent is collapsed. Both types yield simple voiced obstruent initials in modern Mienic languages. In Hmongic, by contrast, the Proto Hmong Mien tight prenasalized obstruents are preserved, while the Proto Hmong Mien loose prenasalized obstruents have simply lost their prenasalization. This is illustrated in Table 2.15.

Another distinguishing feature of Hmongic is a ‘dramatic merger of a large number of Proto Hmong Mien rimes into a small number of Proto Hmongic rimes’ (Ratliff 2010: 6). More than 125 rimes are reconstructed for Proto Hmong Mien. These collapsed ‘into only 28 rimes in the development of Proto Hmongic’ (Ratliff 2010: 6). As a result, Hmongic languages feature the smallest rime inventories of all MSEA languages (see Chapter 4).

Table 2.16 *Evidence for a correspondence between historical prefixes (also found in Tibeto-Burman) and modern voiceless initials in Hmongic languages*

Meaning	Proto forms
moon/month	TB *s/g-la HM *hlaH
sun/day	TB *s-nøy ‘sun’ Hmongic *hneŋ ^A ; Mienic *hnyɔi ^A
to slice	TB *s-lep ‘to slice’ HM *hlep
to sniff at	TB *s-nam ‘to smell (tr.)’ HM *hmjɔmH

After Ratliff 2010: 13. Tibeto-Burman forms are from Matisoff 2003a.

At the same time, Hmong Mien languages are notable for the large size of their inventories of initial consonants (see section 4.1). Large initial consonant inventories are in part due to ‘segmental and suprasegmental traces of old morphology’ (Ratliff 2010: 6). While these languages are monosyllabic, Ratliff’s reconstruction assumes that the proto language had a more complex structure. Ratliff reconstructs voiceless sonorant initials (such as devoiced nasals) as *hC sequences, rather than single unvoiced segments. This is intended to reflect the idea that a pre initial element was the historical cause of the sonorant devoicing. Ratliff (2010: 13) assumes that this pre initial element was likely an *s. She gives evidence from Tibeto Burman correspondences with an *s prefix, a multi purpose prefix derived ultimately from a root meaning ‘animal, flesh’ (see Benedict 1987: 44; see also section 2.6.3, below). Table 2.16 shows some examples.

The Hmongic branch includes three well established major groups: West Hmongic (Chuanqiandian), East Hmongic (Qiandong), and North Hmongic (Xiangxi) (Ratliff 2010, Wang and Mao 1995). A number of Hmongic languages have presented challenges for classification within this scheme, being placed differently within the family by different authors. These languages include Pa Hng, Xiong, Kiong Nai, and She (Ho Ne). These are undoubtedly Hmong Mien, but it is not clear where they fit within that classification.

The ambiguous position of She (Ho Ne) has even led some researchers to posit a third primary branch of Hmong Mien, alongside Hmongic and Mienic (see Wang and Mao 1995). By contrast, Ratliff (1998) presents evidence for classifying She as a Hmongic language, retaining the standard binary top level branching structure of Hmong Mien. Recent work is seeing the pendulum swing yet again: with ‘computer based lexicostatistics that utilizes Bayesian inference’ on 210 meaning items, Taguchi (2013) proposes that Xiong (Xiangxi) and PaHng are descended from a common ancestor that split from Proto Hmong Mien earlier than the rest of Hmongic. The question remains open.

While the Mienic group as a whole is relatively solid, research on Mienic languages has also seen controversies over lower level sub grouping. Wang and Mao (1995) and Mao (2004: 10) propose a four way split under Mienic, with Mien, Kim Mun, Biao Min, and Zao Min branches, as shown in Figure 2.8.

Thongkum (1993) presents a classification of a subset of the languages included in Wang and Mao’s model, focusing only on the Mien and Kim Mun branches in Figure 2.8 (omitting Biao Min and Zao Min). Her analysis differs from that in Figure 2.8 in that it takes Muen (aka Ao Yao, referred to in Figure 2.8 as Luoxiang vernacular in Wang and Mao 1995 and Mao 2004) and places it as a separate branching under Proto Mjuenic (= Proto Mienic), distinct from Mien and Mun. This is illustrated in Figure 2.9.

Ratliff (2010: 3) presents a more consolidated picture, positing a single Mien Mun subgroup (combining the Mien and Kim Mun subgroups from Wang and Mao 1995; see Figure 2.8), as a sister branch to Biao Min and Zao Min. This is illustrated in Figure 2.10.

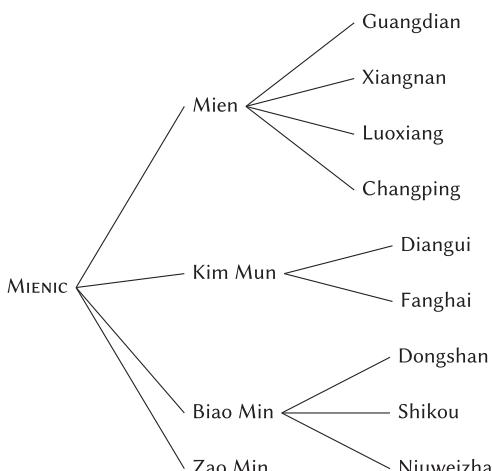


Figure 2.8 *Mienic subgrouping with four main subgroups*
As proposed by Wang and Mao (1995) and Mao (2004: 10).

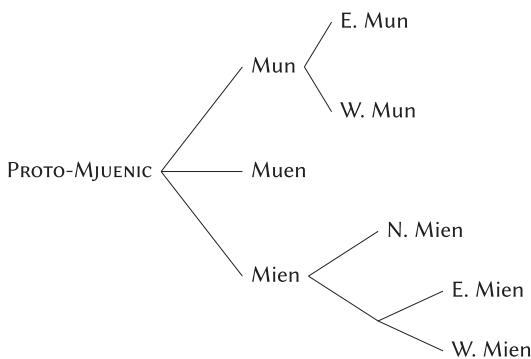


Figure 2.9 '*Classification of Mjuenic dialects and sub-dialects*'

After Thongkum 1993: 170.

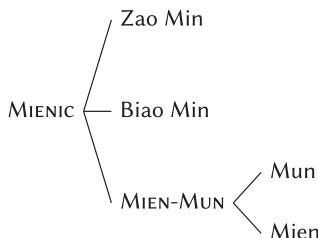


Figure 2.10 *Mienic, with three subgroups*

After Rathiff 2010: 3.

The three competing classifications of Mienic languages just cited were arrived at using the historical comparative method. We can note a further proposed version of the subgrouping of Mienic, offered by Hsiu (2018). This subgrouping, based on computational phylogenetic methods, is shown in Figure 2.11.

Languages in Hsiu's proposed Greater Biao Mon branch are classified as Mien by Wang and Mao (1995; also Mao 2004: 10), but Hsiu elevates them to a top level branch that descends directly from Mienic, as a sister of the Iu Mien and Kim Mun subgroups. This analysis partially overlaps with Thongkum (1993) in not considering Luoxiang Mien as a Mien vernacular, but belonging in a branch that is a sister of Mien.

The case of Hmong Mien again illustrates a general theme in MSEA historical comparative linguistics, reiterating what our review of other major families has shown. Scholars agree as to which languages fit in the family as a whole. And they generally agree as to certain clear mid level groupings, such as Hmongic versus Mienic. But controversy persists in relation to lower level groupings and possible internal relationships of groups within the family, as well as in relation to potential macro level relationships with other major language families. See below for further discussion of these general points.

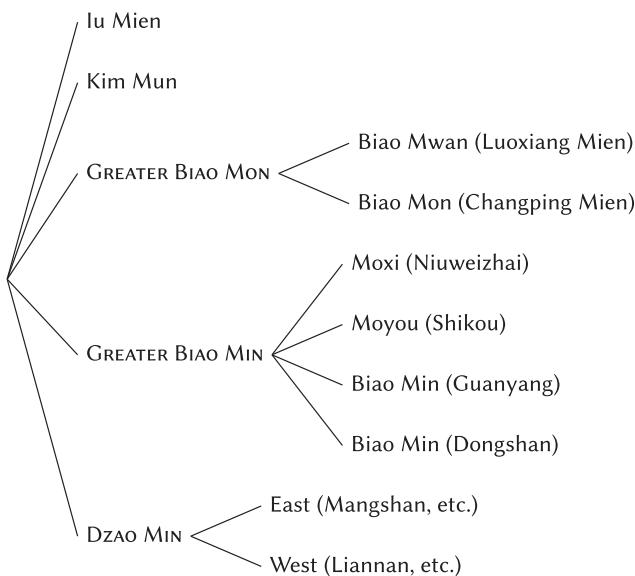


Figure 2.11 *Subclassification of Mienic languages*

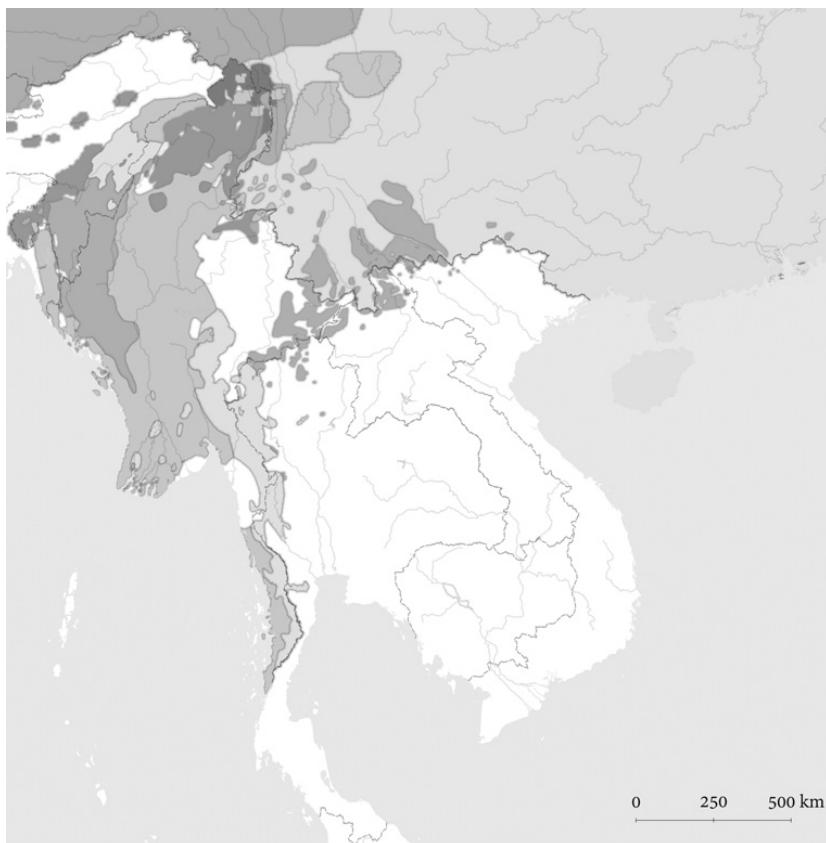
After Hsiu 2018.

2.6 Sino-Tibetan

Of the five language families discussed in this book, Sino Tibetan is the largest, most complex, and probably least well established at the highest level. Burling (1983: 1) remarked that ‘the subclassification of the Tibeto Burman languages has been a subject fraught with considerable mystery’. Three and a half decades later, an authoritative overview of the family is similarly tentative, with the subgrouping of Sino Tibetan languages said to be ‘still in its infancy’, despite decades of sustained and intensive work (Thurgood 2017: 31).

The full geographical spread of Sino Tibetan languages extends well beyond the limits of MSEA, reaching as far west as Kashmir (with the western extent of Tibetan) and further north than Mongolia, in the northeast of China (with the northern extent of Chinese). More than 280 Sino Tibetan languages are estimated to be spoken in greater MSEA. Map 2.7 shows the approximate geographical distribution of Sino Tibetan languages in and around core MSEA.

Not only does the Sino Tibetan family have the highest number of languages and the biggest population of speakers of the major language families in MSEA, it has more speakers than any other language phylum in the world. This is mostly due to the size of the Chinese speaking population. The current population of China is approaching 1.4 billion. Other notably large Sino Tibetan languages include Burmese (spoken as a first language by well over 30 million people) and Tibetan (with varieties spoken by up to 6 million people).



Map 2.7 *Approximate distribution of Sino-Tibetan languages in MSEA*
Sub-branches are represented in different shades.

Evidence for recognizing a Sino Tibetan language family comes from a range of sources. Some examples of the criteria that identify Sino Tibetan languages as a whole include the following (LaPolla 2012: 124):

- shared numerals
- voicing/aspiration distinction in initial consonants ‘marking intransitive vs. transitive or simplex vs. causative’
- **s* causative prefix
- **t* transitivizing suffix
- **n* and **s* formative suffixes

Most importantly, of course, there is lexical evidence for Sino Tibetan. Table 2.17 provides some example correspondences (and see Matisoff 2015a for a large set of comparanda and protoforms).

Table 2.17 ‘*Sino-Tibetan comparisons*’

	MC	OC	WT	WB	Bodo	Trung
‘I’	nguo	ngag	nga	ŋa	aŋ	ŋà
‘you’	n̄zjwo:	njag		naŋ	nəŋ	nā
‘not’	mju	mjag	ma	ma’		mà
‘two’	n̄zji-	njid	gnyis	hnac	nəŋ	ã-ni
‘three’	sam	səm	gsum	sūm	tam	ã-səm
‘five’	nguo:	ngag	lnga	ŋâ	ba	p̄əŋ-ŋjà
‘six’	ljuk	ljøkw	drug	khrok		khlu
‘nine’	kjau:	kjøgw	dgu	küi		da-gò
‘sun/day’	n̄zjet	njit	nyi-ma	ne		nì
‘tree/wood’	sjen	sjin	shing	sac		
‘year’	nien	nin	-ning	hnac		nip
‘name’	mjäng	mjing	ming	ə-maň	mun	
‘eye’	mjuk	mjøkw	mig	myak	megón	miè
‘ear’	n̄zi:	njøg	rna-ba	na	na:-	ã-nà
‘breast’	n̄zju:	njug	nu-mu	nui’		nuŋ
‘joint’	tsiet	tsit	tshigs	ə-chac		tsi
‘fish’	ngjwo	ngjag	nya	ŋâ	ná	ŋa
‘dog’	khiwen:	khwin	khyi	khwe		da-gai
‘insect/worm’	bjau	bjøgw	’bu	püi		bà ‘snake’
‘bitter’	kuho:	khag	kha	khâ	ká	kha ‘salty’
‘cold’	ljang	gljang	grang		gazaj	glanj
‘kill’	šat	srat	bsat	sat		sat
‘die’	si:	sjid	shi-ba	se	tøy	ei
‘poison’	duok	døkw	dug	tok		

MC Middle Chinese, OC Old Chinese, WT Written Tibetan, WB Written Burmese, Bodo a Bodo-Garo language in the Sal subfamily of Tibeto-Burman, Trung (Dulong) a Nungish language of the Central subfamily of Tibeto-Burman

The comparisons are given by Norman (1988: 13) to show lexical links across the family.

2.6.1 Sinitic

Sinitic languages are often referred to as ‘dialects of Chinese’. But they are not dialects in the straightforward linguistic sense of being mutually intelligible varieties. They are not all mutually intelligible, and so they should be regarded as separate languages. Sinitic languages can be grouped together on lexical grounds, as distinct from the rest of Sino Tibetan. We will use the term Tibeto Burman for all non Sinitic languages in this family, taken together. But this usage is not intended to entail that there is a highest level binary branching of the family between Sinitic versus the rest. That said, the belief that

there is this split has become ‘the standard view’ in Sino Tibetan linguistics (Matisoff 2015a: xxxi). This idea is notably represented in proposals by Benedict (1972, 1976b; see Figure 2.14 below). See Figure 2.12.

The standard view is not uncontested. Numerous authors suggest that Sinitic may be grouped with other branches that are currently under the Tibeto Burman rubric (see, for example, Blench and Post 2014: 92):

The standard view is now being questioned by certain scholars who claim that there are no clear innovations shared by Tibeto Burman as a whole with respect to Chinese, or that there is an especially close relationship between Chinese and a particular branch of Tibeto Burman. If we adopt such a point of view, that would make Sinitic just another branch of a superfamily we might call ‘Sino Tibeto Burman’ (STB).

(Matisoff 2015a: xxxi)

The position of Sinitic in the larger family tree aside, the idea that the Chinese languages form a group distinct from all Tibeto Burman families ‘is unambiguous and widely accepted, despite a dwindling number of older scholars who still see the connection as not yet proven’ (Thurgood 2003a: 6). Here are four example features that have been proposed to distinguish Sinitic versus Tibeto Burman languages (LaPolla 2012: 124):

- Proto Tibeto Burman has two series of stops, while Proto Chinese has four
- Negator *ta only reconstructs to Proto Tibeto Burman, not to Proto Chinese
- *^p initial negators occur in Proto Chinese, not in Tibeto Burman
- Lexical isoglosses (e.g., Proto Chinese *tin ‘sky’ vs. Proto Tibeto Burman *r məw)

The Chinese languages share core vocabulary which allows them to be related back to a common proto language. Some lexical comparisons across the family are presented in Table 2.18.

The Chinese languages are generally classified into three groups: Northern, Central, and Southern (Norman 1988: 183, Thurgood 2017: 7). This is shown in Figure 2.13. Other proposals, based on alternative criteria for classification, are reviewed by Tang (2018).

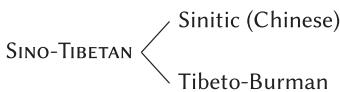


Figure 2.12 *The ‘standard view’ of Sino-Tibetan as consisting of two top-level branches: Tibeto-Burman and Sinitic*

After Thurgood 2017: 6, Matisoff 1991c: 470.

Table 2.18 Selected core vocabulary of Chinese varieties

Old Chinese	Middle Chinese	Mandarin (Beijing)	Wu (Suzhou)	Xiang (Changsha)	Min (Xiamen)	Cantonese (Guangzhou)	Gan (Nanchang)	Hakka (Meixian)	Jin (Taiyuan)	Ping (Guilin)	Hui (Jixi)
'I' ‘not’ ‘two’	* ⁵ ŋai? *pə *m[i]-s; *p rŋ?	ngwX piu ⁵¹ nyiH; ljang X	uo ²¹³ fɔ ²⁴ ŋ ²³¹ (casual), ɿŋ ²¹⁴	ŋo ⁴¹ pu ²⁴ ə ¹¹ ; lian ⁴¹ lian ²¹⁴	ŋua ⁵³ bo ²⁴ lŋ ²²	ŋo ²³ pat ⁵ t ²² , lœŋ ³⁵	ŋo ²¹³ pʰit ⁵ θ ²¹ , liŋ ²¹³	ŋai ¹¹ m ¹¹ ŋi ⁵³ , loŋ ³¹	ŋv ⁵³ pə ² ər ⁴⁵ , lia ⁵³	u ²³ pu ²² i ²¹	ɔ ²¹³ pʰp ³² ɛ ³¹ ; ŋi ²¹³
‘three’	*s rum	sam	san ⁵⁵	san ³³	san ⁴²	sam ³³	səŋ ²¹	səŋ ⁴²	sɔ ⁵⁵	v <u>²¹³</u>	v <u>²¹³</u>
‘five’	*C ʃa?	ngnX	u ²¹³	u ⁴¹	g ²²	ŋ ²³	ŋ ⁵³	ŋ ²³	ŋ ²³	ŋ ²³	ŋ ²³
‘tree’	*m-toʔ-s	dzyuH	zŋ ²³¹	ey ¹¹	tſiŋu ²²	sv ²²	su ⁴⁵	ey ²¹	ey ³¹	ey ³¹	ey ³¹
‘year’	*C n'iŋ]	nien ³⁵	ŋi ¹³	ŋiɛ ¹³	he ²² - t̪bau ²⁴	niŋ ²¹	ŋien ⁴⁵	niɛ ²¹	liɛ ²²	ŋiɛ ⁴⁴	ŋiɛ ⁴⁴
‘name’	*C meŋ	mjjeng	min ³⁵ tsŋ ⁰	min ¹³ tsŋ ⁰	mia ²⁴⁻ t̪au ²²	miaŋ ⁴⁵ tsŋ ⁰	mian ¹¹ tsŋ ⁵³	mijŋ ²¹ tsŋ ⁴⁵	met ²² tsŋ ²¹	miɛ ⁴⁴⁻³⁵	tsb ³¹
‘eye’	*[ŋ] ⁵ <ŋ>- [n] ²	ngeanX	ien ²¹³⁻²¹	ŋjan ⁴⁻⁴⁴ tein ⁰	bak ³⁻³²	ŋjan ²³ tsɛŋ ⁵⁵	ŋjan ³¹ tsu ³³	ie ⁵³ tein ⁰	ŋjɛ ²³ tsɛŋ ⁴⁴⁻	ŋjɔ ²³⁻	ŋjɔ ²³⁻
‘ear’	*C naŋ?	nyiX	teŋ ⁵⁵	u ⁴⁴ tsq ⁴⁴⁻²¹	tsiu ⁵⁵	ŋjan ²³ tsɛŋ ⁵⁵	ŋjan ³¹ tsu ³³	ie ⁵³ tein ⁰	ŋjɛ ²³ tsɛŋ ⁴⁴⁻	ŋjɔ ²³⁻	ŋjɔ ²³⁻
‘fish’	*f ⁵ ŋja?	ngio	y ³⁵	y ¹³	h ²⁴	y ²¹	ŋje ⁴⁵	ni ¹¹	ŋɔ ³³	ŋv ⁴⁴	ŋv ⁴⁴
‘dog’	*C ɔ kfrō?	kuwX	kou ²¹⁴	kaŋ ⁴	kau ⁵³	kaŋ ³⁵	kieu ²¹³	kau ⁵³	kai ²³	kai ²³	kai ²³
‘worm’	*C hrŋj	driuwng	tʂ ³⁵	tsəŋ ¹³	tʂŋ ²⁴⁻²²	tsʰŋ ²⁴	tsʰŋ ²⁴	tsʰŋ ²¹	tswŋ ²²	tswŋ ²²	tswŋ ²²
‘kill’	*s<r>at	sreat	ʂ ^A ⁵⁵	səf ⁴	ʈ̪ar ²⁴	səf ³	səf ⁵	sət ¹	sət ²	sət ⁵⁵	sət ²²
‘die’	*sij?	sijX	ʂ ¹ ²¹³	ʂ ¹ ⁵²	ʂ ¹ ⁴¹	ʂ ¹ ³⁵	ʂ ¹ ²¹³	ʂ ¹ ³¹	ʂ ¹ ²³	ʂ ¹ ²³	ʂ ¹ ²³

Words in Modern Chinese varieties are from Liu *et al.* 2007, reconstructions from Baxter and Sagart 2014

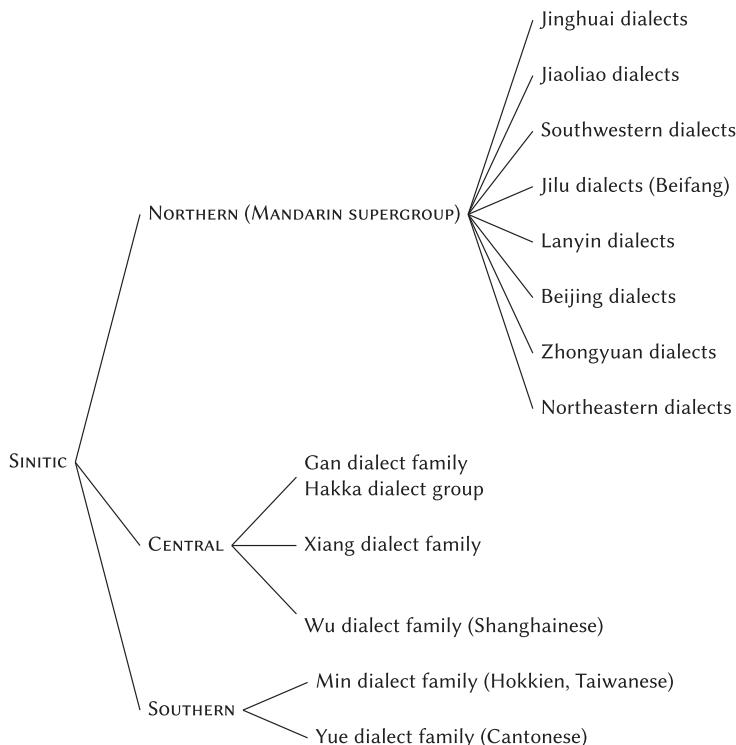


Figure 2.13 ‘Chinese “dialects” (*languages*)’

After Thurgood 2017: 7, updated from Norman 1988: 183.

Details of the history of Sinitic go beyond our scope here, given that the bulk of this family of languages is spoken outside of the MSEA area as we have defined it (with the exception of varieties of Southwestern Mandarin spoken in Yunnan, and Yue and Min varieties spoken in Guangxi and Hainan respectively). For background and references in relation to historical comparative Sinitic, with special attention to the unique role of written records in scholarship on the history of Chinese, see Chao (1941), Karlgren (1954), DeFrancis (1984), Ramsey (1987: 116–42), and Norman (1988: 23–57).

The phonological systems of Chinese languages vary widely, but it is often said that their grammatical systems are the same. Chao (1968: 13) stated that ‘there is practically one universal Chinese grammar’. The truth is that Sinitic languages show ample grammatical diversity, for example in question formation, descriptive complementation, comparative constructions, basic patterns of tense aspect modality marking, and more (see Chappell 2001).

2.6.2 Tibeto Burman Subgroupings

Numerous subgroups of Tibeto Burman are well established and generally accepted. But beyond certain subgroupings, there are few signs of consensus emerging in scholarship of these languages.

Bradley (2002: 74) refers to two classifications of Tibeto Burman languages that were ‘long established and widely cited’ by the end of the twentieth century. One is Shafer’s (1974) proposal of four main groups:

1. Bodic
2. Baric
3. Burmic
4. Karenic

The other was Benedict’s (1972: 5) proposal of seven ‘primary divisions or nuclei’ of Tibeto Burman:

1. Tibetan Kanauri (Bodish Himalayish)
2. Bahing Vayu (Kiranti)
3. Abo Miri Dafla (Mirish)
4. Kachin
5. Burmese Lolo (Burmish)
6. Bodo Garo (Barish)
7. Kuki Naga (Kukish)

Benedict’s comments on these divisions convey the diversity and widely varying texture of the family, and the ambiguity of certain languages and groups:

The seven divisions above range in diversity from the complex Tibetan Kanauri, Burmese Lolo, and Kuki Naga supergroups, each with a multitude of languages and dialects, through the fairly compact Bahing Vayu, Abor Miri Dafla, and Bodo Garo groups, down to Kachin, which consists only of the modern dialects of the language . . . Kachin, however, stands at the linguistic ‘crossroads’ of Tibeto Burman, thus occupying a linguistic position comparable with its geographical setting (northern Burma). Both lexically and morphologically, Kachin ties in with Tibetan, Bahing, and other northern languages as well as with Burmese, Bodo, Lushei, and other southern languages. From Kachin at this linguistic centre of diversification, transitions are afforded by Nung to Burmese Lolo on the east, and by the Konyak or ‘Naked Naga’ languages to Bodo Garo on the west.

(Benedict 1972: 5)

Benedict illustrated this conception of Sino Tibetan as shown in Figure 2.14.

Benedict's 'schematic chart' is striking for its lack of hierarchical organization below Tibeto Burman. Matisoff's late twentieth century proposals show a similarly star like or flat branching of Tibeto Burman groups (Matisoff 1991c: 481, 2003a: 5), with a top level Sinitic/Tibeto Burman split and an entirely flat structure under Tibeto Burman.

Thirty years after Benedict's *conspicetus*, with the benefit of new data from languages of China and northeastern India, Bradley (2002: 75) proposed a scheme with five branches, labelled Western, Sal, Central, Southeastern, and Northeastern. Bradley (2002) provides nuanced and helpful discussion of many of the languages within each of these higher level groups, as well as many of the issues and controversies around subclassification.

Historical classificatory surveys including Benedict (1972), Bradley (1997, 2002), and Matisoff (2015a) assume that with adequate data it should be possible to arrive at definitive, fine grained, multi level subgroupings which accurately assign each of the hundreds of Sino Tibetan languages to specified groups, subgroups, or sub subgroups. Bradley (2002: 110), for example, writes that 'every major subgroup awaits a full comparative treatment', suggesting that this will be possible when data gaps are filled.

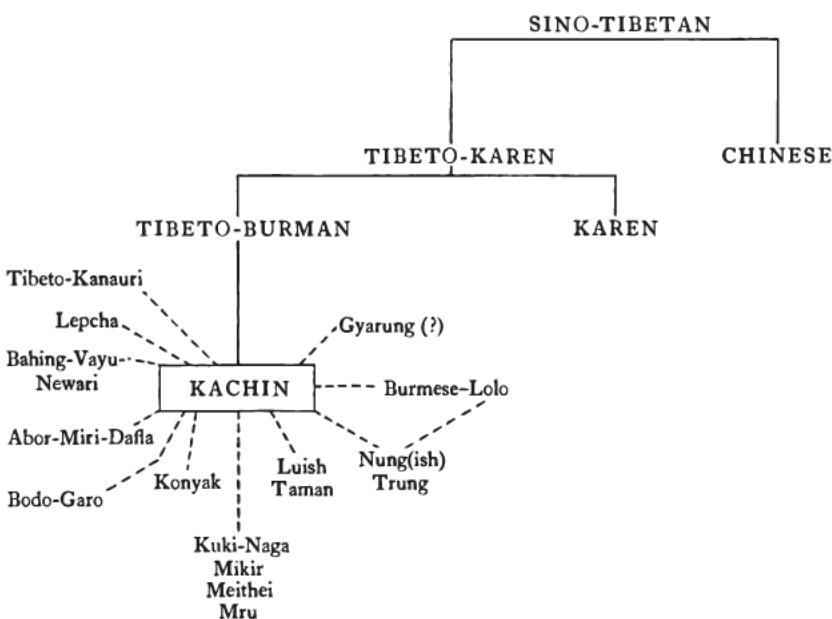


Figure 2.14 'Schematic chart of Sino-Tibetan groups'
Source: Benedict 1972: 6.

But discussion and commentary in these works make it clear that inadequate data is not the only problem. Certain fundamental limitations may mitigate against definitive fine grained subgroupings.

In Chapter 1, I described the processes of century upon century of migration, multilingualism, language contact, intergroup conflict and hegemony, displacement and assimilation of peoples, languages, and cultures in MSEA. This history of sustained and intensive disruption raises the possibility that no amount of new data will allow us to definitely assign certain modern languages to certain (sub sub) groups. One of the significant challenges of historical work throughout MSEA concerns the effects of historical contact between people speaking languages of distinct subgroups. At numerous points in his overview of Tibeto Burman subgrouping, Bradley (2002) remarks on contact effects. For example:

- ‘Most Himalayan languages have been heavily influenced by Indic languages’ (p. 76)
- The Bai language ‘has been heavily influenced for millennia’ by Chinese (p. 77)
- ‘The classification of some languages is uncertain, in most cases due to very extensive contact: Bai and Tujia with Chinese, the Nungish and Naxi languages with Burmese Lolo languages, Pepcha with a variety of Tibeto Burman languages, Dhimal with Kiranti and possibly other languages, Newari due to its position at the crossroads of Nepal, and so on’ (p. 75)
- ‘Some scholars … have suggested that part or all of the relationships between Sinitic and Tibeto Burman is of a contact nature’ (p. 75)

Certain languages, or language groups, may be inherently difficult to classify because they provide genuinely ambiguous signals. For example, some of the southernmost languages of Bradley’s Northeastern group ‘are lexically transitional to Burmese Lolo, but are phonologically and morphologically more typical of Northeastern’ (Bradley 2002: 77). This kind of situation is attested over and over across the family.

Still, some authors are willing to posit greater depth of nesting in the Sino Tibetan family tree than was seen in the earlier works cited above. Between 1991 when he remarked on how young the field of Sino Tibetan linguistics was (Matisoff 1991c: 469) and 2015 when a huge amount of new data had been brought to light, Matisoff became confident enough to propose a Sino Tibetan tree of greater depth and explicit internal structure than seen previously (see his ‘Updated subgrouping model of Sino Tibetan/Tibeto Burman relationships’; Matisoff 2015a: xxxii). Others have maintained a more tentative approach. Thurgood (2017) lists eighteen distinct subgroups of Sino Tibetan, in addition to a list of ungrouped languages (Lepcha,

Newar, Pyu, Tujia, and Bai). He does not offer groupings of these established families, ‘on the one hand because the lower level foundations are lacking, on the other because the necessary linguistic database is lacking’ (Thurgood 2017: 5). Here are Thurgood’s eighteen Tibeto Burman subgroups:

- Chinese
- Lolo Burmese
- Bodish (Tibetan, East Bodish, Tshangla, probably Tamangic)
- Qiangic
- Sal
- Northeastern Border
- Rung
- Rawang (Nungish)
- Kiranti
- West Himalayan
- Kham, Magar, Chepang
- Karenic
- Tani (Mirish)
- Kamengic
- Idu Digaru
- Kaman Meyor
- Koro Milang
- Raute, Raji

Van Driem (2013: 19) goes yet further, presenting the entire Tibeto Burman family (incorporating Sinitic) as a set of forty two subgroups arrayed two dimensionally. This avowedly agnostic approach to inter relations among established lower level groups elides any explicit claims to common ancestry, though it does imply both geographical and genealogical links through the relative placement of subgroups on the page (van Driem 2013: 19; cf. van Driem 2001, 2002).

Returning to the question of where Sinitic is placed, we note that the recent lists given by Thurgood and by van Driem both include Sinitic as one of a list of branches, rather than as an entity separate from all the rest of Tibeto Burman. Van Driem’s (1997) ‘Tibeto Burman hypothesis’ puts Sinitic under Tibeto Burman. In a recent restatement of the idea he uses the term Trans Himalayan (van Driem 2013; cf. Blench and Post 2014). Van Driem argues that there is evidence for a Sino Bodic subgroup, which places Sinitic well below its standard place as one of two primary branches of Sino Tibetan (van Driem 1997; for a critique, see Matisoff 2000a; for a response, see van Driem 2005). This is shown in Figure 2.15.

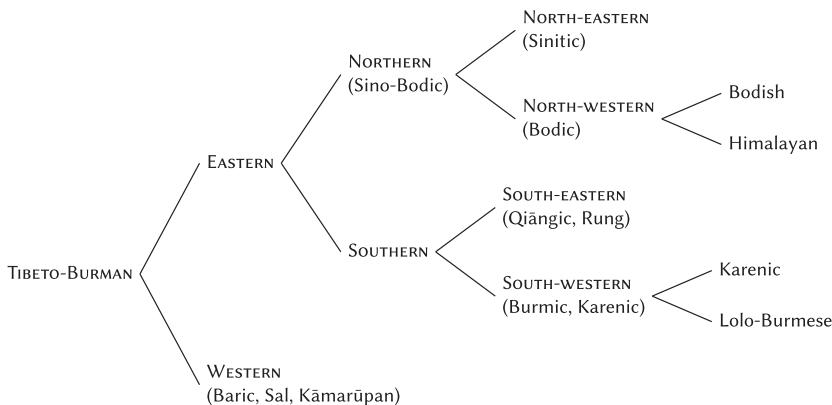


Figure 2.15 '*The Tibeto-Burman language family tree, reflecting the Sino-Bodic and Western Tibeto-Burman hypotheses*'

After van Driem 1997: 463. Note the placement of Sinitic deep in the tree, by stark contrast with 'the standard view' (cf. figures 2.12, 2.14).

In another recent proposal that places Sinitic far below its top level position counter to 'the standard view' Blench and Post (2013: 92) include Sinitic as a sister of Bodish and Lolo Burmish Naxi with a West Tibeto Burman branch. Finally, Sagart *et al.* (2019) present several alternative trees for Sino Tibetan, using both historical comparative linguistic methods and computational phylogenetic methods. While some of their models find that Sinitic is a distinct, top level subgroup, others do not. One model groups Sal and Sinitic languages together under a single branch (Sagart *et al.* 2019: 10320).

The trend suggests that the standard view of a highest level bifurcation between Sinitic and Tibeto Burman is more open to question than ever. Hence Matisoff's (2015: xxxi) advice on the question of Sinitic's place in relation to Tibeto Burman: 'It seems prudent ... to treat this as an open question for the time being'.

2.6.3 Affixes in Historical Tibeto Burman

Tibeto Burman languages – particularly on and beyond the northwestern fringe of greater MSEA – feature more affixal morphology than most core MSEA languages (see DeLancey 2011 and references therein). Wolfenden's pioneering work outlined the basics of morphology in Classical Tibetan and the languages of the Kachin, Bodo, Naga, Kuki Chin, and Burma groups (Wolfenden 1929).

Affixal morphology is relevant to a number of considerations in comparative historical Tibeto Burman linguistics.¹⁵ One question concerns the general character of

¹⁵ Thanks to Mark Post for helpful input here.

languages or subgroups as prefixing or suffixing (or isolating), and the directionalities and motivations of change. A second question concerns the reconstruction of affixes themselves to proto languages at various levels. This has met with some success, and there is now ‘broad agreement on the form and pattern of verb agreement in Proto Tibeto Burman’ (DeLancey 2011). A third question, related to the first two, concerns cycles of morphophonological change whereby root compounding evolves into prefixation with subsequent collapse into cluster/tone/voicing alternation (DeLancey 2011). While some modern Tibeto Burman languages show inflectional verb agreement (e.g., Nung, rGalryong, Eastern Kiranti, Kham Magar, and Kuki Chin; DeLancey 2011), many of the languages have minimal morphology but show evidence of phonological effects of the presence of affixation in earlier forms of the language.

An example is Lolo Burmese (Shafer 1966, Burling 1967, Matisoff 1972, Bradley 1979, Thurgood 1974, 1977, 1982, 2017). Thurgood (1977: 168), drawing on and refining an account by Matisoff (1972), explains that Proto Lolo Burmese had a two manner series of initial simple plosives. This is shown in Table 2.19.

The original plosive contrasts shown in Table 2.19 have evolved into distinct systems in modern Lolo Burmese languages. A first point of distinction between the modern languages concerns the values of the original distinction in voicing. In some languages, both proto forms are retained: in Akha, we see $*b \rightarrow b$ and $*p \rightarrow p$. In some languages the voiced stop became voiceless: in Lahu and Burmese, we see $*b \rightarrow p$. And in some languages the voiceless stop became aspirated: in Burmese, Lahu, and Lisu, we see $*p \rightarrow ph$. These correspondences are shown in Table 2.20.

Table 2.19 *Proto Lolo-Burmese initial simple plosives*

$*p-$	$*t-$	$*ts-$	$*c-$	$*k-$
$*b-$	$*d-$	$*dz-$	$*d\check{z}$	$*g-$

After Thurgood 1977: 168.

Table 2.20 *Modern reflexes of Proto Lolo-Burmese initial bilabial stops*

PLB	Written Burmese	Lahu	Lisu	Akha
$*p$	ph	ph	ph	p
$*b$	p	p	b	b

After Thurgood 1977: 169.

Table 2.21 shows example lexemes illustrating these correspondences.

However, in many cases the expected correspondences are not observed. The explanation has to do with the earlier presence of prefixes. These prefixes had simple and likely phonologically dependent forms, such as *a, *k, *sə, *m, *s, *b, and *r. In many modern reflexes of Proto Lolo Burmese forms, an apparently non regular correspondence can be explained by the historical presence of a prefix, which has merged with a lexical initial consonant. The proto prefix is no longer present as separate segmental material but has left behind an effect on the initial segment of the syllable, the tone of the syllable, or both. As an example, Akha retains the proto *k ‘animal prefix’ in many words,¹⁶ including *k'a hm*, ‘bear’, *k'a la*, ‘tiger’, *k'a pa*, ‘frog’ (Thurgood 1977: 152). In other cases, the prefix is gone, leaving behind a ‘distinctive tonal reflex’: ‘In Akha, when the *k preceded a checked syllable with an initial voiced stop, the reflex was a high rising laryngealized tone’ (Thurgood 1977: 152).

Table 2.22 shows an expanded form of Table 2.20, with the same Lolo Burmese languages, but with added rows for proto forms that featured either the voiceless fricative prefix *s or the voiced nasal prefix *m.

Table 2.21 *Lexical correspondences illustrating modern reflexes of Proto Lolo-Burmese initial bilabial stops*

Proto initial	Meaning	PLB	WB	Lahu	Lisu	Akha
*p-	OPEN	*pwaj ³	phwanj'	phɔ	phu	pahM
*b-	CHIN/CHEEK	*ba ²	pâ	pâ	bà	baL

After Thurgood 1977: 191, 198.

Table 2.22 *Modern reflexes of Proto Lolo-Burmese voiced and voiceless initial bilabial stops, with and without the *s- and *m- prefixes*

		Written		Lahu	Lisu	Akha
		PLB	Burmese			
1	*p	ph	ph	ph	ph	p
2	*s-p	ph	p	p	p	p
3	*s-b					
4	*b	p	p	b	b	b
5	*m-p	p	b	b	b	b
6	*m-b					

After Thurgood 1977: 170.

¹⁶ Incidentally, this prefix is argued to be borrowed from Austroasiatic (see Smith 1975).

Table 2.23 shows example lexemes illustrating the correspondences from Table 2.22.

Table 2.23 *Lexical correspondences illustrating modern reflexes of Proto Lolo-Burmese voiced and voiceless initial bilabial stops, with and without the *s- and *m- prefixes*

	Lexeme	PLB	WB	Lahu	Lisu	Akha
1	PRICE	*pəw ²	phūi	phū	phù	poeL
2	FROG	*s-pa ²	phā	pā	ó-pa	paH
3	ABLE	*s-brik	phrac	pí	pó	pyeuLS
4	THIN	*ba ²	pâ	pâ	bà	baL
5	SATIATED/FULL	*m-pup		bû?	bj	
6	CLEAR WEATHER	*m-ba ³	pa'	ba	ba	ba ('shine')

Data in rows 1–6 are from Thurgood 1977: 200, 194, 188, 204, 200, and 192 respectively.

2.7 Proposed Connections among the Families

In historical comparative linguistics in MSEA today, as illustrated in the previous five subsections covering work in the Austroasiatic, Tai Kadai, Austronesian, Hmong Mien, and Sino Tibetan families, we see a general pattern in the degree and kind of agreement or consensus among researchers. Experts in each major language family largely agree on (1) which languages do and do not belong in the family and (2) the existence of certain mid level subgroups. But they will often disagree as to (1) which languages fit where at lower levels of subgrouping and (2) the existence of relationships, if any, between established subgroups or families. As discussed already, scholarly disagreement partly results from limitations in the quality and quantity of available data, as well as from the fact that some languages are inherently difficult to classify. Effects of inter group contact and socio historical disruption are more pronounced in some languages, sub groups, and areas than in others. It is in the nature of some languages that their lexical and phonological structures provide mixed signals for the task of classification.

Disagreements associated with inadequate or inscrutable data relate not only to the placement of languages at lower levels in a family or in mid level subgroups, but also to possible connections above the level of established major language families. This takes us to the realm of putative macro families. The idea that the major language families of MSEA may share deep ancestry has long played a role in research and thought on the historical linguistics of the area. So far in this chapter we have worked in terms of five distinct, well established major language families. All known MSEA languages are readily identified as belonging to one or another of the families. But there are no agreed macro groups, despite numerous proposals of higher level relationships among the major language families.

Bradley (2002: 73) defines the ‘megalo’ approach to historical/comparative linguistics in MSEA as one in which the researcher tries to establish ‘very remote’ relationships between established language families. The megalο approach is ‘highly speculative’ and faces the great difficulty of separating ‘cognate material from contact material or chance resemblance’ (Bradley 2002: 73).

I will not go into any depth or breadth here on the various macro family proposals. The interested reader may wish to consult Reid (2005), Blust (2013a), and Ostapirat (2018) for reviews of some of the main claims. By way of illustration, we will consider only a few of the many proposals that have been made.

Prominent among megalο approaches are Sagart’s efforts over the last three decades to establish a relationship between Sinitic and Austronesian languages (Sagart 1990, 1993), broadening to link Sino Tibetan and Austronesian with Austric (the grouping of Austroasiatic with Austronesian), and suggesting that the macro family may include Hmong Mien and Tai Kadai as well (Sagart 1994). This line of inquiry naturally ends in the speculation of an all encompassing family with all five major MSEA families under a single family (see Starosta 2005: 183).

No attempt to link Austronesian to any other major MSEA language family has been generally accepted. Blust (2013a: 695–714) evaluates eight proposals of genealogical connections between Austronesian and other language families. He concludes that most of them have fatal methodological problems, with the exception of the Austro Tai and Austric proposals.

Austric refers to a proposed macro family that links Austronesian and Austroasiatic languages. This idea has its roots in the work of Schmidt (1906). When Schmidt was conducting his research, the constituent families were not yet established, and so his proposals are quite different in kind from similar ones today. He was simply trying to establish relationships where none had yet been proposed. By contrast, attempts to establish groups like Austric today are attempts to show relationships between groups whose existence is known to be solidly established. Some apparent morphological resemblances between Austroasiatic and Austronesian have been taken to be suggestive of a genealogical relationship (Blust 2013: 698), as have certain lexical resemblances (Diffloth 1994: 312ff.). After considering the available evidence for Austric, Diffloth concluded as follows:

The Austric hypothesis presents a real challenge to comparativists, and it will only receive its real test when we have full dictionaries of several languages in each branch of Mon Khmer. We are very far from having this at the moment, and all we can do is grope while blind folded and report that there seems to be some kind of object there.

(Diffloth 1994: 320)

Reid (2005) is open to the idea of a relationship between Austronesian and Austroasiatic that does not group them together as a unit. ‘The concept of “Austric” may eventually need to be abandoned in favour of a wider language family which can be shown to include both Austroasiatic and Austronesian, but not necessarily as sisters of a common ancestor’ (Reid 2005: 150).

The Austro Tai hypothesis proposes that Austronesian and Tai Kadai form a single genealogical unit. An early proponent of this idea was Benedict (1942, 1975, 1976a, 1990; for more recent work, see Ostapirat 2005, 2013, 2018). Again, the early macro proposals had a different status from today’s. They were made at a time when the major constituent families were not yet separately established. Benedict (1942: 588) cited the idea that Chinese and Tai constitute a single Eastern division of the Sino Tibetan or Indo Chinese stock, in opposition to a Tibeto Burman or Western division, and referred to this as ‘the generally accepted view’. As later commentators have shown, however, the view is not generally accepted.

Turning to relations between Chinese and families outside of Tibeto Burman, Benedict (1942: 589) notes numerous lexical resemblances between Chinese and Tai. Examples include certain numerals, body part terms, animal names, and some ‘cultural objects’. But Benedict does not accept these as evidence for a genealogical relationship between the languages. Rather, they are evidence of borrowing. Noting that the words for ‘horse’, ‘ride’, and ‘saddle’ are cognate in Chinese and Thai, Benedict remarks: ‘These correspondences strongly indicate that the Thai peoples borrowed the horse complex directly from the Chinese’ (Benedict 1942: 590).

Reviews of the various MSEA macro family hypotheses, including Reid (2005), Blust (2013a), Orlandi (2018), and Ostapirot (2018), review data that are as tantalizing as they are elusive, and endlessly malleable to the hopes of the beholder. The pursuit of speculative macro family hypotheses based on limited and ambiguous data is especially prone to confirmation bias. Blust (2013a) sums up this general point in relation to the claim that Chinese and Austronesian are related:

Most scholars have been left with the impression that the Sino Austronesian hypothesis is the product of an *idée fixe*: once the claim was made that Chinese and Austronesian are genetically related, no stone has been left unturned in trying to find further support for it.

(Blust 2013a: 711)

2.8 Problems and Challenges for Comparative-Historical Research

Comparative historical linguistic research on MSEA languages has attracted the sustained efforts of hundreds of researchers over many decades. Many of the pioneers in the

methods and theories of historical linguistics and related fields globally have worked on problems in MSEA linguistics. It is therefore little wonder that in the context of historical MSEA language research, a number of central issues in the methodology and theory of historical/comparative linguistics have been aired. It is useful to consider some of these issues, as they are relevant not only to MSEA linguistics but to historical comparative linguistics more broadly.

A methodological point concerns the status of external evidence in reasoning about language history. Such external evidence might include dates and locations for technological innovations such as iron, livelihood practices, cultivation of millet, or the use and domestication of animals such as the yak or the horse. Blench and Post (2014: 74) argue that reconstructions offered without consideration of this kind of non linguistic evidence are ‘unjustified and potentially misleading’. Most historical linguists would not disagree with this point but might express it with a different emphasis. Thurgood (2017: 3), for example, remarks that ‘only linguistic data constitutes evidence for a linguistic subgrouping – not geography, not ethnography, not folklore’, but then goes on to acknowledge the linguist’s accountability to non linguistic facts: ‘Of course if the resulting subgrouping is at variance with known history, for instance, either or both should be carefully re examined’ (Thurgood 2017: 3).

LaPolla (2012) outlines several factors that he argues are hampering progress in the historical linguistics of the Sino Tibetan area. One of these concerns epistemological access and standpoint. A researcher inevitably knows some languages and language families better than others. We learn some languages first, and they serve as reference points when we learn other languages later. This introduces biases through the ways in which we use old information to make inferences over new information. So, it is little wonder that with such large and sprawling language families, we see significant disagreement in analysis and emphasis, in part associated with the personal path that each scholar took in coming to the language family.

With respect to ‘subgrouping and reconstruction work’, LaPolla (2012: 120ff.) cites three problem areas. The first concerns data. Besides the obvious lack of adequate data, there is also ‘misuse of data that there is by people trying to make a particular point’. The second problem concerns geography. There is a readiness to assume that if languages are spoken in close geographical proximity, they are more likely to be closely related. When researchers label a language family or subgroup with a geographical label – for example ‘Western Himalayan’ – the designations ‘assume that either the languages involved have always been at that location or that all the languages developed from a single ancestor which migrated to that location at some time in the past’ (LaPolla 2012: 120). Yet as LaPolla has outlined in relation to many Sino Tibetan populations, ‘we know that there have been waves of migration, particularly into and/or through certain regions’ (LaPolla 2013: 464).

The third problem is ‘lack of consistent and clear standards and principles for subgrouping’:

Though there is some excellent work done using the comparative method, and there have been arguments for more rigorous application of the comparative method (using sets of unusual shared innovations Thurgood 1982), subgrouping within Sino Tibetan is often based on certain features that the languages are said to share, or on a few shared lexical items, or even on the fieldworker’s intuitions, or on how remote speakers feel different languages are (the degree of mutual intelligibility), or, as we saw above, because the languages just happen to be in the same geographic area.

(LaPolla 2012: 121)

The key challenge is the constant difficulty of achieving and maintaining some degree of objectivity. Not only is this due to researchers’ individual biases and intuitions, it is also an outcome of sociopolitical bias (no doubt often unconscious in nature) in favour of ‘citations from “major” languages, and reference to epigraphic and written sources for earlier forms of certain such languages, e.g., Chinese, Tibetan, and Burmese’ (Blench and Post 2014: 72). Blench and Post (2014: 73) make this point in their critique of ‘the standard view’ in Sino Tibetan linguistics – by which the Sinitic group stands apart from the rest – arguing that as there is no ‘unambiguous published evidence’ to support the view, rather ‘Sinitic was likely to have been historically set apart for cultural and/or other non linguistic reasons’. The pull of major, well known languages is as great as the need to resist that pull. Luckily, the biases that we naturally bring to research can be largely overcome by collectively adhering to agreed scientific principles of empirical and analytical method and argument.

2.9 Signed Languages

Sign languages of the deaf/Deaf develop in various forms worldwide wherever there are deaf people (see Pfau *et al.* 2012 for a detailed overview of sign language linguistics). Sign language is not the same as hand gestures used alongside spoken language (Stokoe 1960, Emmorey 2001; cf. Goldin Meadow 2003a, Kendon 2004; for research on gestures in Laos see Enfield 2001, 2003, 2004, 2009). Nor is it the same as the finger spelling systems that many sign language users employ to denote words from local spoken/written languages.¹⁷ A general message from the scant literature about sign

¹⁷ The MSEA languages referred to in this section have means of finger-spelling words from local spoken written languages such as Burmese, Khmer, and Thai.

languages of the deaf/Deaf in MSEA is that little is known. We report here on limited information that there is, organized around the four basic kinds of sign language system that can be recognized.

2.9.1 Home Sign Systems

Home sign systems are highly localized systems in which non hearing people and their immediate family members or neighbours innovate a manual system of communication, enabling non hearing and hearing people to communicate typically, a deaf child and her hearing caregivers (Goldin Meadow 2003b). In these situations, home signs systems are created *de novo*, combining ad hoc innovations, hand gestures, and general principles of iconicity to create locally conventionalized ways of communicating exclusively in the visual modality using the hands, face, and body. While home sign systems are not full blown linguistic systems, they share many important properties with fully developed sign languages.

Home sign systems are highly localized in both space and time. They can have a similar status to within family idioms. They disappear as quickly as they appear. They are tied to the biographies of the family members involved. Their transience is associated with the low frequency of deafness in human populations. Approximately one person out of every thousand is born profoundly deaf or becomes profoundly deaf at an early age (Reardon *et al.* 2004: 8). This means that while many villages may have one or two deaf individuals, in the kinds of non centralized societies that have characterized MSEA for the vast majority of its history, there would be no critical mass of deaf populations to enable a home sign system to conventionalize and form into a language that could then be passed on across generations.

Without doubt, many thousands of home sign systems must have appeared and disappeared across MSEA over the thousands of years that humans have populated the area, but we have no evidence of them. There are no studies of these systems in MSEA to date.

2.9.2 Village Sign Systems

Because deafness can be hereditary, sometimes a community may naturally have an unusually high proportion of profoundly deaf individuals living together. In these cases, a group of non hearing people can form a signing community, communicating directly with each other and developing a linguistically complex system over time (Zeshan 2004, 2006, Zeshan and de Vos 2012). In addition, hearing people in the same community may learn and use the language as well (Nonaka 2004, 2007, 2009).

In MSEA, one example of this is the sign language of Ban Khor, a rural village situated in Nakhon Phanom Province of northeast Thailand. The rate of congenital hearing loss in this village is much higher than the average incidence of 1/1,000. At least 16 out of fewer than 3,000 people in this village are reported to have congenital hearing loss (Nonaka 2009: 214).

Ban Khor Sign Language is recognized as an indigenous sign language of Thailand. It has no affiliation with other sign languages of the country. ‘Ban Khor Sign Language patterns differently from other sign languages in Thailand because the history of deaf people in Ban Khor is not shared by Thai deaf people who use [other sign languages of Thailand]’ (Woodward 2003: 290):

Ban Khor Sign Language has developed over the last 80 years as a unique response to a substantial increase in the number of deaf villagers. Some villages in Ban Khor such as Mu 2 and Mu 10 have more than one deaf person per 100 people, which is from five to ten times the expected population of one to two in every 1,000 and from five to ten times the actual percentage of deaf people in the other six communities [of sign language users in Thailand and Vietnam]. Deaf and hearing people in Ban Khor have chosen to respond to this increase in ways that are similar to responses of other small communities found in many parts of the world with similarly large proportions of deaf people . . . The great majority of hearing people in Ban Khor have adopted either neutral or positive attitudes toward deaf people. As a result of these attitudes, hearing people and deaf people have developed a purely indigenous sign language for use in the local area – a sign language that developed without any significant outside contact and that is maintained without outside contact and influence. This sign language is used by the overwhelming majority of deaf people in Ban Khor as their first and only language and by the large majority of hearing people bilingually. Ethnographic field observations would suggest that, although both hearing and deaf people are aware of other sign languages in Thailand, they have no desire to change or replace their indigenous sign language for one that did not develop inside their own local community.

(Woodward 2003: 289–90)

Subsequent work by Nonaka (2004, 2007, 2009) has looked in considerable depth at the structure and dynamics of the speech/sign community in Ban Khor and has offered descriptions of elements of the linguistic system, including toponyms (Nonaka 2015) and interrogatives (Nonaka 2010).

Nonaka (2004) highlights the status of Ban Khor as both indigenous and endangered. She shows that the language is under threat from the shift to Thai Sign Language (Nonaka 2014). Nonaka's research on Ban Khor Sign Language has been part of a recent set of advances in scholarship on these previously undocumented languages worldwide. It has helped to derive some general principles for understanding how these types of systems develop, and in what circumstances:

The communities where indigenous sign languages have appeared are geographically, culturally, and linguistically diverse, yet they share a remarkably similar constellation of socioeconomic and demographic features, including:

- (1) high degrees of real/biological or fictive/nonbiological kinship
 - (2) labor intensive, pre industrial local economies
 - (3) low intra community educational differentiation between deaf and hearing people
 - (4) low intra community occupational differentiation between deaf and hearing people.
- ...

Three things at the nexus of language, language ideology, and language practice, are the hallmarks of speech/sign communities, namely:

- widespread fluency in the local sign language among deaf and hearing people
- neutral to positive attitudes toward deafness, deaf people, and sign language
- a high degree of integration of deaf people into the mainstream of village life

(Nonaka 2009: 212–13)

There is no doubt that these conditions have applied in many cases in MSEA over thousands of years. But we have no records. Again, the sociometrics of such low populations in the context of a 1:1,000 baseline ratio between deaf and hearing people means that even a strong village sign system is unlikely to achieve the critical mass needed for a full blown sign language to emerge and be transmitted indefinitely.

2.9.3 Alternate Sign Languages

In special circumstances, a visual form of communication that resembles sign language may emerge among hearing people. Alternate sign languages are ‘developed by people already competent in some spoken language . . . for use as an alternative to speech in

circumstances where, for whatever reason, speech is not used' (Kendon 1988: 4). For example, in Central Australian language communities, forms of alternate sign language have developed for use in interactions among kinspeople who have an avoidance relationship, and in relation to restrictions on speaking during periods of mourning (Kendon 1988, Green and Wilkins 2015). In native North America, an alternate sign language known as Plains Indian Sign Language emerged in the Great Plains Cultural Areas of Native North America (United States and Canada) and served as a lingua franca (Farnell 1995, Mithun 1999: 292–4, Davis 2010, 2015). In monastic traditions such as that of the Benedictine monks who hold a vow of silence, alternate sign languages have emerged for communication (Quay 2015, Banham 2015). Sawmill Sign Language was developed in North America for communication in environments too noisy for spoken language (Meissner and Philpott 1975). And special sign languages have emerged in the context of hunting, allowing hunters to communicate without making noise that might alert potential prey (see Mohr 2015 on the Tshaukak'ui hunting signs of the Ts'ixa people in northern Botswana).

The literature on alternate sign systems does not include any reference to MSEA communities. It is unknown if any alternate sign language systems exist in the area.

2.9.4 National Level Sign Languages

The 2015 handbook *Sign Languages of the World* provides information for core MSEA countries, along with China and Hong Kong (Jepsen *et al.* 2015: 23–6).¹⁸ See Table 2.24.

Table 2.24 Data on selected sign languages of core MSEA

Ban Khor SL	Thailand	No estimate available
Cambodian SL	Cambodia	Possibly around 13,000 deaf, of which 1,500 are estimated to be signers
Chinese SL	China	20+ million deaf and hard of hearing unknown how many use SL
Ha Noi SL	Viet Nam	Up to 39,000 users
Hai Phong SL	Viet Nam	Up to 1,800 users
Ho Chi Minh City SL	Viet Nam	Up to 45,000 users
Hong Kong SL	Hong Kong	20,000
Laos SL	Laos	No estimate available
Modern Thai SL	Thailand	Up to 67,000 users
Original Bangkok SL	Thailand	Up to 500 users
Original Chiang Mai SL	Thailand	An estimated 19 deaf signers

After Jepsen *et al.* 2015: 23–6.

¹⁸ A significant amount of research has been done on Chinese Sign Language and Hong Kong Sign Language (see Yang 2015, Tang 2007, 2015), but as that work does not focus on the MSEA area as such, we concentrate in this section on sign languages of the countries of core MSEA.

Jepsen *et al.* (2015) features brief overview articles on Cambodian Sign Language (Woodward *et al.* 2015a), Chinese Sign Language (Yang 2015), Ha Noi Sign Language (Woodward *et al.* 2015c), Hai Phong Sign Language (Woodward 2015), Ho Chi Minh City Sign Language (Woodward *et al.* 2015b), Modern Thai Sign Language (Woodward *et al.* 2015d), Original Bangkok Sign Language (Woodward and Suwanarat 2015), and Original Chiang Mai Sign Language (Woodward and Wongchai 2015). Laos Sign Language is listed, but no information is given. Myanmar is not mentioned (see discussion of Myanmar sign language below). Some typological literature makes reference to sign languages of MSEA. For example, Wilkinson's global typological study of kinship terminology in sign languages includes Ho Chi Minh City Sign Language and Thai Sign Language (see Wilkinson 2009: 391–405).

At a general typological level, there is no indication that MSEA sign languages display unique or surprising features. Woodward *et al.* (2015a) remark that Cambodian Sign Language, 'like all sign languages, has a sub lexical level of structure analogous to but not dependent on the phonological structure of spoken languages':

Handshapes, orientations, locations, and movements, and non manual expressions follow expected patterns found in other sign languages ... [Cambodian Sign Language] exhibits all the common phonological processes and changes found in the world's signed and spoken languages: fluidity, deletion, assimilation, as well as the less common process of epenthesis.

(Woodward *et al.* 2015a: 163)

Cambodian Sign Language is SOV in simple clauses and head initial in other kinds of phrase (Woodward *et al.* 2015a: 171–2). Content question words occur in final position (Woodward *et al.* 2015a: 174). These are widely attested features in sign languages globally. At the same time, Cambodian Sign Language has some unique features:

There are some handshapes that are not commonly found in the world's sign languages. In particular there is a handshape that uses the ring and pinky fingers with the thumb extended. In addition, there are also two complex R handshapes that have a space between the index finger and the mid finger.

(Woodward *et al.* 2015a: 163)

There are no indications from the rather minimal literature that the general profile of Cambodian Sign Language is different from that of other sign languages of MSEA.

Thailand

Modern Standard Thai Sign Language (MSTSL, or TSL for Thai Sign Language) is associated with the National Association of the Deaf in Thailand and is used across the country. ‘It is generally recognized that MSTSL has been influenced by signs from ASL [American Sign Language], which were brought to Thailand in the 1950s by some hearing Thai individuals who were interested in formal education for deaf individuals in Thailand’ (Woodward 1996: 228). Woodward (1996) conducted a cognate comparison test directly comparing MSTSL and ASL. He reports a ‘57% rate (42/74 pairs) of possible cognates between ASL and MSTSL’. He concludes on this basis that ‘ASL and MSTSL should be classified as distinct languages that are closely related historically and that belong to the same language family’ (Woodward 1996: 231). Cognate comparisons with older generation signers in Bangkok and Chiang Mai suggest that ‘there were at least two original sign languages in Thailand prior to ASL influence: Original Chiangmai Sign Language (OCMSL) and Original Bangkok Sign Language (OBSL)’ (Woodward 1996: 239).

Woodward (2000) later included the village sign language of Ban Khor in this comparison. Woodward’s conclusion is that there are ‘four separate languages that belong to three separate language families’. These are illustrated in Figure 2.16.

Modern Thai Sign Language is not grouped with the original sign languages in Thailand in Figure 2.16. Woodward attributes this distinctness to the intensive contact effects of ‘the introduction of vocabulary from ASL into schools for the Deaf in Thailand’, resulting in a high ‘rate of cognates between basic vocabulary in Modern Thai Sign Language and ASL’:

Given the great amount of foreign contact and borrowing that has influenced Modern Thai Sign Language’s development and use and the lack of such contact and borrowing in other sign languages in Thailand, there should be little doubt why Modern Thai Sign Language is not closely related to any other sign language in Thailand and why it belongs to a separate language family from any other sign language in Thailand.

(Woodward 2000: 31)

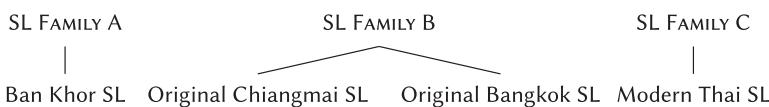


Figure 2.16 ‘Sign languages in Thailand classified by traditional language family trees’
After Woodward 2000: 29.

Vietnam

Three sign languages of Vietnam – Ha Noi Sign Language, Ho Chi Minh City Sign Language, and Hai Phong Sign Language (see Woodward *et al.* 2015b, 2015c, Woodward 2015) – are historically related, with evidence that they are descended from a common ancestor. See Figure 2.17.

A key event in the history of the sign languages of Vietnam was the establishment in 1886 of the country's first school for deaf people, in Lai Thieu, Binh Duong Province (Woodward *et al.* 2015b: 392). French Sign Language was introduced and became mixed with the original sign languages used in southern Vietnam at the time. The Lai Thieu school was the only school for deaf people in the country for the next ninety years. In 1976, a second school was established in the north, in Hai Phong. Between 1945 and 1976, deaf people in Hai Phong had no contact with the signers of south Vietnam, due to the political partitioning of north and south Vietnam, and had no access to deaf education (see Woodward 2015).

Thailand/Vietnam Comparison

Woodward (2000) extends his earlier lexical comparison to include the Thailand and Vietnam sign language data together. He finds that Modern Thai Sign Language is more closely related to sign languages of Vietnam than it is to either Original Bangkok Sign Language or Original Chiang Mai Sign Language, both used in Thailand:

The relation of Modern Thai Sign Language to sign languages in Viet Nam is in fact not a result of direct contact but of indirect contact. Ha Noi Sign Language, Ho Chi Minh Sign Language, and Hai Phong Sign Language all show very strong influences from French Sign Language, which was introduced into schools for the Deaf in Viet Nam. French Sign Language and ASL have a 61% rate of cognates in basic vocabulary and therefore belong to the same language family (Woodward, 1978). Thus, the influence of ASL on Modern Thai Sign Language and the influence of French Sign Language on Ha Noi Sign Language, Ho Chi Minh Sign Language, and Hai Phong Sign Language result in a large number of shared cognates between Modern Thai Sign Language and sign languages in Viet Nam.

(Woodward 2000: 31)

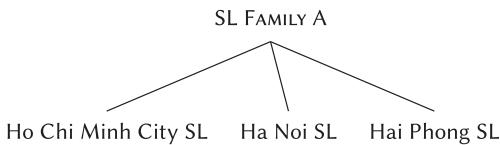


Figure 2.17 'Sign languages in Vietnam classified by a traditional language family tree'
After Woodward 2000: 29.

Woodward (2000) also finds that Hai Phong Sign Language is more closely related to the two original sign languages of Thailand than it is to other sign languages of Vietnam. He suggests that Hai Phong sign language shows a higher degree of cognacy with the original Thai sign languages (which he states '*have not* been influenced by French Sign Language or ASL'; 2000: 32) because of greater isolation. He speculates that there were 'original Southeast Asian signs' (2000: 32) and that these are preserved in the older sign languages across the two countries.

Hai Phong Sign Language is ambiguous in its affiliation, with a relatively high level of shared vocabulary not only with the two other Viet Nam sign languages (Ha Noi and Ho Chi Minh City) but also with the two Old/Indigenous Thai sign languages (Bangkok and Chiang Mai), with cognate levels around fifty per cent. Woodward hypothesizes that Hai Phong Sign Language is a 'link language' between indigenous Thai sign languages and other sign languages of Vietnam. This is illustrated in Figure 2.18.

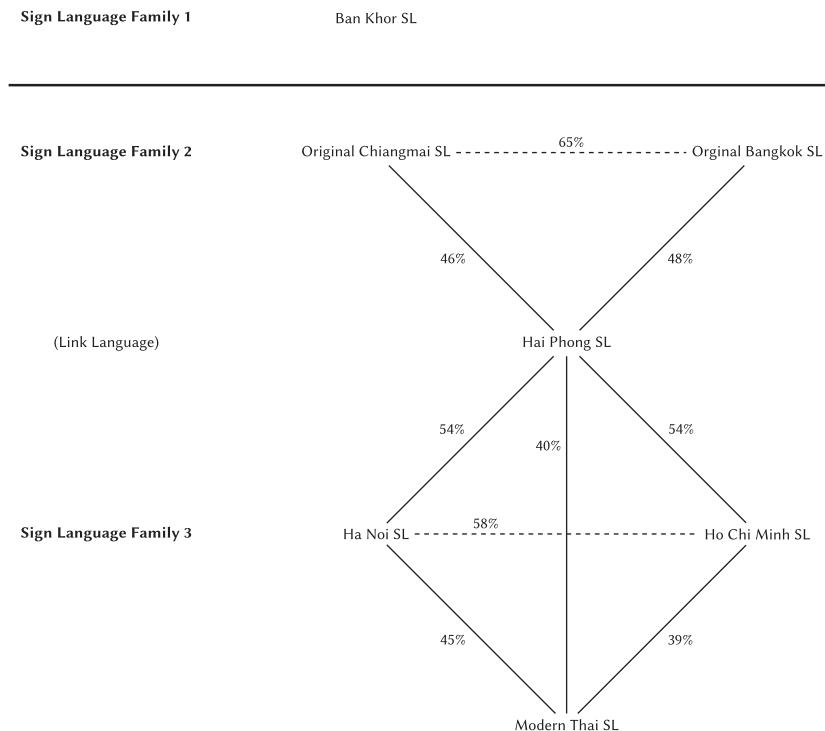


Figure 2.18 '*Linguistic relations between sign languages in Thailand and Vietnam*' After Woodward (2000: 32). The percentages along connecting lines refer to cognacy measures in basic vocabulary.

These findings imply that the modern sign languages are ‘mixtures, probably creolizations, of original sign languages with French Sign Language, ASL, or both’ (Woodward 2000: 33). This suggests that Hai Phong has an original Southeast Asia substrate and a French Sign Language/ASL superstrate.

Cambodia

The first school for deaf people in Cambodia was established in 1997. Cambodian Sign Language shows a ‘significant relationship’ with Modern Thai Sign Language, as measured by lexical similarity ratings on a comparative set of basic vocabulary (Woodward *et al.* 2015a). That said, Cambodian Sign Language and Modern Thai Sign Language are nevertheless ‘very distinct languages’ (Woodward *et al.* 2015a). Woodward *et al.* (2015a) assumes that an indigenous substrate accounts for those elements of Cambodian Sign Language that are unlike the strongly ASL influenced Modern Thai Sign Language:

Cambodian Sign Language shows no significant relationship with any of the sign languages in Viet Nam (Ha Noi Sign Language, Hai Phong Sign Language, Ho Chi Minh City Sign Language) and shows no significant relationship with any of the original sign languages in Thailand (Original Bangkok Sign Language and Original Chiangmai Sign Language). Therefore CBDSL must have originated independently of these sign languages.

(Woodward *et al.* 2015a: 161)

Myanmar

Watkins (2010) presents some basic facts and observations on sign language of Yangon (Rangoon), in connection with the long established Mary Chapman School for the Deaf and the Yangon Deaf Association. Influences that have shaped the sign language observed in Yangon include ‘spoken Burmese and ASL, Korean sign language and possibly also an old BSL [British Sign Language] substratum’. ‘An indigenous system which preceded this one was in use in the school in the 1970s and 1980s and is still known and used by Deaf signers in Yangon who were in the school at that time’ (Watkins 2010). Kamei (2004, cited by Watkins 2010) suggests that there may be distinct northern (Mandalay) and southern (Yangon) sign language varieties.

2.9.5 Prospects

MSEA sign languages are badly understudied. And they are in a state of rapid change, like so many other minority languages of MSEA. They are subject to similar pressure as endangered languages around the world (Nonaka 2004):

Modern sign languages have already replaced original sign languages among younger signers in Thailand and in Viet Nam. Within 50 years, it is highly likely that all original sign languages in Thailand and Viet Nam will be extinct, dying out with the users who still remember them.

(Woodward 2000: 33)

Indigenous and original sign language varieties represent ‘crucial pieces in the puzzle of constructing “sign language families” (Woodward 1993, Zeshan 2003)’ (Nonaka 2004: 738). They are also central to questions of sign language typology and of linguistic typology more broadly.

2.10 Classical Languages

When Hindu and Buddhist religious influences spread into the MSEA area early in the modern era, they brought with them the written Indic languages Sanskrit and Pali. These languages were used in religious practice and epigraphy. Pali was the main language of the Buddhist teachings and traditions that came to prevail in the area, and which are now part of everyday life in Myanmar, Thailand, Cambodia, and Laos.

The ‘oldest surviving examples of Pali’ are found in MSEA (Stargardt 2000: 25). The Golden Pali Text and the Great Silver Reliquary are treasures of the mid fifth or sixth century CE, buried in Myanmar in a relic chamber under a stupa in the Pyū city of Śrī Kṣetra, the last and southernmost of the Pyū city states. The Golden Pali Text is ‘a small replica in solid gold of a palm leaf manuscript’ (Stargardt 2000: 22), containing twenty leaves of gold inscribed with excerpts from the Pali canon: a total of sixty lines of ‘pure canonical Pali’ (Stargardt 2000: 24). The Great Silver Reliquary – a hollow cylindrical container made of a thin, brittle, silver sheet – has a single line Pali inscription running around the lid.

Pali as a classical language is today read by specialists in Buddhist literature and the religious canon. Villagers across Thailand, Laos, Burma, and Cambodia encounter Pali on an almost daily basis, but few can be said to know or understand the language. In everyday temple rituals, chants are done in Pali, both by monks and by lay people. Lay people can recite many Pali words and phrases from these formulaic texts, but they are seldom able to translate the meanings.

Sanskrit is associated with Hinduism, and it has less of a presence in daily life in MSEA. That said, Sanskrit rather than Pali has been the source of official neologisms in modern languages of the Theravāda Buddhist nations of Cambodia, Laos, Myanmar, and Thailand (on Thai, see Gedney 1947, Prasithrathsint 1994, Diller 2007). The Thai Royal Institute, established in 1933 on the model of the French Academy, was ‘responsible for word coining and for periodic editions of the official dictionary’ of the Thai language (Diller 2007: 19). The Royal Institute created ‘new Indic (Pali, Sanskrit, or

mixed Pali/Sanskrit) calques based either on direct translation of Latin and Greek roots in Western vocabulary or more broadly on the semantics of the concept to be represented' (Diller 2007: 22). Table 2.25 shows some examples.

Table 2.25 '*Neologisms and lexical specifications*' based on Pali/Sanskrit in Thai

Targeted English	Proposed substitute(s)
art	sin ⁴ lapa ¹ , sin ⁴ (<Sanskrit <i>silpa</i> 'art')
city	buri: (<Sanskrit <i>puri</i> 'city')
effect	phon ⁴ (<Sanskrit <i>phala</i> 'fruit')
service	bo'ri ³ ka:n (Sanskrit <i>pari-kāra</i> 'around + action')
consume	bo'ri ³ pho:k ² (Pali <i>pari-bhōga</i> 'partake of food'; cf. Sanskrit 'engage in sex')
context	bo'ri ³ bot ¹ (Sanskrit <i>pari-pada</i> 'trap, snare')
culture	wat ³ thana ³ tham (Sanskrit <i>vadhna-dharma</i> as though 'change + doctrine')
reform	pat ¹ tiru:p ² (Sanskrit as though <i>pari-rūpa</i> 're + figure')

After Diller 2007: 22.

Another language that may be said to have classical language status in MSEA is Chinese. Much of the historical record of MSEA history (reviewed above) has come from records of Chinese dynasties over the last two thousand years. Vietnamese itself was long written in Chinese characters (and still is today in some contexts; see DeFrancis 1977, and next section).

2.11 Scripts and Writing Systems

Most languages of MSEA have no written form. In those MSEA languages that do have a written form, the uses of writing are mostly restricted to a small set of social and functional contexts and are used by only a minority in the speech community. Only a tiny minority of MSEA languages are exceptions to this. These are the dominant languages of MSEA nation states – Burmese, Thai, Khmer, Lao, and Vietnamese. These written forms of language observed in national media and formal education systems are used in a maximally broad range of contexts, for many practical purposes. These writing systems are used in bookkeeping, street signage, formal and personal correspondence (letters, emails, texts, or social media posts), instruction manuals, research reports, advertising, product labelling, and literature, among much else. Very few writing systems for non-national languages have come to be used for any of these functions. An example is the Hmong Romanized Popular Alphabet (see below).

Writing systems have two main types of function, which I will here term informational and indexical. The informational function of writing is the function of encoding

and conveying information that would otherwise be semantically encoded in the words and constructions of spoken language. This might refer to the price of some produce on a market stand, the amount of tax collected on a land title, or an article of personal news such as a birth or a death. The indexical function of writing is independent of semantic content. By using a form of writing quite apart from the informational content of what is being written a person can index that they are of a certain social or ethnic identity, have a certain degree of education or modernization, or make a certain claim to legitimacy (Goody 1977, Ong 1982). The informational and indexical functions of writing are theoretically independent from each other, but in practice they are usually intertwined.

The best known and most widely used forms of writing in core MSEA are those of the national language systems of Burmese, Khmer, Thai, Lao, and Vietnamese (as well as Chinese, along the north of the area). All of these but Vietnamese and Chinese are Indic in origin.

Indic based writing systems are alphabetic. They feature distinct characters for consonants and vowels. Syllables and words are written on a line, from left to right, normally without spaces between words. Many characters have alternative forms, depending on their immediate context in a syllable or word. For example, vowels in Indic based systems often have more than one distinct form. An independent form may be used, for example, when the vowel is a syllable coda. A bound or fused form may be used, for example, when the vowel is syllable medial. Often, two or more symbols can combine to signify a distinct sound. Because of the large number of vowel phonemes in the languages of MSEA (see Chapter 4), the Indic alphabetic system requires some adaptation in order to capture those sounds. For example, in Thai, the /ə/ vowel phoneme is signified by a combination of the Indic sourced forms for /e/ and /i/; the /u/ vowel phoneme in Thai is signified by the /i/ symbol with an added stroke.

The Indic scripts are all ultimately derived from the Pali/Sanskrit type of orthography, which in turn comes from the older Brāhmī script (Hartmann 1986: 7; see Masica 1991: 133–51 for an overview of the history and development of Brāhmī based Indic scripts). An example of an Indic based MSEA script can be illustrated by the array of Cham consonants shown in Table 2.26.

The letters shown in Table 2.26 are arranged, in part, following a linguistically grounded organization of the consonants in terms of place and manner of articulation. In Table 2.26, each of the first four rows of consonants are grouped according to place of articulation, starting at the back of the vocal tract and moving forward. The first four rows are at velar, palatal, apico dental, and bilabial place of articulation, respectively. Below these is a row of liquids/fricatives, plus a couple of other letters, and then the independent vowel letters. (Vowels also have dependent/abbreviated forms, which attach directly to consonant letters.) Within each of the top four rows, the left to right

Table 2.26 ‘Cham alphabet letters: Consonants’

Vowels									
៥៥	៥៥៥៥	៥៥	៥៥៥៥	៥៥	៥៥				
a /a/	i /i/	u /u/	é /ɛ/	ai /aɪ/	o /o/				
Consonants									
៥៥	៥៥៥៥	៥៥	៥៥៥៥	៥៥	៥៥				
k /k/	kh /kʰ/	g /g̚/	gh /gʰ/	ng /ŋ/	ng /ŋ/				
៥៥	៥៥៥៥	៥៥	៥៥៥៥	៥៥៥៥	៥៥៥៥				
c /c/	ch /tʃ/	j /dʒ/	jh /tɕ/	ny /ɲ/	ny /ɲ/				
៥៥	៥៥	៥៥	៥៥	៥៥	៥៥				
t /t/	th /tʰ/	d /d/	dh /dʰ/	n /n/	n /n/				
៥៥ የ៥៥	የ៥៥	ᅕᅕ	ᅕᅕ	ᅕᅕ	ᅕᅕ				
p /p/	ph /f/	b /b/	bh /bʰ/	m /m/	mb /b/				
ᅕᅕ	ᅕᅕ	ᅕᅕ	ᅕᅕ	ᅕᅕ	ᅕᅕ				
y /j/	r /r/	l /l/	w /w/	s /s/	h /h/				
Final consonants									
៥៥-k	የ៥៥-ŋg	የ៥ᖷ-c	የᅕᖷ-t	የᅕᖷ-n	የᅕᖷ-p	የᅕᖷ-y	የᅕᖷ-r	የᅕᖷ-l	
/-k//-/?	/-ŋ/	/-ar?/	/-t/	/-n/	/-p/	/-ar/	/-r/	/-l/	
የᅕᖷ-w	የᅕᖷ-s	የᅕᖷ-h	የᅕᖷ-m	የᅕᖷ-ng					
/-ao//ɔ-/	/-əs/	/-h/	/-m/	/-ŋ/					
Numerals									
ᅕᖷ 0	የᅕᖷ 1	የᅕᖷ 2	የᅕᖷ 3	የᅕᖷ 4	የᅕᖷ 5	የᅕᖷ 6	የᅕᖷ 7	የᅕᖷ 8	የᅕᖷ 9
thaoh	sa	dua	klau	pak	lamā	nem	tajuh	dalapan	salapan
/t̚əh/	/sa:/	/doa:/	/klau/	/pa:?	/la-mɔ:/	/nʌm/	/ta-dzuh/	/də-la-pʌn/	/sa-la-pʌn/

From <https://nguoicham.com/cdict>

ordering of letters also follows a consistent logical order, this time in terms of different manners of articulation. (Note the velar gap – no voiced ‘g’ – at the end of the first row.) This resembles the basic logic of all of the Indic based systems, including the Devanagari scripts used for writing Sanskrit and Hindi. The Cham alphabet features more symbols than just these consonants, and there are complexities in spelling, including alternate forms of certain letters, digraphs, and other kinds of combination.

In the Hinduized early states of MSEA, the language used by scribes and rulers was classical Sanskrit. Sanskrit inscriptions on stone or copper plates have been found in

Cambodia dating from the fifth century CE. Polities associated with Buddhism used Pali. Sinhalese monks carrying out Buddhist missionary work, most intensively in the eleventh and twelfth centuries, brought Theravāda scriptures to MSEA. These were written mostly in Pali. Pyū inscriptions in Myanmar date from the sixth century CE, Mon inscriptions in Thailand from the eighth century (Stargardt 2000, Hartmann 1987: 7). The orthographies used for writing these classical languages were soon adopted for writing local vernacular languages, for example in the cases of Burmese and Thai:

The Burmese, recently emerged as a distinct ethnic group at that time, developed a script for their own language in the twelfth century. The Thais, infiltrating the area between the Burmese and the Khmer, devised an alphabet of their own in the late thirteenth century.

(Hartmann 1987: 7)

The main physical medium for the spread of these orthographies was palm leaf, from palmyra palms (*Borassus* spp.) and talipot palms (*Corypha umbraculifera*). Scriptures and other forms of literature were transcribed by monks in village temples on palm leaves and bundled together into book form. The manuscripts naturally biodegrade within a few decades. Scribes would therefore be required to preserve the texts by regularly copying them onto fresh manuscripts. In this way, they maintained the faithful transmission of the texts, along with the classical languages and their orthographies. They also would introduce minor alterations, by accident or design.

Palm leaf inscription was done using a metal stylus. Because of the physical nature of the palm leaves, sharp angles in the lettering could split the leaves along their grain.

၆ မှ ၁၂။ ထိပ်ကြီးအား ဦးဆောက်ပုံးအမည်ယူလေသာ တယောက်သာ မြတ်နိုင်ယူ ဖို့ရား
သော်ယို၏။ ထိပ်ဆောက်ပုံးအမည်ယူလေသာ မိမိရားသည် ဘုရားမြတ်မြတ်၏။ အကျိုးစီးပါးကို ရှားပို့ဆေ
ယို၏။ အလုပ်စံသာ ကိစ္စတို့ လိမ်းသာ။ ထိပို့ရားတွင် အမည်အားမြင် ရာကျားမာရီ၏သောတယောက်
သောသားသည်ယို၏။ ထိပ်သားသည် မင်းမြှုပ်ရေးတို့၏ မင်းအနေတွေ ကြော်ကြေားထုတ်ပြုလုပ်
တတ်သော အမတ်လုပ်မြင်၏။ သတ်ကရာဏ်ယို၏။ အားမြင် ပုံးပြည်သာမြို့၏။ ထိုတိုင် မင်းကြေား
ကြိုးသည်။ ပို့ရားအား ကြည့်လို့သည်မြင်၏။ အဲခေါ်သိမ်း ကျွန်အသံအောင်ပြု၏ သုတေသန၏
ရွှေ့ကိုပေးသာနားတော်မူလျှော့။

Figure 2.19 *Modern Burmese writing*

ມະນາດຖຸກຄົນເກີດມາມີກົງດັບຕັກສິໄລທີ່/ແລືພູບ
ແລະອາວຸນສມີຍັງບໍ່ເຖິງຫຼາຍ້ນັ້ນ. ບຸກງົາຄົນມີ
ເຫດດີ້ນີ້ແລະອາວຸນຄົດລາວມເຕັກສົນເຖິງຂອງໃຈ
ຂອງໜັນ/ແຕ່ວ່າມະນຸງທຸກໆຄົດລາວມປະເປີດຕັກສິກັນ
ກັບເປັນອ້ອງຫ້າຍ້ນັ້ນ.

Figure 2.20 Modern Lao writing

Curved lettering developed as a way to avoid this. Modern scripts like Burmese and Lao have retained the curved style. See Figures 2.19 and 2.20.

In the first millennium states in the areas of present day Myanmar and Thailand, distinct scripts for both Mon and Khmer had emerged by around the beginning of the seventh century (Finot 1959, Hall 1966). When Tai speaking peoples moved into MSEA, ‘they probably first encountered and adopted Mon type script’ (Hartmann 1987: 8). New orthographies based on these Indic scripts were subsequently developed and spread across the Tai speaking areas of MSEA. These include the *Tham* script and related religious forms of writing in northern Thailand and Laos, scripts used for writing Lue and Shan in southern China and upland Eastern Myanmar, and scripts used by the White, Black, and Red Tai, in northern Laos and Vietnam. The history of development of these latter systems is not well understood. What is known is that these minority Tai scripts, unlike the better known scripts of national languages like Burmese, Thai, Lao, and Khmer, developed independent of any link to Pali/Sanskrit linguistics and literature. ‘The Black Tai (Tai Dam) alphabet might be considered a folk adaptation of an Indic script the White, Black, and Red Thai borrowed from their Lao or Lanna Thai neighbours’ (Hartmann 1987: 10). Figure 2.21 shows an example of the modern Black Tai writing.

To the northwest of core MSEA, in northeast India and western Myanmar, Tai speaking groups including the Shan, Khamti, and Ahom also have Indic derived scripts but ‘appear to form a separate group [from other Tai languages] both in terms of alphabets and other historical developments’ (Hartmann 1987: 12).

Indic based Tai language writing systems can be grouped into two main types, on the basis of their specific manners of indicating sounds (F. K. Li 1977: 22–3, drawing on Cœdès 1925). One is the Siamese type. This type features a full complement of vowel sounds, dealing with the fact that the Indic languages, for which the source scripts were originally suited, have fewer vowel sounds than the average MSEA language. The Siamese type also preserves historical distinctions in voicing of early Tai initials, where

ມາ ອົງທຶນ ໜັນດີນ ມະນີ ມີເວິ ມີແວິ ວິ ມີກັນ ວິ ອົງທຶນ
ດູນທຽບ ພິໄສ ແລະ ທຽບ ຮິນ ມີຊື່ ຂີ່ເປົ້າພື້ນແອຸນທຽບ ຢິນເມີນ ນົງ ແກ່ນ ມະນີ
ພື້ນທຽບ ຊັກ ວິ ວິ ຮິນ ແລະ ທຽບທຽບ ນີ້ ດັກ ດັກ ດັກ ດັກ ຕ່າງໆ ຕ່າງໆ ຕ່າງໆ
ກຳນົດ ມະນີ ສົກ
ນີ້ ດັກ
ນີ້ ດັກ
ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ
ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ
ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ ດັກ

Figure 2.21 *Modern Lao Song (Black Tai) writing*
From Mukdawijitra 2019.

that voicing is now lost, but is embodied in a lexical distinction of tone (see sections 2.3.2 and 4.6). Thus, in Lao, the monosyllabic words for ‘leg’ and ‘stuck’ have the same segmental structure [k^ha:] but are pronounced with different pitch contours. The tone difference is signalled orthographically by the choice of initial letter. The first letter in the word for ‘leg’ is in the ‘low series’ of consonants, while the initial letter in ‘stuck’ is in the ‘high series’.

- (2.1) ຂ້າ (kh_{HIGH SERIES}+a:) /kha:¹³/ ‘leg’
 ຂ້າ (kh_{LOW SERIES}+a:) /kha:³⁵/ ‘stuck’

Diacritics can also indicate tone in Lao. The monosyllabic words for ‘galangal (*Aplinia galanga*)’ and ‘to kill’ have the same segmental structure [k^ha:] but have different pitch contours. The tone difference is signalled orthographically by tone marking diacritics:

- (2.2) ຂ້າጀ (kh_{HIGH SERIES}+tone 1 marker+a:) /kha:³³/ ‘galangal’
 ຂ້າጀ (kh_{HIGH SERIES}+tone 2 marker+a:) /kha:³¹/ ‘to kill’

The other main type of Indic based writing system for Tai languages is the Shan type. This type ‘does not provide enough symbols for the more complicated Tai vowels and diphthongs, and does not differentiate the old voiced consonants from the voiceless ones . . . It usually has no tone marks’ (F. K. Li 1977: 23).

We can note, finally, that some Tai speaking peoples have used non Indic writing systems. Tai speakers of the border region between China and northern Vietnam in the thirteenth century developed writing systems based on Chinese. This was ‘inspired and implemented, no doubt, by the same forces that led to the appearance of Chữ Nôm script [in Vietnam]’ (Hartmann 1987: 11; see below).

The development and spread of Indic based scripts in MSEA is summarized in Figure 2.22.

Vietnamese is written today in a Latin based script known as Chữ Quốc Ngữ (see DeFrancis 1977 for a historical overview.) For most of the period of formal Chinese domination over Vietnam (111 BCE – 938 CE), Vietnamese remained an oral language. From the second century BCE, the only written language in Vietnam was Chinese, in Chinese characters (referred to as Chữ Hán in Vietnamese). By the eighth century CE, a Chinese character based system had emerged, called Chữ Nôm, for writing the Vietnamese language. Both Chữ Nôm and Chữ Hán are still used in Vietnam today, but in restricted contexts. Chinese characters in Vietnam are more for ritual, symbolic, and decorative purposes than for general communication of information.

Vietnamese is most widely written using the official Latin based modern standard form of romanization. The Chữ Quốc Ngữ system has its roots in the sixteenth century work of Portuguese and Italian Jesuit missionaries in Vietnam. The Portuguese

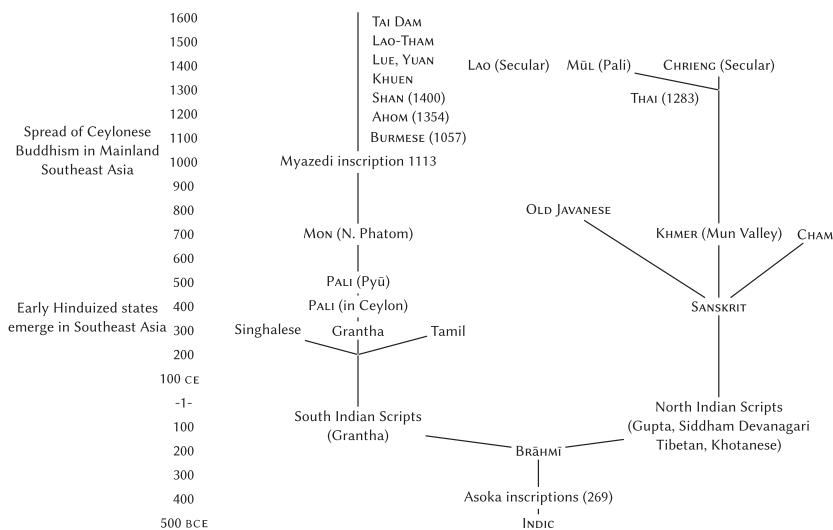


Figure 2.22 '*Historical development of Indic scripts in Southeast Asia*'

Adapted from Hartmann 1986: 13.

missionary Francisco de Pina initiated the development of the Latin based script, partly for the purpose of learning the Vietnamese language. This work was continued in the seventeenth century by Alexandre de Rhodes, who wrote the first Vietnamese Portuguese Latin dictionary (DeFrancis 1977: 52ff. Jacques 2002, Haudricourt 1949/2010). For two centuries, both this Latin based system and the Chinese based Chữ Nôm system were used for writing Vietnamese. In the nineteenth century, French colonial authorities took strong measures to promote the Latin based system and to discourage the use of Chinese characters. By the early twentieth century, Chinese characters had been removed from formal education and administration in Vietnam, entirely replaced by the Latin based Chữ Quốc Ngữ.

Chữ Quốc Ngữ writing consists of the Latin alphabet with diacritics and special marks to capture Vietnamese specific sounds. One function of these marks is to distinguish between certain consonants, for example ‘d’ /z/ (or /j/ in certain dialects) versus ‘đ’ /d/. Another function is to distinguish between certain vowels, for example ‘a’ /a/ versus ‘â’ /ʌ/, ‘u’ /u/ versus ‘ü’ /w/, ‘o’ /ɔ/ versus ‘ö’ /ø/ versus ‘ô’ /o/. Five diacritics – four combining above, one below – mark lexical tones (where one tone is unmarked; for more on Vietnamese tones, see section 4.5):

- (2.3) *ma* ‘ghost’ [ma:³²]
 má ‘mother’ [ma:³⁵]

- mā* ‘horse’ [ma[?]a³⁵]
mǎ ‘tomb’ [ma.²¹³]
mà ‘which’ [ma.³¹]
ma ‘rice seedling’ [ma?²¹]

The spelling of Vietnamese words shows a regular and consistent mapping of written form to sound.

Turning to minority languages of MSEA, most are unwritten. A trope among many minority communities is that they once had a writing system but it was lost. Across the area, ‘the apparent absence of indigenous literacy is lamented through tales in which writing was lost through profligacy, carelessness, treachery or acts of desperation’ (Kelly 2018: 1):

The Akha, it is said, kept their writing on buffalo skins but were forced to consume them while fleeing the invading Tai armies. The Wa wrote on ox hide that was eaten in hard times, while the Lahu wrote their letters on cakes that met the same fate. The writing of the Hmong was destroyed by the Chinese or eaten by horses, while Karen writing was variously stolen, eaten or left to rot. Similar stories of being cheated out of literacy and its presumed benefits are told by the Kachin, the Chin and the Khmu (for summaries see Enwall 1994; Rastorfer 1994; Scott 2009: 221–2).

(Kelly 2018: 1)

In his exploration of uplanders’ avoidance of state control, Scott (2009) argues that writing systems go against widespread uplander values of freedom, transience, and sociocultural flexibility. Because writing has its ultimate origins in accounting, it is associated with forms of political control such as taxation (Goody 1977, Ong 1982). And because it fixes information and allows it to be transported away from its original context of expression, it is antithetical to impermanence. Because ‘the permanency of a written text is a potential drag on adaptive cultural dynamism’, then perhaps ‘it was preferable for literacy to be destroyed than for its practitioners to suffer their own destruction’ (Kelly 2018: 3).

Alongside those groups who may have abandoned their writing systems, and the many who have never written their languages, a number of minority groups of MSEA have independently developed unique writing systems, bringing about a ‘return of writing’. Kelly (2018) reviews a number of examples of this – among the Karen, Jruq, Chin, and Hmong – and discerns features that are common to all of the cases, despite the communities coming from quite distinct linguistic and cultural backgrounds. These are listed in Table 2.27.

Table 2.27 *Features common to cases of the independent development of orthography in ‘Zomian’ minority languages*

-
- storytellers narrate a critical event in which literacy, and its presumed advantages, was lost
 - a messiah or king is prophesied
 - an external upheaval manifests itself as an internal crisis of leadership and allegiance
 - a charismatic individual rises from the chaos attracting followers by calling for a restoration of a past order
 - revelational events bring writing back to the communities in the form of a distinctive script
 - the script is deployed by an emergent politico-religious institution
-

After Kelly 2018: 14.

An illustration is the case of a part invented, part inspired script consisting of some 300 characters called Khom (Sidwell 2008). This script was used for writing Loven, a Bahnaric language of southern Laos. Khom script was invented by a Loven man named Kommadam who led an anti French rebellion on the Boloven Plateau between 1901 and 1936. In interviews in the late twentieth century, older Loven people ‘proudly recall that the Kommadam gave them a script for their language, like a “real” nation’ (Sidwell 2008: 19). Legend had it that the script was revealed by the revolutionary leader Kommadam in the following way:

Kommandame [sic] would go into a trance (a possible demonic possession), he would repeat the sound over and over again, as the symbol appeared on the naked flesh of his chest. Once the scribe had recorded it, there would be another. After these sessions, Kommandame would be utterly exhausted, and sometimes sleep for the next day.

(Sidwell 2008: 18–19, quoting a letter by John Davis)

Sidwell’s account of the history and symbolic role of this writing system gets to the heart of the very different logic that has driven the development of scripts among the people of Zomia (see Scott 2009, Kelly 2018). While the major scripts of MSEA such as those used for writing Vietnamese, Burmese, Cambodian, Lao, and Thai are both practical and symbolic, scripts like Kommadam’s Khom script for Loven functioned primarily to symbolize power and identity rather than to communicate information:

Khom script was perhaps never intended as a real language planning tool, but merely a very elaborate symbol of mystical power, part of the grand religious justification of Kommadam’s place at the head of the anti colonial rebellion.

(Sidwell 2008: 24)

The Hmong language has seen similar developments. Hickmann (2019) details some of the stories that describe loss of orthographies:

One version of the story has it that when Hmong lost the great war and were fleeing for their lives, their books fell into the great river they had to cross when they sought refuge. The books were washed away, and Hmong writing was lost with the current. Another version of the story tells how, after the Chinese won the war, they stole the writing system from Hmong people, who were then destined to become an illiterate people with a strong oral tradition, having lost their great society as the mandate of heaven passed to the Chinese ‘younger brother.’ Yet another version of this origin myth describes how subjugated Hmong were forced to hide their characters in the intricate embroidery that then became so central to Hmong material culture, as an ‘ethnic minority.’

(Hickmann 2019: 1)

The Is Npis Mis Nus (INMN) group in Thailand takes its name from ‘the title of the founding prophet, who had received divine revelations and established a religious commune in a Hmong village complex in Sayabouri province in Laos in the 1960s and 1970s’ (Hickmann 2019: 13). This group has an associated orthography. Another example is the Yaj Soob Lwm, ‘The Mother of Writing’ (Smalley *et al.* 1990). Yaj Soob Lwm ‘also received heavenly visions in the 1950s, and had established what would come to be the most commonly used non Romanized orthography for Hmong’. This is the Hmong Romanized Popular Alphabet (RPA), now ‘the most common international standard for writing White (Hmoob Dawb) and Green Hmong (Moob Ntsuab)’ (Hickmann 2019: 2). The Hmong RPA is unique in having letters at the ends of syllables for tones (see Figure 4.6 and example (4.10) in section 4.5). It also stands apart from all of the other ‘returned’ scripts of MSEA in so far as it is widely used for purely practical purposes by large populations.

For completeness, we may mention the use of Latin based orthographies in the work of descriptive linguistics. For example, while languages like Kri and Semelai do not have written forms that are used by their speakers, writing systems have been developed by linguists to transcribe the languages phonemically. Many of the examples provided in chapters 4–7, below, are rendered in these technical orthographies.

In concluding this section, I emphasize that scripts and writing systems are secondary representations of language. They are politicised and highly salient markers of ethnic identity. Often, their indexical functions are more important than their informational functions. Hartmann’s remarks on Tai are illustrative:

The appearance of writing systems among Tai and Vietnamese in the late twelfth and early thirteenth centuries seems to be part of a cultural wave of awakening or assertion of ethnicity among all of these various groups following the early precedent and eventual decline of the Mon and Khmer as political and cultural forces in mainland Southeast Asia. An alphabetic renaissance seems to have been developing at the time. Each group appears to have wanted to assert its political and cultural independence and uniqueness from its neighbours or patron power – the Mon, Khmer, or Chinese.

(Hartmann 1987: 11)

This political and ideological power of orthography helps to explain numerous grass roots orthography projects that some Zomian groups have embarked on over recent centuries.

It is also important to emphasize that despite its indexical and symbolic power, a script is not connected to any kind of essence in a language. The cases discussed in this section all illustrate a constantly recurring theme in the language and culture of MSEA over millennia. Aspects of culture – from livelihoods, to language, to food and dress, and orthography – are adopted from outside a group, later becoming entirely appropriated and internalized as a sign of native cultural identity. Consider Blood's remarks on the Cham writing system in Vietnam: 'When Hindu grammarians came to the Southeast Asian peninsula in the second century with their own writing system, they brought with them a legacy to the Cham people of south Vietnam which is for them a vital symbol of cultural unity today' (Blood 1980: 35). Over time, something that was entirely external to the linguistic and cultural context has become an integral part of it. The Cham script now 'acts as a cohesive device binding together the people of Champa', writes Blood (1980: 43): 'By 1975 it had become a symbol of their uniqueness in a hostile world that eats away at their identity.'

Typological Overview

Linguistic typology is the study of universals and diversity in human language. In seeking to determine the range of variation, typologists chart the possibilities of human language. Various types of linguistic structure have been proposed, and typologists are becoming more confident in their statements about what is expected or usual and what is exceptional in human language (Song 2011, Aikhenvald and Dixon 2017). As in any area of the world, MSEA languages are in some ways ordinary, having properties shared by many languages around the world. But in other ways they are unusual, displaying rare or marked features. What makes any area unique in typological terms is not any single feature but a characteristic constellation of features.

3.1 Preliminaries

This chapter, with subsequent chapters in the book, constitutes a typological survey of the languages of MSEA. We begin here with an overview of some basic structural features of MSEA languages in the context of traditional linguistic typology, including the relative order of phrasal elements. Chapter 4 surveys sounds and sound systems in MSEA, examining the diversity of consonant and vowel inventories, and with special attention to the intertwined phenomena of register and tone. Chapter 5 examines principles of word formation, showing the many ways in which MSEA languages depart from the stereotype of an extreme isolating/analytic profile. Chapter 6 surveys the main forms and structures that speakers of MSEA languages use for formulating reference to persons, places, and things. And Chapter 7 looks at the description of events, especially by means of verbs, their marking, and their role in clauses, with special attention to the MSEA wide phenomenon of multi verb constructions.

Every spoken language provides its speakers with an inventory and set of rules for forming syllables and words from combinations of consonants and vowels, and other elements of sound such as pitch and duration. In MSEA languages, most words consist of a single stressed syllable. Suprasegmental features play an important role in MSEA phonological systems, especially in relation to tone and phonation type

register (see Chapter 4). The typical definition of a tone system holds that a word must be pronounced with the same pitch contour each time. Changing the pitch can signal a different word (for example in Lao, if *kaj* is pronounced with a low rising pitch, it means ‘far’, but with a high falling pitch it means ‘near’). In tone languages of MSEA, there are usually between four and six defined pitch patterns used for distinguishing between words (see Chapter 4). *Register* is a similar phenomenon, where the most important distinguishing feature is not pitch but voice quality – for example breathy phonation versus clear phonation. In Kri (spoken in upland central Laos), if *cii?* is pronounced with breathy phonation it means ‘elder sister’, but with clear phonation it means ‘head louse’. We shall see in Chapter 4 that matters are more complex than this neat distinction between ‘tone’ and ‘register’ suggests.¹

All languages have means of word formation. We shall see that affixation is limited in MSEA, but that compounding and reduplication are relied on heavily. We shall also see that the distinction between morphology and syntax has less utility in MSEA languages than it does in many other languages of the world. In the domain of word classes, the major classes noun and verb do not present problems for identification. Adjectives are typically a subclass of verbs.

Let me introduce a few basic terms and concepts I will use in subsequent chapters.

Verbs are central in all languages in the construction of *clauses*. These are the structures that make up, or form the basis of, sentences. Clauses are at the core of sentence level syntax. They describe an action, state, or property being characterized at each point in a discourse. A verb, as the primary constituent of a clause, has *valency*. It specifies the number and type of participants in the event or situation described. A verb is *monovalent* (or *intransitive* or *one place*) when there is just one of those participants, as in, for example, *Minh fell*. When there are two participants as in *Minh saw me* the verb is *bivalent* (or *transitive* or *two place*). While I will sometimes use the terms *intransitive* and *transitive* with these meanings, I note that the term *transitivity* can be useful for referring to a concept that is distinct from valency, and complements it. Valency is a discrete distinction, referring to whether a verb specifies one, two, or more participants. Transitivity is a scalar notion, capturing the idea that, say, within the set of verbal expressions that are bivalent, all having two arguments, some are more transitive than others. Those that are more transitive will tend to imply a more controlling actor and a more affected undergoer, among other things. So, while *Minh demolished the house* and *Minh noticed the house* are both transitive (or bivalent) they differ in their degree of transitivity.

¹ We shall also see that register has two other meanings: in phonology, register can refer to the splitting of a tone into two new tones, a higher and lower ‘register’; and in sociolinguistics, register refers to a specified style or mode of speech, with a defined degree of formality or appropriateness to context.

Verbs present two main types of interpretative problem that the grammatical system of any language must solve. The first is the *argument structure* problem, linking nominal arguments to verbs in terms of their roles in events and situations, and the second is the *grounding* problem, signalling the relations – for example, in time, place, and person – between the speech event and the event or situation being described.

The argument structure problem is the problem of making it clear who did what to whom. When an event denoted by a verb has two participants, typically there is a strong semantic asymmetry between the participants, and hearers need to identify which participant fills which role. For instance, if the verb is *praise* and there are two people involved in the event of praising – one person praises, the other is praised – the grammar of a language should be able to make it clear who did the praising and who was praised. In order to talk about how this problem is solved, it is useful to have terminology that allows us to distinguish between the syntactic functions that a participant can play in a basic clause. In a canonical transitive clause, one of the two participants is more agentive. One of the participants instigates and controls the event, while the other is an undergoer that is affected by the event. We can label these participants A and O, respectively. In the clause *Minh drank water*, the A argument is *Minh*, the O argument is *water*. Intransitive clauses have just one participant, which we can label S.

Languages of the world offer three main types of strategy for distinguishing between A, S, and O. One is constituent order: in English, *Minh saw you* is crucially different from *You saw Minh*. The rule is that A is placed before the main verb, while O comes after it. MSEA languages are widely thought to rely most heavily on constituent order for signalling grammatical functions. We shall see that to an extent this is true, but there is considerable leeway for constituent order to vary for expressive purposes.

Another argument structure strategy is case marking, that is, direct marking on nominal arguments to show their grammatical function in relation to the verb. This is found in the English pronoun system, in which most pronouns have two forms, an accusative (e.g., *them*, when in O function) and a nominative (e.g., *they*, when in A or S function). So, the pronoun in *Minh saw them* is different from that in *They saw Minh*, even though the same group is referred to. In English intransitive clauses, S function is marked in the same way as A in a transitive clause: we say *They sneezed* and not **Them sneezed*. We shall see that there is no case marking in MSEA in any usual sense of that term, although occasionally the term *case*, or a related term, is used (see section 3.2).

The third main way in which languages allow people to distinguish who from whom is to mark the verb directly with information that can identify which participant fulfils which role. This type of system is often termed *agreement*. English has a limited system of agreement. In the simple present tense, English speakers add a suffix *s* when the S or A argument of the verb is third person singular. Thus, the verb takes agreement marking in *She sneeze s* (in which the S argument is third person singular) or *She see s them* (in

which the A argument is third person singular), but not in *They see ø her* (in which only the O argument is third person singular). We shall see that there is next to no agreement in the grammars of MSEA languages.

The English system treats S and A arguments similarly. For example, S and A both trigger verb agreement, while O does not. Both S and A occur preverbally, while O does not. When S and A are grouped together by being treated in the same way in a grammatical system, we refer to this group as *subject* (in one important sense of this term, which has other meanings). Many languages have grammatical systems that do not group S and A together, but rather group S and O together. This is referred to as an *ergative* arrangement. There is little if anything that could be called ergativity in MSEA (but see LaPolla 1992, LaPolla 1995, Morey 2013 for discussion of some relevant phenomena in Tibeto Burman languages of the northwest fringe of the greater MSEA area).

Secondary to the basic organization of argument structure by which who is distinguished from whom are mechanisms referred to as *voice*. Voice provides grammatical ways to reduce, increase, or rearrange the mentioned participants in an event, or the relations between them. For example, a passive derivation provides a way to take a two place predicate, such as *praise* in *Minh praised Anh*, and change it to a one place predicate, as in *Anh was praised*. The previous A is demoted, either not expressed or expressed in a ‘by’ phrase that is external to the core of the clause (*Anh was praised by Minh*). The previous O (transitive object) becomes now an S (intransitive subject). Another valency changing derivation found in many languages is the antipassive derivation, which English does not have. The antipassive is a logical alternative to the passive. It takes a two place predicate and changes it to a one place predicate, but where it is the previous O that is demoted (either omitted or expressed in a ‘by’ phrase), and the previous A that is now an S. If English had an antipassive derivation, it would take the sentence *Minh praised Anh* and yield *Minh praised ANTIPASS*, meaning ‘Minh praised’ or something like ‘Minh did some praising’, with the possible addition of a phrase like *of Anh*. No form of productive grammatical alternation akin to antipassive is found in MSEA languages.

Related to voice is the phenomenon of *ambitransitivity*. This is where a single verb can be used both transitively or intransitively. (Ambitransitive verbs are also sometimes referred to as *labile*.) An example is the English verb *break*. It can be used intransitively, as in *It broke*, or transitively, as in *I broke it*. Another example is *eat*. Both *I ate* and *I ate it* are fine in English. The pattern of ambitransitivity is different in these two verbs. *Break* is S O ambitransitive. In the intransitive usage the S argument is an affected entity, while in the transitive usage that entity is in O function. *Eat* is S A ambitransitive. In the intransitive usage the S argument is an actor, who instigates and controls an activity. This entity is in A function in the transitive usage of the verb.

Further terms and concepts are explicated where needed.

3.2 Typological Overview

The following chapters of this book describe some of the most prominent structural properties of MSEA languages. In this overview chapter, we begin with two summary lists of typological features, in tables 3.1 and 3.2.²

Some of the most noteworthy commonalities among MSEA languages have to do with their *lack* of marking of certain semantico grammatical categories. Most notably, as remarked upon in the list above, the languages lack inflectional morphology in the usual sense of that term (i.e., agreement, case, gender/number/definiteness on noun phrases, tense marking on verbs, etc.). It may be said that a lack of a feature should not be treated as a feature. We would not group English with Swahili on the basis that they each lack a category of dual pronouns, or verbal marking of evidentiality. But categories like dual pronouns and evidentiality, while common in the world's languages, are in a sense marked categories, while the things that MSEA languages

Table 3.1 *Overview of sound system features in MSEA languages*

-
1. Vowel systems are large, showing many distinctions (it is sometimes difficult to determine how many vowels a system has; there are alternative analyses of features such as diphthongs and phonation splits).
 2. Vowel phoneme systems often show a common underlying canonical structure: nine-place, symmetrical, hi-mid-low by front-central-back.
 3. Vowel systems often show a long versus short vowel distinction.
 4. Many more consonants are possible in initial position than in final position.³
 5. There is a preference for one (major) syllable per word, with many languages featuring minor syllables or pre-syllables in an iambic pattern (see Pittayaporn 2015, Butler 2015, Post 2015, and Brunelle and Kirby 2015).
 6. Lexical contrast is marked by laryngeal features including pitch and phonation type, often in combination; lexically contrastive pitch and phonation type are strongly correlated in functional and historical terms.
 7. Tone systems are complex (number of tones ranges from four to nine or possibly more, with counts for a language differing depending on the analysis chosen).
 8. Phonation type systems usually distinguish two registers, e.g., ‘clear’ versus ‘breathy’ or similar.
 9. There is often a gap in voiced stop series at velar place of articulation (no voiced ‘g’).⁴
 10. Implosive consonants are found more frequently than elsewhere in the world.
-

² The lists draw mostly from Enfield 2005: 186–90 and Enfield 2011b: 69–70, which in turn draw from many references cited there; see Clark 1985, 1989, Clark and Prasithrathsint 1985, Matisoff 1991a, 1991b, 2001, Bisang 1996, Comrie 2007, Jenny *et al.* 2015.

³ It is a universal of human language that there are fewer possible phonological distinctions in coda position than in initial position, but in MSEA languages the ratio is especially high (Rousset 2004, Gordon 2016: 98–104; see section 4.4).

⁴ Universally, voiced velar stops are slightly less likely to occur in phoneme inventories than either voiced bilabial stops or voiced dental/alveolar stops (Gordon 2016: 46).

Table 3.2 *Overview of morphosyntax-semantics system features*

-
1. There is no inflectional morphology (no case, gender, number, or definiteness marked on noun phrases, no agreement or tense-marking on verbs); but note that *derivational* morphology is widespread and sometimes highly productive in Austroasiatic languages of MSEA (see Alves 2015, Jenny *et al.* 2015).
 2. Open class items mostly nouns and verbs often serve functions that are expressed by dedicated functional morphemes (including bound morphology) in other languages, e.g., nominals as prepositions, verbs as aspect markers, comparative markers, adversative passive markers, and valence-changing devices (Clark and Prasithrathsint 1985, Kölver 1991, Ansaldi 1999).
 3. Verb serialization (meaning a range of different kinds of predicative structures that use combinations of verbs) is used extensively, with a rich array of types and functions in each language (Bisang 1991).
 4. Order of major constituents of the clause tends to be relatively flexible within languages, sensitive to pragmatic factors (though verb-object constituent order is dominant); noun phrases tend to be head-initial, and may have discontinuous constituents, especially when classifiers are involved ('Constituent order at the clause level is fluid'; Kruspe 2004: 5).
 5. Zero anaphora: noun phrases may often be ellipsed when their referents are contextually retrievable (this, combined with flexibility in constituent order, results in variable surface options; for a case study see Enfield 2007: 271–84).
 6. Topic-comment structure is used extensively at the clause level.
 7. Languages often have a large set of labile or ambitransitive verbs, especially of the S O type (sometimes called causative/inchoative or unaccusative; e.g., Lao *hak2* can mean transitive ‘snap’ or intransitive ‘is/has been snapped’).
 8. Rich inventories of sentence-final particles make subtle distinctions in sentence type, stance, evidentiality, and combinations thereof.
 9. There are rich expressive forms, including large inventories of ideophones (or ‘expressives’), rhyming four-syllable expressions, and productive elaborative rhyming devices.
 10. There are numeral classifiers and related systems of nominal classification.
 11. A number of languages feature complex pronominal systems, with multi-level social-deictic meaning distinctions.
-

happen to lack are much more general and unmarked types of language structure, in particular the absence of any form of inflectional morphology. We can proceed with caution.

3.3 Morphological Profile

Morphological typology is concerned with how languages put words together (Payne 2017). Every language has an *index of synthesis*. This is the average number of morphemes per word in that language (Sapir 1921: 128, Payne 2017). At the far end of the scale of morphological synthesis are polysynthetic languages such as those of the Inuit Aleut family, which have a morpheme to word ratio of ‘well over 3’ (Payne 2017:

87). English has a synthesis index of ‘approximately 1.7’ morphemes per word (Payne 2017: 86). At the low end of the scale is the logical extreme: a morpheme to word ratio of exactly 1, in other words, where all morphemes are stand alone words. Languages that are positioned towards this end of the scale are said to be *isolating* in morphological type. MSEA languages are put forward as classic examples of languages with this low index of synthesis.

We shall see in later chapters that MSEA languages do not, in fact, quite show the logical extreme ratio of one morpheme per word. One reason for this is that the MSEA languages that are typically cited in this regard include Vietnamese, Thai, and Chinese, and these languages are in a minority in MSEA in this respect. Austroasiatic languages, which make up nearly half of the languages in core MSEA, show a higher index of synthesis, with their widespread use of derivational morphology (see Chapter 4). And even in oft cited extreme isolating languages such as Vietnamese and Thai, certain rich and productive morphological processes are often overlooked in typological work. Further, a ‘word’ may be defined either in phonological or grammatical terms, implying different results of a morpheme to word ratio count (see section 4.9).

When typologists think of complex morphology, they tend to think of inflectional markings for categories such as case (e.g., accusative versus nominative), person (e.g., first, second, or third person), or number (e.g., plural versus singular). MSEA languages lack morphological marking for these categories, but they make up for it with marking of meanings in other domains, such as the compounding and elaborate forms of reduplication described in Chapter 4.

In his famous characterization of morphological type, Sapir (1921) used Vietnamese (which he termed Annamese) as the paradigm case of a language in which words are not morphologically complex. Many people still think this (but see Chapter 4 below for a case study of a complex derivational morphological system in Vietnamese). It is probably true that these languages are as far as you get on that scale, but they do not approach the logical extreme – that is, the situation in which there are no words in the language with more than one morpheme. No language shows a purely isolating profile.

An underappreciated domain of morphological complexity in MSEA is elaborative reduplication. It is found in Tai languages like Lao, as well as Austroasiatic languages like Vietnamese and Semelai. Another domain of morphological complexity in MSEA is the derivational morphology found in languages like Khmer (though unproductive), Semelai and Kmhmu (fairly productive), and Kri (somewhere in between). The real generalization that can be made is that there is no obligatory inflectional morphology in MSEA languages: no case marking, no agreement marking, and no grammatically obligatory tense or aspect marking.

Another important point to register is that MSEA languages do not rely in a straightforward way on a simple constituent order rule to distinguish between core grammatical functions A, S, and O. The widespread idea that they do is, I think, based on the presumption that it *must be* the case, since there is no agreement or case marking to rely on for distinguishing between *who* and *whom*. It is assumed that hearers of isolating/analytic languages have no choice but to rely on strict constituent order to maintain informational coherence in predicate argument relations. This claim is, however, weak at best, as (1) a language may have an array of different argument structure constructions, with variation in the mapping of functional role to syntactic position, (2) extensive ellipsis (including zero anaphora) and movement create widespread surface ambiguity as to the mapping of functional role to syntactic position, even within examples of a single construction, and (3) this variation and ambiguity does not appear to hamper communication. Accordingly, for Mandarin, Li and Thompson (1981: 26) state that ‘no basic word order can be established’. On Riau Indonesian, Gil (2005) argues that observed constituent order patterns in the language are epiphenomenal (cf. also LaPolla 1994).

Writing about constituent order in Khmu clauses, Premsrirat (1987: 60–1) writes that A V O is ‘the normal order of transitive clauses’. She adds, however, that ‘for emphasis’ the object can often occur clause initially, resulting in O A V order being ‘also common’. See the following examples:

- (3.1) *rpà:ng nò: tám*
 gong they beat
 ‘The gong, they beat (it).’
 Khmu | AA | Premsrirat 1987: 60

- (3.2) *kó:n mà? nò:k*
 child mother beat
 ‘The child was beaten by the mother.’
 Khmu | AA | Premsrirat 1987: 61

Premsrirat adds that the ‘reversed’ order of the unmarked AVO – that is, O V A may be used ‘for special emphasis’. Systematic analysis of corpus data is needed in order to understand what ‘special emphasis’ means. Here are two examples with OVA order:

- (3.3) *s?ɔ:y péy ?o?*
 tree cut.down 1
 ‘I cut down a tree.’
 Khmu | AA | Premsrirat 1987: 61

- (3.4) *ré² r̩əm ɔ²*
 farm clear 1
 ‘I clear the land for crops.’
 Khmu | AA | Premsrirat 1987: 61

See Enfield (2007) for details of a similar situation in Lao.

Having said that there is no case marking in MSEA languages, I will note here that Kruspe describes a system in Semelai in which transitive subject arguments are marked with a proclitic (Kruspe 2004: 259). Here is an example (with the proclitic in bold):

- (3.5) *ki goy **la** knlək hn bantal*
 3A bring A husband 3POSS pillow
 ‘Her husband brought the pillow’
 Semelai | AA | Kruspe 2004: 259

If it is a case system, it is a simple one. Note also that the relation between the proclitic and the A function in clausal syntax is not straightforward: ‘Not all post verbal NPs marked with *la* are coding the A; *la* also expresses causation or reason’ (Kruspe 2004: 259). This seems to suggest that *la* marks the participant that is responsible for the occurrence of an event, regardless of whether it is core or peripheral in a syntactic sense.⁵

Languages on the western fringe of MSEA show phenomena that are described using the word ‘case’, but these do not appear to be morphological case inflections in the core sense of that term (i.e., affixes that mark the function of noun phrases in their clause; Dixon 2010: 43). The terminology of case marking has been used in LaPolla’s (2003) description of an ‘anti ergative’ marker *thà?* in Lahu. The function of this marker is ‘to mark an animate argument that might otherwise be interpreted as an actor as being something other than an actor’ (LaPolla 2003: 34). Here are two examples:

- (3.6) *ŋà thà? tâ dʒ?*
 1SG OBJ NEG.IMP hit
 ‘Don’t hit me’
 Lahu | TB | Matisoff 1973a: 156; glossed by LaPolla (2003: 34)

- (3.7) *li? chi ŋà thà? pî?*
 book that 1SG OBJ give
 ‘Give me that book’
 Lahu | TB | Matisoff 1973a: 157; glossed by LaPolla (2003: 34).

⁵ I thank Weijian Meng for suggesting this interpretation.

The morpheme *thà?* marks *yà* as a patient in the first example, and as a recipient in the second. Being at the top of the animacy hierarchy, ‘I’ might normally be expected to refer to a prominently agentive entity, and the ‘anti ergative’ marker here serves to discount this possibility. Jinghpaw (TB, NE India/Myanmar/China) has a similar marker. The Jinghpaw marker *phé?* is used in a transitive clause ‘when there is a possibility that [an O argument] may be misconstrued with A’ (Kurabe 2017: 1001). (Jinghpaw has nine other relational markers essentially, postpositions all of which mark oblique relations such as locative, ablative, allative, and comitative.) In a kind of mirror image of this system, Mongsen Ao (TB, Nagaland, northeast India) has a single ‘case marking’ clitic for marking core arguments (Coupe 2017: 289–90). The agentive *nō* is used ‘in pragmatically determined situations to mark the actor arguments of both monovalent and bivalent predicates’, suggesting that it is a subject marker. Seven other relational clitics mark oblique relations: instrumental, dative, benefactive, comitative, allative, ablative, and locative. These are grammaticalized forms of body part terms and relational nouns in erstwhile N N compounds (Coupe 2017: 290). They are similar to what are described as denominal adpositions in many other languages.

These examples from Tibeto Burman languages on the western fringe of MSEA are systems of dependent marking, but they do not fit a straightforward definition of ‘case’. They involve independent markers rather than bound inflectional morphemes, and they are generally sensitive to semantic and pragmatic factors such as agency and animacy rather than having purely syntactic status.

A final example of a claim that there is ‘case’ marking in MSEA languages concerns Chinese. The Mandarin Chinese morpheme *ba* has sometimes been described as a special kind of case marker. Examples (3.8) and (3.9) have the same verbal and nominal elements. In example (3.8), the compound verb *qi si* ‘angry dead’ serves as a high effect transitive verb. Example (3.9) shows a rephrasing of this using *ba* as a direct marker of the effected undergoer in a reshuffled constructional variant:

- (3.8) *Wo qi si ta le*
 I angry dead s/he ASP
 ‘I made him/her angry to death.’
 Chinese | Sinitic | Liu 2013: 2230

- (3.9) *Wo ba ta qi si le*
 I ‘ba’ him/her angry dead ASP
 ‘I made him/her angry to death.’
 Chinese | Sinitic | Liu 2013: 2230

Li and Yip (1979) argue that *ba* is an absolutive marker, but Liu (2013) argues against this position, and treats it as an object marker. Many other things have been said about

ba, including that it is not a case marker at all, but a grammaticalised serial verb (see Chapter 6).

3.4 Syntax

Information on topics that come under the rubric of syntax is distributed across a number of relevant sections of this book. Chapter 6 deals with the linguistic function of reference and is therefore concerned with matters of nominal syntax, including classifier constructions, relative clause constructions, and head modifier relations. Chapter 7 deals with the linguistic function of predication and is therefore concerned with matters of clausal syntax, including verb serialization, complementation, and valency changing derivations such as syntactic causatives. By way of orientation, in this section I summarize some of these syntactic phenomena and point to the sections where they are further discussed.

In section 3.1, above, I defined the basic elements of clausal syntax in terms of the syntactic primitives A, S, and O. Two of the three basic ways in which languages distinguish A and O in transitive clauses – case marking and verb agreement – are almost entirely absent from the grammars of MSEA languages (apart from Tibeto-Burman languages at the western edge of MSEA). The third main way is constituent order. MSEA languages do make use of constituent order for distinguishing A and O, but there is also a high tolerance for ambiguities that arise from zero anaphora and pragmatically conditioned constituent order variation (see discussion in section 3.3 and the beginning of Chapter 6).

Relative clause constructions allow speakers to take a clausal predication and derive from it a modifier of an argument (as in the noun phrase ‘the rice that we bought’). MSEA languages have relative clause constructions that are relatively simple in structure and quite liberal in the range of argument roles that they can relativize upon (see section 6.1.2).

Complement clause constructions allow speakers to take a clausal predication and use it as the complement of another clause (as in ‘We saw [you drinking beer]’ or ‘He knows [we are here]’). MSEA strategies for clause complementation are relatively simple in structure, mostly involving unmarked sequences of verbs, in a form of serial verb construction or multi verb construction. Complement clause constructions are discussed in section 7.3.4.

Serial verb constructions more broadly are heavily relied on in MSEA languages for a range of functions which in other languages might be handled by disparate clause combining strategies. Numerous types of serial verb construction are described in section 7.3. These range from clause chain like complex event descriptions, with

coordinating properties, to adverbial/depictive like constructions in which predicates show a more subordinating type of relation.

Voice like phenomena provide speakers with ways to deal with non canonical alignments of arguments to syntactic roles, as well as increases or reductions in the number of arguments that are expressed with a given predicate. Reflexives and reciprocals are discussed in section 5.5.3 and section 7.4. These are typically handled in MSEA languages using dedicated stand alone nominals translatable as ‘self’ and/or ‘each other’. Valency increasing strategies (causatives) are discussed in section 7.4. Valency decreasing strategies (akin to passives in some ways) are discussed at the beginning of Chapter 7.

3.5 Basic Phrasal Constituent Order Patterns

We examine in later chapters many of the typological features listed above. The rest of this chapter presents a preliminary overview of some of the basic features of constituent order in the organization of phrases and clauses in MSEA languages.

Austroasiatic and Tai languages are generally head initial, while Chinese, Tibeto Burman, and Hmong Mien exhibit a mix of head initial and head final structures. Languages to the north and west of the core MSEA area are generally head final (Japanese, Mongolian, Uralic, Indo Iranian, Dravidian). Here we see that the area is firmly split in terms of the direction of headedness in phrases (Comrie 2007: 23). Tibeto Burman languages, like East Asian languages (e.g., Japanese, Korean, Mongolian, but not Sinitic) and Indic languages, tend to show object verb order in main clauses. They are, accordingly, head final in other areas of phrasal grammar, such as in nominal constructions including relative clause constructions and possessive constructions. By contrast, languages of core MSEA, such as Khmer, Brou, or Semelai, show verb object order in main clauses and are similarly head initial in other areas of phrasal grammar. We shall see that not only does the region show a differential distribution of head initial and head final structures, but such a differential distribution is observed within languages as well, resulting in systems whose mixed appearance is often attributed to language contact.

For example, while Tibeto Burman languages are almost exclusively verb final in main clauses, Karenic languages depart from this pattern, in having verb object clausal constituent order. Here are two examples from Sgaw Karen:

(3.10) *jə¹ tɔ² jə¹ Jɔ¹*

1.SG hit 1.SG older brother

‘I hit my older brother.’

Sgaw Karen | TB | Lucey Weinhold 2011: 79

- (3.11) *mr¹tr¹ya⁶ pɔ⁵ pʰɔ¹θa⁵ lɛ⁵*
 who eat apple Q
 ‘Who ate the apple?’
 Sgaw Karen | TB | Chapman 2011: 88

Here are two examples from Pwo Karen:

- (3.12) *jə ke tha: lai:*
 1SG write PVP letter
 ‘I wrote a letter (for a certain purpose).’
 Pwo Karen | TB | Kato 2003: 637
- (3.13) *θa ɻwa the thwi:*
 Thawa kick dog
 ‘Thawa kicked the dog.’
 Pwo Karen | TB | Kato 2003: 639

And two examples from Eastern Kayah Li:

- (3.14) *ɻa chū ɻa thwi*
 3 stab 3 dog
 ‘He stabbed his own dog.’
 Eastern Kayah Li | TB | Solnit 1997: 149
- (3.15) *lɔ be ɻa kuklɔ*
 rock strike 3 head
 ‘A rock struck his head.’
 Eastern Kayah Li | TB | Solnit 1997: 130

This distinct profile of Karenic languages may be an effect of language contact, given that these languages have long been in close contact with Mon and Tai languages (Matisoff 1991b).

Another example of diversity in constituent ordering within a language family comes from the Tai Kadai languages. The basic organization of phrases in Tai Kadai languages spoken in China is different from those spoken further south. While Tai Kadai languages are verb object in clause structure, there is variation in headedness of other types of structure, including relative clause constructions and classifier constructions (see below). Again, language contact effects are often posited as the likely cause.

Now we turn to some examples illustrating differences in constituent order in MSEA languages in basic grammatical structures.

3.5.1 Relative Position of Verb and Object

In ordering constituents in a basic clausal predicate, languages of most MSEA language families show verb object order, the only exception being Tibeto Burman (except for the Karenic branch).⁶

V-O

Following are examples showing V O ordering in a wide range of MSEA languages.

- (3.16) *dèèng3 kin3 khaw5 lèèw4*

D. eat rice finish

‘Deng has eaten.’

Lao | TK

- (3.17) *mae: hēn khon nan*

mother see person that

‘Mother saw that person.’

Thai | TK

- (3.18) *khñom sɔm²aat phteh*

1SG clean house

‘I clean the house.’

Khmer | AA | Ehrman 1972: 29

- (3.19) *Tus dev tom tus npua*

CLF dog bite CLF pig

‘The dog bit the pig.’

Hmong | HM | Clark 1989: 178

- (3.20) *kéuih ngoi ngóh*

3SG love 1SG

‘S/he loves me.’

Cantonese | SN | Matthews and Yip 1994: 68

- (3.21) *mẹ yiêu con*

mother love child

‘I love you.’ (said by mother to child)

Vietnamese | AA

⁶ Bradley (2002: 77) also notes that Bai, often classed as a Burmese-Lolo language, shows SVO syntax ‘as an alternative possibility’ (see also Bradley 2002: 96: ‘some dialects [of Bai] have SVO order’).

- (3.22) *tɔŋgaj trənau kau jɔ' tɔlaat*
 tomorrow 1SG go market
 'Tomorrow I will go to the market.'
 Ngeq | AA | Enfield fieldnotes

- (3.23) *tsɛ² mɔ⁶ chi¹ fja¹*
 sister he comb hair
 'His sister was combing her hair.'
 Mulao | TK | Wang and Zheng 1993: 79

O-V

Less common is O V ordering, as shown in these Tibeto Burman language examples:

- (3.24) *y᷑ y᷑ te ch᷑ ve*
 3SG house make PROG
 'He is building a house.'
 Lahu | TB | Matisoff 1991a: 404

- (3.25) *Thu sa yei pa te*
 3SG.M letter write polite v.s.
 'He wrote a letter.'
 Burmese | TB

3.5.2 Relative Position of Modifier and Head in Noun Phrases

In constructions that connect an adjective or equivalent property concept word to a head noun, again most languages show head initial order. Notable exceptions are Sinitic languages, which show head final order in many modifier modified phrases. This is also the case in some non Sinitic languages spoken in China.

N-A

The following examples show the predominant ordering of modifer and head in noun phrases:

- (3.26) *khon2 suung3*
 person be.tall
 'tall person'
 Lao | TK

- (3.27) *menūs l̪aa*
 man good
 ‘good man’
 Khmer | AA | Capell 1979: 5
- (3.28) *rapyā kariār*
 girl beautiful
 ‘the beautiful girl’
 Palaung | AA | Milne 1921: 38
- (3.29) *hɔe² hnòk*
 house large
 ‘a large house’
 Mon | AA | Bauer 1982: 333
- (3.30) *nhà nhó*
 house small
 ‘a small house’
 Vietnamese | AA | Thompson 1965: 221
- (3.31) *tsən¹ ya:i³*
 hair long
 ‘long hair’
 Mulao | TK | Wang and Zheng 1993: 82
- (3.32) *yei æi or yei ei*
 water cold water cold
 ‘cold water’
 Burmese | TB | Okell 1969: 43

A-N

Less common in MSEA is adjective noun order in noun phrases. This pattern is standard in Sinitic languages, such as in this example from Taiwanese, in which the adjectival modifier precedes its nominal head:

- (3.33) *chit tâi sin chhia*
 this CLF new car
 ‘this new car’
 Taiwanese | SN | Lin 2015: 150

This sort of structure is also found closer to the core MSEA area, often with a particle that links the modifier and the head. In this example from Cantonese, the linking particle *ge* is used between the nominal head and its preposed modifier:

- (3.34) *dākyi ge gáu jái*
 cute LP puppy
 ‘cute puppy’
 Cantonese | SN | Matthews and Yip 1994: 88

Similar structures are found in Kam Tai languages (which are otherwise head initial) spoken in contact with Sinitic languages. Here is an example from Mulao (a Kam Sui language spoken in China):

- (3.35) *səp⁸ at⁷ hən³ lo⁴ kɔ ku³*
 ten CLF very big PCL drum
 ‘ten very big drums’
 Mulao | TK | Wang and Zheng 1993: 87

There are cases in which speakers of a single language can be observed using both head final and head initial structures for the same type of grammatical function. See section 6.1 for further discussion.

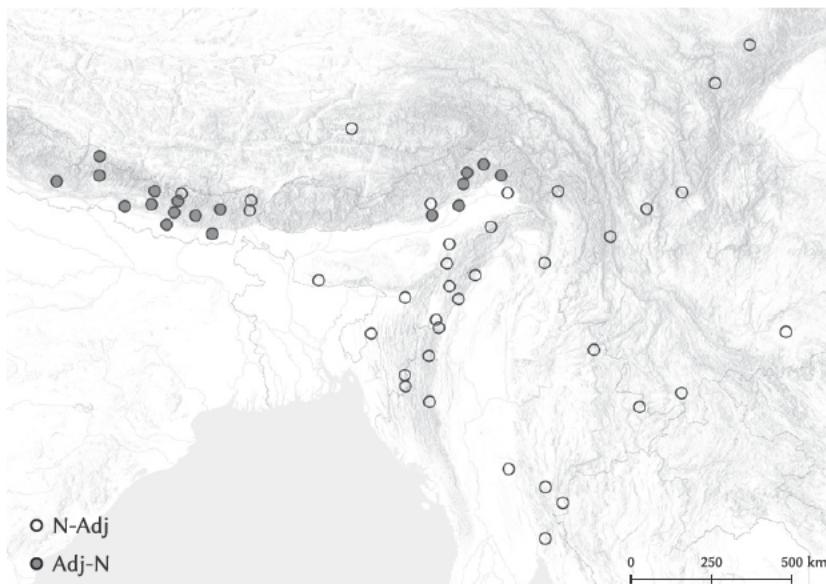
Among Tibeto Burman languages in MSEA, head initial N A order in noun phrases is the usual pattern. Tibeto Burman languages spoken to the west of the northern part of Myanmar (considered here as part of greater MSEA but outside of core MSEA) do show A N order. Map 3.1 shows that A N order in Tibeto Burman languages is found only to the west of the Myanmar/India border (Dryer 2003: 45)

In Idu, for example (spoken in Arunachal Pradesh, northeast India), ‘Adjectives generally precede the noun qualified’ (Pulu 1978: 16):

- (3.36) *icia mazi*
 small goat
 ‘small goat’
 Idu | TB | Pulu 1978: 16

In Tshangla (spoken in Eastern Bhutan), going just beyond the perimeter of MSEA, adjectives ‘may either precede or follow the head noun’. ‘Pre nominal adjectives are restrictive; post nominal adjectives are non restrictive and descriptive of the head’ (Andvik 2003: 441). Here are examples showing the two structures, respectively:

- (3.37) *dukpu waktsa khepa*
 poor child TOP
 ‘the poor child’ (identification)
 Tshangla | TB | Andvik 2003: 441



Map 3.1 *Order of adjective and noun in Tibeto-Burman languages of northwestern MSEA and beyond*
After Dryer 2003: 46, Map 3.1.

- (3.38) *waktsa dukpu khēpa*
 child poor TOP
 ‘the child, who is poor’ (description)
 Tshangla | TB | Andvik 2003: 442

3.5.3 Relative Position of Noun and Relative Clause

The placement of relative clauses in relation to the head nouns that they modify generally shows the same pattern as head modifier relations in noun phrases, discussed in the last section. In some languages, however, the Sinitic head final pattern is followed.

N-Rc

The following are examples from MSEA languages in which a relative clause or similar structure follows the head that it modifies:

- (3.39) *qan3 khōòj5 hēn3*
 CLF 1SG see
 ‘the thing I saw’
 Lao | TK

- (3.40) *sách mà anh mua hôm quyển ở đâu*
 book ALP 2SG buy yesterday stay where
 ‘Where is the book you bought yesterday?’
 Vietnamese | AA | Capell 1979: 10
- (3.41) *siøvphòu dael lo:k msøl ten nøрю ?ae na*
 book which 2SG yesterday buy where
 ‘Where is the book you bought yesterday?’
 Khmer | AA | Capell 1979: 10
- (3.42) *ib tug dev tom npua*
 one CLF dog bite pig
 ‘a dog who bites pigs’
 Hmong | HM | Clark 1989: 183
- (3.43) *ŋwa¹ lik⁸ na:n⁴*
 dog chase deer
 ‘a (deer)hunting dog’
 Mulao | TK | Wang and Zheng 1993: 29

Re-N

In a subset of MSEA languages, mostly Sinitic or Tibeto Burman, relative clauses precede their heads. Here are some examples:

- (3.44) *ngóh chéng ge gūngyāhn*
 1SG hire LP maid
 ‘the maid I hire’
 Cantonese | SN | Matthews and Yip 1994: 88
- (3.45) *mə ‘hma tε’ ‘sheshəya*
 not mistake makes doctor
 ‘a doctor who makes no mistakes’
 Burmese | TB | Capell 1979: 11
- (3.46) *ɳa:u³ twa³ jəu⁶ fɔŋ¹ jəu⁶ kya:ŋ¹ kɔ ɳa:n² ɳwa⁴*
 one CLF both tall and bright PCL house tile
 ‘a tiled house which is both tall and bright’
 Mulao | TK | Wang and Zheng 1993: 87

Mulao, a Kam Sui language spoken in Guangxi, appears in both lists, showing a mixed profile. See section 6.1 for further discussion.

3.5.4 Relative Position of Possessor and Possessed

In possessive phrases, such as *Kim's house*, possessors and possessees (i.e., possessed entities) are combined in a single phrase. (These are labelled Pr and Pe respectively; Aikhenevald and Dixon 2013.) The same structures are used when expressing part whole relations (e.g., *the arrow's tip*). A possessor and a possessed entity can be combined by morphological means, for example using an affix (such as English 's in *Kim's house*) or a dedicated pronoun (such as English *his* in *his house*). Or it may be done by syntactic/constructional means (such as English *of* in *the roof of the house*). Languages vary in their relative ordering of possessor and possessed in these phrases.

Pe-Pr

Following are some examples of possessive phrases in MSEA languages in which the possessee precedes the possessor, following the head initial structure:

- (3.47) *huan2 (khòòng3) khòòj5*
 house of 1SG
 'my house'
 Lao | TK

- (3.48) *ptěah (rəbəh) kñom*
 house of 1SG
 'my house'
 Khmer | AA

- (3.49) *nhà (cùa) tôi*
 house of 1SG
 'my house'
 Vietnamese | AA

- (3.50) *pu⁴ niu²*
 father 1SG
 'my father'
 Mulao | TK | Wang and Zheng 1993: 59

Pr-Pe

And here are some languages that use the opposite ordering. These include not only Sinitic and Tibeto Burman languages, but also Hmong Mien languages and those Tai languages that are in intensive contact with Sinitic:

- (3.51) *koj (lub) tsev*
 2SG CLF house
 ‘your house’
 Hmong | HM | Clark 1989: 185
- (3.52) *ngóh sailóu*
 1SG YBR
 ‘my younger brother’
 Cantonese | SN | Matthews and Yip 1994: 107
- (3.53) *hohksāang ge gājéung*
 student POSS parents
 ‘the student’s parents’
 Cantonese | SN | Matthews and Yip 1994: 107
- (3.54) *na² təm¹ tau²*
 2SG chest
 ‘your chest’
 Mulao | TK | Wang and Zheng 1993: 62
- (3.55) *ù tháñ yé ámyiñ*
 U T POSS view
 ‘U Thant’s view’
 Burmese | TB | Okell 1969: 169

3.5.5 Relative Position of Adposition and Noun

Non core phrases express elements of a situation such as time or place. When added to the core of a clause, these are often marked by an adposition or equivalent. English non core phrases are marked by prepositions such as *at* or *in* (e.g., *John left at three o’clock* or *They cooked dinner in a hut*). Languages differ with respect to whether their adpositions appear before or after the phrase that they link to the clausal core (hence, prepositions versus postpositions).

Preposition-N

Numerous MSEA languages follow the head initial principle in prepositional (or equivalent) phrases, as shown in the following examples:⁷

⁷ In the Lao example, both *juu1* ‘be.at’ (a verb) and *theng2* ‘top of’ (a noun) can be considered as prepositions in functional terms.

- (3.56) *man² kin³ khaw⁵ juu¹ theng² phuu²*
 3SG eat rice be.at top mountain
 ‘S/he eats rice on top of the mountain.’
 Lao | TK

- (3.57) *sui⁵³ ηau³³ te³²³ met³¹ ta³³*
 sit at below tree that
 ‘Sit underneath that tree.’
 Dong | TK | Long and Zheng 1998: 141

N-Postposition

Other languages notably, in the Sinitic and Tibeto Burman language families feature postpositions (though not necessarily exclusively, with varieties of Chinese combining noun based postpositions with verb based prepositions). Here are some examples:

- (3.58) *kéuih hái jáudim chēutbihn dáng ngóh*
 3SG be.at hotel outside wait 1SG
 ‘S/he is waiting for me outside the hotel.’
 Cantonese | SN | Matthews and Yip 1994: 118

- (3.59) *chò kà?*
 here LOC
 ‘here, hither, hence.’
 Lahu | TB | Matisoff 1973a: 164

3.5.6 Relative Position of Adjective and Standard of Comparison

In languages with comparative constructions (e.g., English *this is old* → *this is older than that*), the adjective (or equivalent property concept word) may be placed either before, or after, the standard of comparison.

Adj-SComp

The following examples show MSEA languages in which the standard of comparison follows the property concept word:⁸

⁸ The comparative marker is often derived from a verb meaning ‘surpass’ or ‘exceed’ (Ansaldi 1999). This pattern is observed in languages from other parts of the world as well.

- (3.60) *suung³ kua¹ caw⁴*
 tall surpass 2SG
 ‘taller than you’
 Lao | TK
- (3.61) *fɔŋ¹ ta⁶ ja²*
 tall over 2SG
 ‘taller than you’
 Mulao | TK | Wang and Zheng 1993: 56
- (3.62) *câu hon anh*
 tall more.than 2SG
 ‘taller than you’
 Vietnamese | AA
- (3.63) *A wàhn leng gwo kéuih mùihmúi*
 A W pretty than 3SG sister
 ‘Wan is prettier than her sister.’
 Cantonese | SN | Matthews and Yip 1994: 167
- (3.64) *niəng nih l'ɔɔ ciəng niəng nuh*
 girl this pretty surpass girl that
 ‘This girl is prettier than that girl.’
 Khmer | AA

SComp-Adj

As the following examples show, the opposite ordering is found in Sino Tibetan languages, and in some languages that are in intense contact with them, such as Dong, a Tai language spoken in China:

- (3.65) *A wàhn béi kéuih mùihmúi leng*
 A W than 3SG sister pretty
 ‘Wan is prettier than her (younger) sister.’ (formal)
 Cantonese | SN | Matthews and Yip 1994: 167
- (3.66) *xiǎo Wǎng bǐ tā mèimei piàoliang*
 little W than 3SG younger sister pretty
 ‘Wang is prettier than her younger sister.’
 Mandarin | SN

- (3.67) *mau³³ pi³²³ ηa²¹² phay³⁵*
 3SG compare 2SG tall
 ‘He is taller than you.’
 Dong | TK | Long and Zheng 1998: 141

- (3.68) *ŋà a kε' N V*
 I more.than N v
 ‘n is more v than me’
 Lahu | TB | Matisoff 1973a: 131

- (3.69) *mì kì chi N a kε' a ci' dà? ve*
 chair this N more.than more good PCL
 ‘This chair is better than n.’ (Matisoff 1973a: 131)
 Lahu | TB | Matisoff 1973a: 131

3.6 Sentence-Type Distinctions

All languages need to distinguish between the basic speech acts of asserting, questioning, and commanding. These speech act types correlate with the grammatical sentence types interrogative, imperative, and declarative (Sadock and Zwicky 1985). The languages of MSEA are fairly uniform in terms of how they signal these basic distinctions in sentence type. Interrogatives tend to be marked by *in situ* question words for content questions and sentence final particles for polar questions. Both imperatives and declaratives tend to be marked by sentence final particles.

Among interrogatives, there are two major types: polar questions (‘yes no’ questions) and content questions (‘WH’ questions). Polar questions in MSEA tend to be formed using final particles. Here are examples from all major language families, with the final particle highlighted:

- (3.70) *nɔŋ^{f13j} tɛ^hi vəɻ^{f3-4j}*
 you go Q?
 ‘Did you go?’
 Shanghaiese | SN | Zee and Xu 2003: 142

- (3.71) *dza³³ dza²¹ lə²¹*
 rice eat Q
 ‘Have you eaten?’
 Lisu | TB | Bradley 2003: 234

- (3.72) *Nin daih nza?*
 nin³³ tai³¹ **dza**³³
 3SG come PTC
 ‘Is he coming?’
 Ruyuan Mien | HM | Liu 2016: 130
- (3.73) *pò kù:n plà:ŋ rò:t kì há*
 2PL see PN come PROX Q
 ‘Have you seen Plàaj coming here?’
 Kammu | AA | Svantesson and Holmer 2015: 971
- (3.74) *?uŋ²¹k^hiu²¹ sat²⁴ za:j³³ te?²⁴*
 doctor true come Q
 ‘Is the doctor really coming?’
 Hainan Cham | AN | Thurgood *et al.* 2014: 222
- (3.75) *Chǔm kung rang chuh non rang mok mai*
 animal that 3.INDEF to.hunt that 3.INDEF to.bring towards
bong hu mǐn
 to.eat to.be.able.to YN.QM
 ‘Can they eat animals that they hunted?’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 17
- (3.76) *ɳ² pɔ¹ ta⁶ pak⁷kɪŋ¹ ma⁶*
 2SG go ASP Beijing PART
 ‘Have you been to Beijing?’
 Chadong | TK | J. Li 2008: 610

There are other means of forming polar questions in these languages. For example, in a number of Hmongic languages, a polar question is formed by placing a question word after the subject and before the verb, as in this White Hmong example:

- (3.77) *koj puas nqis los hauv no na?*
 2SG Q descend come inside here PCL
 ‘Are you coming down here?’
 White Hmong | HM | Jarkey 2015: 57

Another option found in Hmongic languages is a V not V strategy, widely known from Sinitic languages (Clark 1989: 209). Here is an example from Miao:

- (3.78) *moy⁵⁵ moy¹¹ a⁵⁵ moy¹¹ nen³⁵?*
 2SG go not go SFP
 ‘Do you want to go or not?’
 Hmong | HM | Clark 1989: 209

We turn now to content questions. English has a typologically unusual pattern of forming content questions. This involves a construction in which question words such as *who* or *what* occupy a distinct syntactic position. We tend not to say *You ate what?* but rather *What did you eat?* In MSEA languages, by contrast, content questions are normally formed using *in situ* question words. The question words appear in the same slot that their corresponding referring expressions would appear in. Here are some examples of MSEA content questions with *in situ* question words:

- (3.79) *nøy^{/13]} ma ts₂^[1 3] saməðzr^{/3-5-1]}*
 you bought ASP what?
 ‘What did you buy?’
 Shanghaiese | SN | Zee and Xu 2003: 142

- (3.80) *a⁴⁴le⁴⁴ (t^hi²¹) ma⁴⁴ la³³ a⁴⁴ (a²¹) [a⁴¹]*
 which? one CL come IMPFV Q
 ‘Which one is coming?’
 Lisu | TB | Bradley 2003: 233

- (3.81) *mouj² kɔŋ¹ haj⁴kai¹*
 2SG say what
 ‘What did you say?’
 Vietnamese Mien | HM | Doan and Mai 1992: 98

- (3.82) *?à? cò pà? máh*
 1DU IRR eat what
 ‘What shall we eat?’
 Kammu | AA | Svantesson and Holmer 2015: 972

- (3.83) *Rang padär hagék manung samrap*
 3.INDEF to.use what some for.the.purpose.of
 ‘What (material) do they use to (hunt animals)?’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 18

- (3.84) *man² van^{1?}qun¹ ɳa:u⁶ ju¹ka:i¹ ts^hjeŋ⁵ ni⁴nam² ?na:k⁷ se¹ ?ni³*
 3SG yesterday at town sing what to 2PL listen
 ‘What was s/he singing to you in town yesterday?’
 Maonan | TK | Lu 2008: 179

- (3.85) *kaw³³ ɻa³³za:y³³ lɔ²¹*
 1P who PFV
 ‘Who am I?’

Hainan Cham | AN | Thurgood *et al.* 2014: 226

3.7 Sentence-Final Particles

All languages make use, to a greater or lesser extent, of sentence final elements, including tag questions and particles, to mark sentence types or to refine the pragmatic meanings of utterances. MSEA languages feature sets of sentence final particles that are relied upon heavily in language usage, especially in dialogue. The above examples of polar questions all illustrated the use of sentence final particles for grammatical marking.

Inventories of sentence final particles in MSEA languages are large and structured systems. In Lao, for example, more than twenty five particles are divided into three main categories depending on the type of sentence function they are associated with interrogative, factive, imperative showing differences in meaning within those categories, invoking subtleties of aspect, evidentiality, and illocutionary force (Enfield 2007: ch. 4). In Cantonese, there are some thirty basic particles, with over seventy possible combinations of two or three of these particles (Kwok 1984, Luke 1990). In their summary of the system, Matthews and Yip (1994: 338–56) identify a set of features that are typical of these systems across MSEA. These are shown in Table 3.3.

As Matthews and Yip point out, sentence final particles present significant challenges for analysts. This is an area for special attention from grammatical and semantic typologists.

Here are further examples, also including other sentence types such as declaratives and directives.

- (3.86) A: *léih mhóu joi máaih yéh la*
 you don’t again buy stuff PCL
 ‘Don’t you go buying more stuff.’
- B: *ngóh tái háh jē ma*
 I look del PCL(playing:down) PCL(of course)
 ‘I’m just going to take a look.’
- Cantonese | SN | Matthews and Yip 1994: 344

- (3.87) *lāahmjái tek bō tūhngmàaih léuihjái tek bō*
 boy kick ball and girl kick ball
móuh māt fānbih^t ge je bo
 not:have any difference PCL(assert) PCL PCL(invite agrmnt)

‘Boys play football and girls play football – what’s the difference?’

Cantonese | SN | Matthews and Yip 1994: 345

Table 3.3 *Features of Cantonese sentence-final particles (SFPs)*

Feature	Description
Speech act functions	SFPs indicate speech-act type distinctions (interrogative, declarative, imperative), as well as distinctions in evidentiality, and affective and emotional colouring.
Similar in function to intonation	SFPs convey the kind of information that is conveyed by intonation in other languages (such as English) (sometimes said to compensate for lack of freedom to intone, due to lexical tone).
Frequent, colloquial	SFPs are very frequent and important in colloquial speech.
Elusive meanings	Analysts find it difficult to agree on whether SFPs belong in semantics, pragmatics, or discourse.
Phonologically free from constraints of tone	SFPs sometimes have a non-arbitrary relationship to tone (some have claimed that there are ‘tonal particles’ (this has been claimed for Mandarin which has only about seven particles, which cannot stack); many of the Cantonese particles have ‘tonal variants’; ‘Typically, the high-tone variants are more tentative, the low-tone ones assertive and the mid-tone ones neutral’).

After Matthews and Yip 1994: 338.

- (3.88) *Kuv tsis hnov koj hais os*

I not hear you say PCL(emph)

‘I didn’t hear what you said!’

Hmong | HM | Clark 1989: 182

- (3.89) *Koj ua dabtsi lawm os ob peb hnub no na*

you do what PFV PCL(emph) two three day this PCL(oh?)

‘What have you been doing these few days?’

Hmong | HM | Clark 1989: 182

- (3.90) *Ngày.mai anh có đi không? Đi chúc*

tomorrow BRO have go PCL(q) go PCL(sure)

‘Are you going tomorrow?’ ‘Of course I’m going.’

Vietnamese | AA | Clark 1989: 182

- (3.91) *sam1 naen1 seic mans nyac jas dung1 xogx laeuc bac*

three CLF cake yellow those DET cook good SFP1 SFP2

‘I guess those three yellow cakes should have been cooked

through (now)?’

Kam | TK | Yang and Edmonson 2008: 570

3.8 Expressive Language

All languages have words, structures, and other resources for expressing aspects of first person experience, such as sensory and emotional experience. Here we survey the phenomena of (1) ideophones and (2) poetic idioms.

3.8.1 Ideophones/Expressives

Many Southeast Asian languages have a distinct word class of ideophones, which have their own syntactic behaviour.⁹ For example, although they are semantically verb like, because they encode information about actions, properties, and states of affairs, they tend to be unlike verbs in their grammatical behaviour. They are often limited in the kinds of tense/aspect/modality marking they can take. Another term for these words is *expressives*. This term is widely used in MSEA linguistics, following Diffloth's pioneering work (Diffloth 1972).

Ideophones can be defined as ‘marked words that depict sensory imagery’ (Dingemanse 2012, 2017). They are found in languages around the world:

Many natural languages have words depictive of sensory perceptions like Japanese *nyoro nyoro* ‘wriggling motion’ and *tsuru tsuru* ‘smooth surface’ (Gomi 1989), Semelai *rɔprap* ‘something large walking through twigs’ and *cərələp* ‘sound of someone/something entering the undergrowth’ (Kruspe 2004), and Gbeya *elele* ‘hair waving gently in a breeze’ and *bakat bakat* ‘sound of sandal flapping’ (Samarin 1970).

(Dingemanse 2017: 364)

As these examples suggest, ideophones often have highly specific semantic values, associated with modification of narrowly defined property concepts (e.g., colours, states) or events. Ideophones are often separate lexical items, usually with two syllables that go together in a rhyming or alliterative pattern. They function as adverbials, often with highly evocative meanings. Writing of Cantonese, Bodomo states that an ideophone ‘mainly serves to help express the vivaciousness of one’s feeling towards the entity one wants to describe’ (Bodomo 2006: 209, original emphasis).

Table 3.4 supplies data from Sgaw Karen, illustrating two of the quintessential properties of ideophones: (1) highly specific imagistic meanings and (2) rhyming and alliterative sound structures.

⁹ This is one of many areas in which MSEA and West African languages pattern similarly.

Table 3.4 *A selection of Sgaw Karen (TB) ideophones*

bla ² bla ²	to have no taste such as in badly prepared soup
səli ² sə ¹ la ¹	to spread out in different directions
ka ²¹ di ² ka ² da ²	to flee from danger in a hasty and unorganized fashion
shye ²³ shyi ² she ² da ²³	(of sound or actions that are) loud and swift
ki ²¹ ki ²	moving (like the surface of water)
pə ²¹ ci ¹ pa ²¹ te ²	(to wait/to see with) impatience
bra ²¹ bri ²³	to have the inside out (such as shirt, bag)
ko ¹ lo ¹	(to come) in groups
po ¹ li ² po ¹ yɔ ²	worn out (like old shirts or rags)
lɔ ² swe ²³ lɔ ² swɔ ²	serious wound that was torn out (as by a tiger)
do ² mo ² do ² kholo ²	bumpy (road)
ka ²¹ ja ² ka ²¹ ja ²	(to cry with) low voice
ka ¹ lo ² ka ¹ la ²¹	to do things in a sloppy way
she ^{23?} ɛ ²³	dishonestly clever
kho ² do ²	(to stay) quietly without care nor enthusiasm
ka ¹ be ²³ ka ¹ be ²³	of actions that are slow
ka ² phu ¹ li ² lu ²	describing actions/situations that are unorderly
cɔ ² klɔ ²¹	describing actions that are clumsy
ka ²¹ swe ² ka ²¹ swo ²	describing soil that can let water pass freely
shi ²³ ke ²	for a little moment
pə ¹ di ²	(to do things) bit by bit
te ² xe ²	see-through (describing paper, cloth etc.)
sə ¹ xi ²¹ di ²¹	(to do things) with care/attention to details

Tabulated from Ratanakul 2001: 131 2, from Ratanakul's 1986 *Thai-Sgaw Karen Dictionary*.

Turning to a Tai Kadai language, some examples from Lao show ways in which ideophones can convey fine distinctions in meaning in specific semantic fields. The following examples show distinct ideophones describing patterns of baldness (as modifiers of the verb/adjective *laan4* ‘bald’):

- (3.92) *laan4 quù4 khùù4*

bald EXPR

‘completely bald’

Lao | TK

- (3.93) *laan4 khim1 mim1*

bald EXPR

‘lightly balding’ (‘with a frog by the pond’)

Lao | TK

- (3.94) *laan4 sameng4 kheng4*
 bald EXPR
 ‘bald across the whole dome’ (‘swidden in the jungle’)
 Lao | TK
- (3.95) *laan4 samook5 khook5*
 bald EXPR
 ‘bald, with fuzz all over, and a protruding forehead and back of head’
 Lao | TK

The next two examples describe contrasting ways in which a person can be seen standing:

- (3.96) *jiùìn3 tok2 pok2*
 stand EXPR
 ‘of old lady, “skinnily” standing, angular’
 Lao | TK
- (3.97) *jiùìn3 sêê1 lêê1*
 stand EXPR
 ‘of someone standing, stooped over “all alone, no one talking to them”’
 Lao | TK

Here are two examples describing manners in which a person can be seen lying down:

- (3.98) *nòòn2 khòò5 lòò5*
 lie EXPR
 ‘lying down like a child asleep’
 Lao | TK
- (3.99) *nòòn2 qêêk1 lêêk4*
 lie EXPR
 ‘lying down lazily, not doing anything’
 Lao | TK

The following three examples describe some ways a person’s hair may appear (note the use of *hêt1* ‘do/make’ as a dummy verb that the ideophone modifies):

- (3.100) *phom3 hêt1 keng4 ceng4*
 hair do/make EXPR
 ‘of messed up hair, “uncombed, not flat, going this way and that”’
 Lao | TK

- (3.101) *phom3 hêt1 phùk1 vuìk1*
 hair do/make EXPR
 ‘of hair bent from being slept on’
 Lao | TK

- (3.102) *phom3 hêt1 khuk1 ñuk1*
 hair do/make EXPR
 ‘of curly, flowing, hair’
 Lao | TK

The next two examples describe different perceptions of the colour white:

- (3.103) *khaaw3 pen4 ven4*
 white EXPR
 ‘white, of face or body, pale, like too much powder applied’
 Lao | TK

- (3.104) *khaaw3 còn1 phòn1*
 white EXPR
 ‘white, of an old person’s completely white hair’
 Lao | TK

And finally, here are some ideophones describing different possible sizes of a hole or perforation:

- (3.105) *huu2 hêt1 cing1 ping1*
 hole do/make EXPR
 ‘tiny, cannot put finger in’
 Lao | TK

- (3.106) *huu2 hêt1 còng1 pòng1*
 hole do/make EXPR
 ‘of a lettuce, with holes in it from snails (can put a finger in)’
 Lao | TK

- (3.107) *huu2 hêt1 côóng pôóng1*
 hole do/make EXPR
 ‘can put hand in’
 Lao | TK

- (3.108) *huu2 hêt1 cang1 paang1*
 hole do/make EXPR
 ‘can fit head through’
 Lao | TK

Table 3.5 *Some ideophones in Chrau*
(Austroasiatic)

hǔl hǔl	‘sitting quietly’
khǔch khǔch	‘many fish’
khyōng khyōng	‘walking stiffly’
klo klo	‘darkly, unclear’
lăq lăq	‘sitting still, sick’
caprăh caprăng	‘scattered’
ravênh rawai	‘dizzy’
mbăq mban	‘unskillfully’
mlăq mlăq	‘dirty’

After Thomas 1971: 155–7.

As these examples should suggest, ideophones are poetic and evocative in nature. Accordingly, their usage is appreciated as a matter of style. Thomas (1971: 155ff.) notes that ‘the judicious use’ of ideophones in Chrau (Austroasiatic) ‘is a mark of an expert story teller’. Table 3.5 shows some Chrau examples.

Further examples of ideophones/expressives from languages across the MSEA area follow:

- (3.109) *tim4* → *tim4.je4.je4*
 ‘sweet’ ‘sweetish’
 Cantonese | SN | Bodomo 2006: 209

- (3.110) *plig plawg* ‘bird rising from a nest on the ground’, as in:
 ib pab noog ya plig.plawg
 one group bird fly IDEO
 ‘a flock of birds fly by making a great flapping’
 White Hmong | HM | Ratliff 1992: 220

- (3.111) *pa:i³ n̥iu⁵n̥iu⁵*
 wag IDEO
 ‘to wag (the tail)’
 Chadong | TK | J. Li 2008: 608

- (3.112) *pia³³ lio²¹ p^hia²¹ p^hia²¹*
 water flow gurgle gurgle
 ‘Water flows gurgling.’
 Hainan Cham | AN | Thurgood *et al.* 2014: 190

(3.113) *cʰŋis*

‘the smell of ammonia, urine’

Semelai | AA | Kruspe 2004: 396

(3.114) *cräläp*

‘the sound of something entering the undergrowth’

Semelai | AA | Kruspe 2004: 396

(3.115) *cnayur*

‘with the appearance of bushiness in several places’

Semai | AA | Kruspe 2004: 396

(3.116) *kərəlɔg lɔg*

‘with the sound of heavy footsteps’

Temiar | AA | Kruspe 2004: 397

3.8.2 Poetic Idioms (Four Syllable Expressions)

Another form of expressive language that is widespread in MSEA languages involves four syllable expressions which combine patterns of rhyming or alliteration with patterns of semantic interrelations such as parallelism, association, and combinations of metaphor and metonymy. These are well known and extensively documented in Sinitic languages (in Mandarin, these four syllable expressions are called *cheng yu*).

Liem (1970) lists over 800 four syllable expressions in Vietnamese, dividing them into a range of formal categories. For example, one set is of the form VERB OBJECT VERB OBJECT, two verb phrases with meanings that together point to the higher level meaning of the expression as a whole. Here are some examples:

(3.117) *cám chợ ngǎn sông*

forbid market block river

‘to be isolationist’

Vietnamese | AA | Liem 1970: 9

(3.118) *buôn mây bán gió*

buy cloud sell wind

‘engage in risky commerce’

Vietnamese | AA | Liem 1970: 9

(3.119) *cày mây cuốc gió*

plough cloud hoe wind

‘have a free life in the open air’

Vietnamese | AA | Liem 1970: 9

Another set takes the form NOUN ADJECTIVE NOUN ADJECTIVE, two noun phrases with meanings that together point to the higher level meaning of the expression as a whole. Here are some examples:

- (3.120) *ma thiêng nước độc*
 ghost powerful water poison
 ‘of the highlands’
 Vietnamese | AA | Liem 1970: 21

- (3.121) *mặt to tai lớn*
 face big ear long
 ‘with power, influence’
 Vietnamese | AA | Liem 1970: 21

- (3.122) *mình gầy mặt vỡ*
 body thin face exhausted
 ‘emaciated’
 Vietnamese | AA | Liem 1970: 21

The following examples show patterns in another language family, Tai Kadai, firstly from Lao:

- (3.123) *khaw5 paa3 qaa3haan3*
 rice fish food
 ‘foodstuffs’
 Lao | TK | Enfield 2007: 305

- (3.124) *kép2 phak2 hak2 nòòl*
 collect vegetables snap.off shoots
 ‘Gather edible vegetables and shoots.’
 Lao | TK | Enfield 2007: 305

- (3.125) *khiaw3 sot2 ngot1 ngaam2*
 green fresh EUPH beautiful
 ‘luscious, fresh and green’ (as the Lao countryside in the rainy season)
 Lao | TK | Enfield 2007: 306

And finally, from Dong (Long and Zheng 1998: 162):

- (3.126) *pen⁵³ ta⁵⁵ pen⁵³ neŋ⁵⁵*
 colourful eye colourful nose
 ‘face is very dirty’
 Dong | TK | Long and Zheng 1998: 162

- (3.127) *weŋ²¹²* *kau³²³* *weŋ²¹²* *kha³⁵*
horizontal head horizontal ear
'do not listen to reason'
Dong | TK | Long and Zheng 1998: 162

Phonological Systems

This chapter surveys phonological systems in spoken MSEA languages, their inventories of consonants and vowels, the phonotactic principles by which these segments are used in building syllables and words, and the laryngeal phenomena (including, but not confined to, pitch) that come under the rubric of what can be called *tonation*, following Bradley (Brunelle and Kirby 2016, Kirby and Brunelle 2017, Brunelle *et al.* 2018; cf. *tonality*; Edmondson and Gregerson 1993). While the MSEA area is often cited as having languages whose words are monosyllabic and tonal, often with reference to the cases of Thai, Vietnamese, and varieties of Chinese (which are in any case neither purely monosyllabic nor purely pitch based in tonation), we shall see that many of the area's languages are neither tonal nor overwhelmingly monosyllabic.

4.1 Consonants

Consonant inventories in MSEA languages are relatively large, in world terms.

In some languages, the reason for large inventories is that there are more distinctions in manner of articulation and voicing/laryngeal distinctions than is typologically usual. For example, there are series of pre nasalized stops and voiceless nasals in Dànanshān Miao (see below). Or there may be more than the typologically usual two distinctions in voicing. For example, in Thai there is a contrast between voiced versus voiceless unaspirated versus voiceless aspirated stops. Speakers need to distinguish between sets of words like *baa* ‘shoulder’ (with voiced /b/), *paa* ‘forest’ (with voiceless unaspirated /p/), and *phaa* ‘machete’ (with voiceless aspirated /ph/).

Places of articulation in MSEA language systems tend to be restricted to the typologically usual positions of bilabial, labio dental, alveolar, palatal, and velar consonants, with only a few exceptions. For example, Kri has retroflex stops and fricatives (see below).¹ There is a tendency across the languages of the world for a larger number of

¹ For simplicity I refer to ‘retroflex’ as a place of articulation, though I note that the term describes tongue shape for apico-postalveolar articulation rather than literally a place of articulation.

distinct consonants to occur in syllable initial position than in syllable final position. In many MSEA languages, this tendency is especially marked, with many languages allowing only a dramatically reduced set of consonant contrasts in syllable final position. Below sections illustrate these tendencies in a number of MSEA languages.

Two small points can be noted before we proceed. First, many of the languages show a gap in the inventory of consonants where a voiced velar stop [g] is expected (see examples below). Second, the presence of implosive stops in MSEA has been noted by typologists, as they are rare in other parts of Eurasia (Maddieson 2013, Jenny *et al.* 2015: 23). Surveying data from the *World Atlas of Language Structures* (Dryer and Haspelmath 2013), Comrie (2007: 32) notes that ‘given the rarity of implosives across the languages of the world they are found in only 13.3% the fact that they are so frequent in Mainland Southeast Asia, and virtually lacking in surrounding areas, is a significant fact of areal distribution’. If an MSEA language has implosive stops, this does not mean that the language has an extra series based on the implosive manner of articulation. In Kri, for example, there are implosive voiced stops but no other voiced series. In this case, the implosive stops occur in place of, rather than in contrast to, voiced stops. MSEA languages with implosive stops are not limited to a single language family. The bilabial and alveolar voiced implosive stops [ɓ] and [ɗ] are found, for example, in the Tibeto Burman language Bwe Karen (Weidert 1987: 321), the Austronesian language Tsou (Wright and Ladefoged 1994: 67), the Austroasiatic languages Vietnamese (Thompson 1965: 4) and Kri (Enfield and Diffloth 2009: 11–12), and the Tai Kadai language Hlai (Baoding dialect; Ostapirat 2008: 626). Implosives are almost non-existent in Hmong Mien languages (indeed Hmongic languages tend to lack any voiced stops), but an implosive [ɗ] has developed from a cluster (*ql) in Mo Piu (West Hmongic, north Vietnam; Ly 2012: 46–7).²

Let us begin our thumbnail survey of consonant inventories in MSEA languages with Khmu³ (Austroasiatic, Laos; Premsrirat 1987). Table 4.1 shows the set of consonants that can occur in initial position in a Khmu syllable. The set illustrates the points made above; in world terms, the places of articulation do not form a large or unusual set labial, alveolar, palatal, velar, and glottal while in the manner of articulation and voicing distinctions we see greater elaboration than is typologically average. There are two liquids, and two glides, and a three way contrast in voicing in the stops (though note the gaps: no voiced stops in palatal and velar positions, and also no [f]).

Table 4.2 shows the syllable initial clusters that occur in Khmu, made possible by combining voiceless stops with the trilled [r] (possible at all places of articulation), and more restrictedly with the lateral [l] and the glide [w].

² Thanks to Martha Ratliff for this reference.

³ Marc Brunelle (pers. comm.) notes that these statements are true of Eastern Khmu. Other dialects do not show a surface voicing contrast.

Table 4.1 *Simple syllable-initial consonants in Khmu (AA)*

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	voiceless unaspirated	p-	t-	c-	k-	?-
	voiceless aspirated	p ^h -	t ^h -	c ^h -	k ^h -	
	voiced	b-	d-			
Fricatives	voiceless		s-			h-
	voiced	m-	n-	jn-	ŋ-	
Nasals	trill		r-			
	lateral		l-			
Glides	voiced	w-			j-	

After Premsrirat 1987: 7.

Table 4.2 *Khmu (AA) initial clusters*

		+Trill	+Lateral	+Glide
Labial	voiceless unaspirated stop	pr-		pl-
	voiceless aspirated stop	p ^h r-		
Alveolar	voiceless unaspirated stop	tr-		
	voiceless aspirated stop	sr-		
Palatal	voiceless unaspirated stop	cr-		
	voiceless aspirated stop	c ^h r-		
Velar	voiceless unaspirated stop	kr-	kl-	kw-
	voiceless aspirated stop	k ^h r-		k ^h w-

After Premsrirat 1987: 10.

In contrast to these many alternatives for beginning a syllable in Khmu, the set of possible syllable final segments is restricted.⁴ As Table 4.3 shows, no distinctions in voicing are made in syllable final position, reducing all stops to voiceless unaspirated (and usually checked/unreleased – which syllable final stops tend to be in most MSEA languages).

Further to the constraints shown in Table 4.3, there are no clusters in syllable final position in Khmu. While this shows that the possibilities for syllable final position in Khmu are reduced in comparison to those in syllable initial position, we shall see below from other MSEA languages that Khmu shows a relatively mild degree of asymmetry between possibilities in initial and final position.

⁴ There is a global tendency for fewer consonant phonemes to occur in coda position than in onset position. In a table of example languages from Rousset (2004; cited in Gordon 2016: 98–104), the MSEA language Wa (Tibeto-Burman) shows the greatest discrepancy, with only one in five of the language's consonant phonemes occurring in coda position.

Table 4.3 *Syllable-final consonants in Khmu (AA)*

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	voiceless unaspirated voiceless aspirated voiced	-p	-t	-c	-k	-ʔ
Fricatives	voiceless		-s [ç/x]			-h
Nasals	voiced	-m	-n	-ɲ	-ɳ	
Liquids	trill		-r		-ɳ	
	lateral		-l			
Glides	voiced	-w			-j	

Dashes indicate syllable-initial consonants that do not occur in final position.

Table 4.4 *Set of consonants possible in the initial position of main syllables in Kri (AA)*

	Labial	Alveolar	Retroflex	Palatal	Velar	Laryngeal
Voiceless unaspirated stops	p-	t-	t̪-	c-	k-	ʔ-
Voiceless aspirated stops	p ^h -	t ^h -	t̪ ^h -		k ^h -	
Voiced (implosive) stops	b-	d-		f-		
Fricatives		s-			ɣ-	h-
Nasal sonorants	m-	n-		ɲ-	ɳ-	
Oral sonorants	v-	l-r/z̪-		-j/j-		

Table 4.5 *Final segments of stressed syllables in Kri (AA)*

	Labial	Alveolar	Retroflex	Palatal	Velar	Laryngeal
Voiceless unaspirated stops	-p	-t		-c	-k	-ʔ
Voiceless aspirated stops						
Voiced (implosive) stops						-h
Fricatives						
Nasal sonorants	-m	-n		-ɲ	-ɳ	
Oral sonorants	-w	l-r/z̪-		-j/j-		

Dashes indicate syllable-initial consonants that do not occur in final position.

Kri provides another example from the Austroasiatic family of an MSEA language in which the initial final asymmetry for consonant phonemes is relatively mild. Table 4.4 shows the full set of consonants possible in the initial position of main syllables in Kri. Table 4.5 shows the consonants possible in syllable final position.

Table 4.6 *Syllable-initial consonants in Lao (TK)*

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	voiceless unaspirated	p-	t-	c-	k-	ʔ-
	voiceless aspirated	p ^h -	t ^h -		k ^h -	
	voiced	b-	d-			
Fricatives	voiceless	f-	s-			h-
Nasals	voiced	m-	n-	jn-	ŋ-	
Liquid	lateral		l-			
Glides	voiced	v-		j-		

Table 4.7 *Syllable-final consonants in Lao (TK)*

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	voiceless unaspirated	-p	-t	-	-k	(-?)
	voiceless aspirated					
	voiced					
Fricatives	voiceless					
Nasals	voiced		-m	-n		-ŋ
Liquids	trill					
	lateral					
Glides	voiced		-w		-j	

Dashes indicate syllable-initial consonants that do not occur in final position.

Compare this with Lao (Tai Kadai; Enfield 2007), which has a slightly different set of simpler syllable initial consonants than Khmu, shown in Table 4.6, but a very limited set of possible initial clusters by comparison (only *kw* and *k^hw* occur). Syllable final consonants are shown in Table 4.7.

In Sgaw Karen (Tibeto Burman), the large set of options in initial position looks similar to the situation in Khmu, with some differences in the details. For example, Sgaw Karen has voiced fricatives. Note also the gaps in the Sgaw Karen system: no voiceless aspirated palatal stop, no voiced velar stop. See Table 4.8. Further, an array of syllable initial clusters is possible, and as Table 4.9 shows, Sgaw Karen allows a larger range of syllable initial clusters than Khmu. On the other hand, the Sgaw Karen system contrasts sharply from the other languages shown so far in its severely limited possibilities for consonants in syllable final position. Only one consonant occurs in syllable final position, namely the glottal stop /ʔ/.⁵ This is shown in Table 4.10.

⁵ Joe Pittayaporn (pers. comm.) reports that the stop is often thought of as a feature of tone, meaning that there are in fact zero final consonant phonemes in this language.

Table 4.8 *Syllable-initial consonants in Sgaw Karen (TB)*

		Bilabial	Alveolar	Palatal	Velar	Glottal
Stops	voiceless unaspirated	p	t	c	k	?
	voiceless aspirated	ph	th		kh	
Fricatives	voiced	b	d	j		
	voiceless		s(θ)	sh	x	h
Sonorants	voiced		z(j)		χ	
	semivowel	w(v)				
	nasal	m	n	ŋ	ŋ	
	lateral			l		
	trill			r		

Reproduced from Ratanakul 2001: 128, Table 1.

Table 4.9 *Syllable-initial consonant clusters in Sgaw Karen (TB)*

	-w-	-l-	-y-	-r-	-j-
p-		pl-	py-	pr-	pj-
ph-		phl-	phy-	phr-	phj-
b-	bw-	bl-	by-	br-	
m-		ml-	my-		mj-
d-	dw-				
c-	cw-		cy-		
n-	nw-				
k-	kw-	kl-	ky-	kr-	
kh-	khw-	khl-	khy-	khr-	
s-	sw-				sj-
sh-	shw-		shy-	shr-	
l-	lw-				
?-	?w-				

Adapted from Ratanakul 2001: 128, Table 2.

Moklen syllable initial phonemes resemble what we have seen already, but note the presence of voiced velar plosive g, and the absence of fricatives, shown in Table 4.11. Moklen syllable final consonants are half in number of the initials. This is shown in Table 4.12. It should also be noted that the Moklen system is exceptional for an Austronesian language. This is another case, alongside Cham, of seemingly radical restructuring of a phonological system as a possible result of sustained language contact. Moklen and related dialects have long been in intensive contact with MSEA languages (Pittayaporn 2005).

Table 4.10 *Syllable-final consonants in Sgaw Karen (TB)*

		Bilabial	Alveolar	Palatal	Velar	Glottal
Stops	voiceless unaspirated					-?
	voiceless aspirated					
	voiced					
Fricatives	voiceless					
	voiced					
Sonorants	semivowel					
	nasal					
	lateral					
	trill					

Dashes indicate positions where syllable-initial consonants that do not occur in final position.

Reproduced from Ratanakul 2001: 128, Table 1.

Table 4.11 *Set of consonants possible in the initial position of main syllables in Moklen*

	Labial	Alveolar	Palatal	Velar	Laryngeal
Voiceless unaspirated stops	p-	t-	c-	k-	?
Voiceless aspirated stops-	p ^h -	t ^h -	c ^h -	k ^h -	
Voiced stops-	b-	d-		g-	
Spirants-					h-
Nasal sonorants-	m-	n-	j-	ŋ-	
Oral sonorants-	w-	l-	j-		

After Larish 2005: 516.

Table 4.12 *Set of consonants possible in the final position of main syllables in Moklen*

	Labial	Alveolar	Palatal	Velar	Laryngeal
Voiced stops					
Voiceless aspirated stops					
Voiceless unaspirated stops	-p	-t		-k	-?
Spirants					-h
Vasal sonorants	-m	-n		-ŋ	
Oral sonorants	-w		-j		

After Larish 2005: 517.

Table 4.13 *Initials of Dànnánshān Miao, Guizhou (HM)*

		Labial	Alveolar	Retroflex	Palatal	Velar	Uvular
Stops	voiceless unaspirated	p-	t-		c-	k-	q-
	voiceless aspirated	p ^h -	t ^h -		c ^h -	k ^h -	q ^h -
Prenasal stops	voiceless unaspirated	m-p-	n-t-		ŋ-c-	ŋ-k-	ŋ-q-
	voiceless aspirated	m-p ^h -	n-t ^h -		ŋ-c ^h -	ŋ-k ^h -	ŋ-q ^h -
Affricates	voiceless unaspirated		ts-	tʂ-	tc-		
	voiceless aspirated		ts ^h -	tʂ ^h -	tc ^h -		
Prenasal affricates	voiceless unaspirated		n-ts-	n-tʂ-	n-tc-		
	voiceless aspirated		n-ts ^h -	n-tʂ ^h -	n-tc ^h -		
Fricatives	voiceless	f-	s-	ʂ-	ç-		
	voiced	v-		ʐ-	z-		
Nasals	voiceless	m-	n-		j-		
	voiced	m-	n-		j-		ŋ-
Liquids	lateral, voiceless		l-				
	lateral, voiced		l-				
Glides	voiced		w-				

After Ramsey 1987: 282.

Further to the extreme of asymmetry between what is possible in syllable initial and syllable final position, in Dànnánshān Miao (Hmong Mien) a very high number of initial consonant contrasts, in combination with a tone system that has eight tonal distinctions, corresponds with a very low number of vowels and finals (from Ramsey 1987: 282). Table 4.13 shows the initials of Dànnánshān Miao, which is among the largest consonant inventories in MSEA. Its large size is due to the presence of distinct prenasalized series of stops and affricates, as well as further elaboration in the system, including voicing contrasts not only in fricatives but in nasals and liquids, and consonants at retroflex and uvular places of articulation.

In contrast to its extravagant set of consonants in initial position, the only permissible consonant finals are n and ŋ, and these are not contrastive. Thus, the only possible syllable rimes that end in consonants in Dànnánshān Miao are [en], [aŋ], and [oŋ]. This might suggest that the only final is an unspecified nasal, whose place of articulation is determined by vowel quality.

4.2 Vowels

Languages of MSEA have vowel inventories that are among the largest in the world. Khmer has eighteen steady state vowels and twelve diphthongs, as shown in Table 4.14.

Table 4.14 *Cambodian (AA) vowel phonemes*

	Front	Central	Back
High	i i:	i i:	u u:
Mid	e e:	ə ə:	o o:
Low	ɛ:	a a:	ɔ ɔ: ɑ:

A thirty-vowel system (including diphthongs) Diphthongs: iə, iø, eə, ei, əw, ae, aə, aj, a:o, ou, ue, œ

Not all MSEA vowel inventories are large. In the northern part of MSEA and beyond, Sinitic and Hmong Mien languages have vowel inventories that are modest by MSEA standards, but around average in world terms. For example, Dànnánshān Miao (Hmong Mien, Guizhou) has five simple vowels, as laid out in Table 4.15.

Table 4.15 *Dànnánshān Miao (HM) vowel phonemes*

	Front	Central	Back
High	i		u
Mid	e		o
Low		a	

A five-vowel system.

After Ramsey 1987: 282.

Among Sinitic languages, those spoken further north, such as Mandarin, tend to have smaller vowel inventories, while further south, for example in Cantonese and Shanghainese, closer to the MSEA area, larger inventories are found. Tables 4.16 and 4.17 illustrate the Cantonese and Shanghainese systems, respectively.

Table 4.16 *Cantonese (SN) vowel phonemes*

	Front	Central	Back
High	i	yu	u
Mid	e	eu	o
Low		a aa	

An eight-vowel system.

Reproduced from Matthews and Yip 1994: 17, Table 1.2.

While these two examples from Sinitic show numerous distinctions in vowel quality, they do not show systematic distinctions in length. Only one length contrast is shown in tables 4.16 and 4.17: the distinction between /a/ and /aa/ in Cantonese.

Table 4.17 *Shanghainese (SN) vowel phonemes*

	Front	Central	Back
High	i y	ɛ	u
Mid	ø		o
Low	ɛ	a	ɔ

A nine-vowel system: [ɪ ʏ ə ʊ] also occur as allophones of /i y ε u/.

Reproduced from Zee and Xu 2003: 133.

A similar situation is observed in Tibeto Burman languages spoken closer to the core MSEA area.⁶ For example, the Sgaw Karen vowel phoneme space shows nine distinctions in vowel quality, with no contrastive length distinction. See Table 4.18.

Table 4.18 *Sgaw Karen (TB) vowel phonemes*

Levels of tongue	Position of lips	Parts of tongue		
		front	central	back
		unrounded	rounded	
high		i	ɪ	u
mid		e	ə	o
low		ɛ	a	ɔ

A nine-vowel system.

Reproduced from Ratanakul 2001: 129, Table 3.

Other language families of MSEA tend towards yet larger inventories, with nine or more contrasts in simple vowel quality, with the addition of vowel length contrasts and complex vowels (diphthongs). Many languages feature unrounded high vowels in central or back position.

The vowel phoneme inventory of (Eastern) Khmu, shown in Table 4.19, features ten distinct vowel qualities, all but one of which split into two due to a short/long distinction. In addition to these steady state vowels, Khmu has three diphthongs, all of which move from a high vowel to a schwa: iø, uø, aø.

⁶ For a comprehensive survey of phonological inventories of Tibeto-Burman languages, see Namkung (1996).

Table 4.19 (*Eastern*) *Khmu (AA)* vowel phonemes

	unrounded			rounded
	Front	Central	Back	
High	i i:		i i:	u u:
Mid	e e:	ə ə: ʌ:		o o:
Low	ɛ ɛ:	a a:		ɔ ɔ:

A 22-vowel system (including diphthongs).

After Premsrirat 1987.

The situation in Lao is almost identical to that in Khmu, with one fewer distinction in the steady state vowels, and with three lowering diphthongs /ia, ua, ua/ (and a rising diphthong /au/ in Northern varieties). See Table 4.20.

Table 4.20 *Lao (TK)* vowel phonemes

	unrounded			rounded
	Front	Central	Back	
High	i i:		u u:	u u:
Mid	e e:	ə ə: ʌ:		o o:
Low	ɛ ɛ:	a a:		ɔ ɔ:

A 23-vowel system (including diphthongs).

After Enfield 2007: 35.

Further illustrating the high degree of convergence in the basic shape and size of vowel systems in core MSEA, Table 4.21 shows the vowel system of a language from yet another language family: Modern Western Cham (Headley 1998: 25ff.; cf. Thurgood 1998). This Austronesian language has had intensive contact with Austroasiatic languages including Khmer.

Table 4.21 *Modern Western Cham (AN)* vowel phonemes

	Front	Central	Back
High	i i:	i i:	u u:
Mid	e	ə ə:	o o:
Low	ɛ ɛ:	a a:	ɔ ɔ:

A 23-vowel system (including diphthongs) Diphthongs: ia, ii, ea, au, ua, oa
Registers: high, low

After Headley 1998: 25ff.

A similar situation is found in Moken, which has the vowel system shown in Table 4.22 (Pittayaporn 2005: 190).

Table 4.22 *Moken (AN) vowels*

	Front	Central	Back
High	i i:	ə	u u:
Mid	e e:	a a:	o o:
Low	ɛ ɛ:		ɔ ɔ:

A 17-vowel system (including diphthongs). Diphthongs: iə, uə.
After Pittayaporn 2005: 190.

To put the Modern Western Cham and Moken vowel inventories into the perspective of their own language family, they have nearly four times as many distinct vowels than would typically be found in an Austronesian language (outside the Chamic and Moken/Moklen families). Compare the vowel inventory of Bahasa Malaysia, shown in Table 4.23, with six steady state vowels and no length distinctions.

Table 4.23 *Bahasa Malaysia (AN) vowels*

	Front	Central	Back
High	i		u
Mid	e	ə	o
Low		a	

A 6-vowel system. After Clynes and Deterding 2011: 263.

The Modern Western Cham vowel inventory looks less like other languages of its family and more like the languages spoken in its immediate surroundings. This kind of situation is often explained in terms of linguistic areal effects. The emergence of a system of just the kind found in neighbouring languages would seem too unlikely to have occurred by independent internal development, given that the same development has not occurred in other Austronesian languages.⁷ As we saw in Chapter 1, Cham has been spoken in the mainland of Southeast Asia for at least 2,000 years (Sidwell 2006).

Similarly, Moklen vowels show an almost identical structure to those in Modern Western Cham (see Table 4.24). They are equally aberrant from the point of view of Austronesian more broadly.

⁷ Marc Brunelle (pers. comm.) reports that Austroasiatic vowels appear to have been introduced into Chamic through loan words from Austroasiatic languages.

Table 4.24 *Moklen (AN) vowel phonemes*

	Front	Central	Back
High	i i:	u u:	u u:
Mid	e e:	ə ə:	o o:
Low	ɛ ɛ:	a a:	ɔ ɔ:

A 26-vowel system (including diphthongs). Diphthongs: ia, uə, uə, εo, aw, aj, əj, ui.

From Larish 2005: 518, after Swastham 1982: 29.

4.3 Phonotactics

Phonotactics in MSEA languages can best be thought of in terms of an /initial + rime/ structure, where the consonant phonemes that can appear in syllable final position constitute a highly limited set, as we saw in the last section. This restriction on distinctions among finals can in part be interpreted as a neutralization of phonemic contrasts in syllable final position. For example, in the case of Lao (see tables 4.6 and 4.7 above) there are three distinctions in voicing of stops in syllable initial position, but these distinctions are always neutralized in syllable final position. While initial position allows /b/, /p/, and /ph/ in Lao, final position only allows a checked/unreleased bilabial stop [p]. Also as shown in the last section, languages of the area differ as to whether they allow consonant clusters in initial position. Austroasiatic languages mostly allow complex clusters, while languages of other families mostly do not. This has consequences for the analysis of these languages' phonotactic templates.

Many languages have an initial unstressed ‘minor syllable’, which can draw only from a reduced set of possible consonant and vowel phonemes. Figure 4.1 shows the segmental structure of Kri (see Hayes 1984 on closely related Thavung).

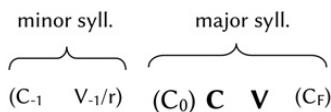


Figure 4.1 Segmental structure of the Kri (AA) word
After Enfield and Diffloth 2009: 10. Non-segmental features are not specified here.

While Kri has a large number of vowel phonemes (see below), the vowels that may occur in minor syllable (in the $V_{.1}$ slot in Figure 4.1) constitute a radically reduced system, with the following properties:

- only three vowels occur: a , i , u (along with r)
- there is no vowel length distinction (only short/unstressed vowels occur)
- there is no register distinction independent from that of the major syllable

While all Kri words have major syllables, minor syllables are relatively infrequent: around a third of all words have a minor syllable. Of those minor syllables that have a vowel, two thirds have a , while the other two are significantly less frequent (one in five have u , only one in fifteen have i . See Enfield and Diffloth 2009).

With reference to these minor major syllable structures, researchers of the sound structure of words in MSEA languages often refer to the idea of *sesquisyllables*, and the property of *sesquisyllabicity*. This terminology was introduced by Matisoff (1973b) to refer to the ‘one and a half syllable’ form of words found in many MSEA languages (see Henderson 1952, Shorto 1960, Thomas 1992, Butler 2015, Pittayaporn 2015).

The term *sesquisyllable* has not always been applied in an exact or consistent way.⁸ In a narrow sense, it can refer to a syllable with schwa epenthesis between elements of an initial consonant cluster; that is, a syllable whose onset is phonologically /CC/ but phonetically [C³C]. In a broad sense, it can refer to any word that has an iambic structure, with the main stressed syllable coming at the end. Consider the following four words in Kri (Enfield and Diffloth 2009):

- (4.1) /cakaŋ/ [caka:ŋ]
 ‘to measure something by handspans’
 Kri | AA | Enfield and Diffloth 2009: 46
- (4.2) /ckaŋ/ [c³ka:ŋ]
 ‘a hand span’
 Kri | AA | Enfield and Diffloth 2009: 46
- (4.3) /caŋ/ [ca:ŋ]
 ‘buttress of a tree’
 Kri | AA | Enfield Kri word list
- (4.4) /kaŋ?/ [ka:ŋ?]
 ‘chin/jaw’
 Kri | AA | Enfield Kri word list

In the broad sense, both (4.1) /cakaŋ/ [caka:ŋ] and (4.2) /ckaŋ/ [c³ka:ŋ] are sesqu syllabic, while in the narrow sense, only (4.2) /ckaŋ/ [c³ka:ŋ] is. Recent work by Butler

⁸ Thomas (1992) argues that there are four types of sesquisyllable: (1) where a fully predictable schwa is inserted in an initial cluster, (2) where an inserted schwa contrasts with the same CC sequence within the language, (3) with a minor syllable that can only contain one of two or three vowel phomemes, and (4) with a minor syllable that allows ‘nearly full vowel contrasts’.

(2015) and Pittayaporn (2015) has made a significant advance by insisting that we be consistent and precise in the use of the term sesquisyllable. By turning to empirical and theoretical accounts in order to offer motivated solutions, this work is making the intuitive idea of sesquisyllabicity accountable to current theory and data in theoretical phonology and articulatory phonetics. Butler (2015) calls for more thoughtful consideration of the terms and seeks to make progress by holding certain phonological ideas of syllable structure accountable to phonetic behaviour that can be experimentally tested. Pittayaporn (2015) takes a broader comparative approach to the problem, offering a typology of sesquisyllabic languages, defining the distinct meanings that this term can have.

Sino Tibetan and Hmong Mien languages show an especially marked asymmetry between initials and rime, allowing large sets of consonants in syllable initial position and few if any consonants in syllable final position. Cantonese is at the liberal end of the Sinitic spectrum, with 20 consonant phonemes possible in syllable initial position, and six possible in final position (/ m, n, ɳ, p, t, k/). Dànnánshān Miao (Guizhou, Ramsey 1987: 282) as noted above has many more possible initials, but only three possible consonant final rimes: / en/, / aŋ/, and / oŋ/. Lahu has no final consonants at all. As Matisoff describes it, Lahu has ‘very simple monosyllables, no initial consonant clusters and no final consonants, and no contrast in vowel length’ (Matisoff 2000b: 88).

4.4 Register

Many languages of the Austroasiatic family in MSEA feature a register system in their phonology. In a register system, words in the language occur in one of two, or occasionally three or even four, phonological registers. Each register is defined by a cluster of phonetic properties, associated with a core distinction between types of phonation defined by different articulatory states of the glottis such as breathy versus clear/modal (Henderson 1952, Gregerson 1976, Smith 1979: 32).

Kuay, a language of the Katuic sub branch of Austroasiatic, is described as having two registers:

Register 1 or tense register: ‘clear voice quality, fewer short vowel phonemes, some allophonic vowel height, strong aspiration, voiceless unaspirated initials, a larger consonant inventory, and higher pitch’

(Preecha 1993: 245)

Register 2 or lax register: ‘breathy voice, more short vowel phonemes, no allophonic vowel height, weak aspiration, voiced initial stops, a smaller consonant inventory, and lower pitch’

(Preecha 1993: 245)

The register of the syllable has an effect on the pronunciation of the syllable's segments. When a vowel is pronounced with different phonation it tends to be affected in quality by the phonation difference (see examples below). Similarly, the realization of some stop phonemes can be affected by phonation register, where markedly distinct allophones are conditioned by the register of the syllable. (4.5) and (4.6) show the phoneme /t/ in syllable initial position, realized as [t] with first register, and as [d] or [d̥] with second register:

- (4.5) /to:m/ [to:m]

'to wrap'

Kuay | AA | Preecha 1993: 248

- (4.6) /ti:/ [d̥i:] or [d̥i̥:]

'middle'

Kuay | AA | Preecha 1993: 248

Phonotactic constraints often apply in relation to register, in similar ways to what is described in the next section in relation to tone. For example, in Kuay, /s/ and /h/ occur in Register 1 only.

Theraphan (1989) describes a similar situation in the closely related language Kui (also a Katuic language). Kui has two registers, also labelled R1 (having 'clear voice and high pitch') and R2 (having 'breathy voice and low pitch'). Thongkum notes that R2 vowels have lower F0 than R1 vowels (Thongkum 1989: 8), pointing to an oft noted association between register and pitch. Further, in Kui, both vowel quality and vowel length are affected by whether the syllable in question is R1 or R2.

Kri (Vietic sub branch of Austroasiatic) has two registers, referred to as heavy and light (Enfield and Diffloth 2009). Table 4.25 gives some minimal pairs, showing that the same underlying vowel phonemes are realized phonetically in quite distinct ways. The 18 phonetically distinct long vowel nuclei in Kri can be grouped into nine pairs, where the members of each pair contrast phonologically in their register and phonetically in terms of their quality. See Table 4.26.

Table 4.25 *Some Kri (AA) (near) minimal pairs contrasting in register alone*

Heavy register	Light register
/ciin? _H / [ci:n ²] 'nine'	/ciin? _L / [c ^e i:n ²] 'cooked'
/kara:i _H / [kar ^a :i:y] 'sand'	/kara:i _L / [kara:i:y] 'sunshine'
/mata:m _H / [mat ^a :a:m] 'son-in-law'	/kata:m _L / [kata:m] 'crab'
/kme:? _H / [kme: ^e ?] 'thumb'	/kme:? _L / [km ^e e: ²] 'female (chicken)'
/ro: ^e ? _H / [z ^e ɔ: ² ?] 'clear'	/ro: ^e ? _L / [z ^e ɑ: ² ?] 'pig basket'
/ru: ^e ? _H / [z ^e u: ² ?] 'fence'	/ro: ^e ? _L / [z ^e o: ² ?] 'to know'

After Enfield and Diffloth 2009: 30.

Table 4.26 *Long vowel phonemes of Kri (AA) showing vowels paired by register*

	Front	Back
High		
Mid		
Low		

After Enfield and Diffloth 2009: 29.

Table 4.27 *Long vowel phonemes of Kri (AA) laid out in a standardized nine-place system*

		Front	Central	Back
High	heavy	i:	i:	u:
	light	e: ɪ:	ə: ɪ:	eo: ʊ:
Mid	heavy	ɪ: e:	ə: ɔ:	ʊ: ɔ:
	light	ɛ: ε:	ɔ: a:	ɔ: a:
Low	heavy	ɛ: ε:	ɔ: a:	ɔ: a:

Each vowel phoneme has two allophones, conditioned by register.

After Enfield and Diffloth 2009: 31.

Emerging from the array in Table 4.26 is an orderly pattern of nine heavy light pairs, giving eighteen vowels in all. The pattern is shown in Table 4.27.

If the distinct phonetic vowels are not viewed as register conditioned allophones of underlying vowel phonemes, then the analyst may posit an especially large number of vowel phonemes. In the case of Kri, a liberal count, including all register conditioned allophones, long and short vowels, and the three diphthongs, would give thirty three vowels. A conservative count would treat these as all derivable from an underlying nine vowel system.

Sedang, a language of the Bahnaric branch of Austroasiatic, has two registers, labelled tense and lax, with the kinds of properties described for Kuay, Kui, and Kri.

Also in Sedang, the distinction between registers is not evenly distributed across the vowel space. The register difference is neutralized in some kinds of syllables, namely those with final stops and final /h/ (Gregerson and Smith 1973: 160, Smith 1979: 32).

In a similar system in Rengao, also of the Bahnaric branch of Austroasiatic, we again see how vowel phonemes can be realized differently depending on register. Table 4.28 shows a selection of long vowel phonemes in Rengao and how they are differently realized in the two registers tense and lax (from Gregerson and Smith 1973: 147).

Table 4.28 *Selection of Rengao (AA) long vowel phonemes and their realization in tense versus lax register*

Vowel phonemes	with lax register	with tense register
/i/	[i]	[e ^t]
/e/	[e]	[ɛ]
/u/	[u]	[o ^u]
/o/	[o]	[ɔ]

Reproduced from Gregerson and Smith 1973: 147.

This effect of phonation register on the quality of vowels can naturally lead to a split in a language's vowel phoneme inventory. If a phonation difference co occurs with a difference in vowel quality, then the role of a phonation difference as a contrastive signal can, in principle, be fully transferred to the associated vowel difference. At a first stage, the phonation is what matters for signalling lexical contrast in the language, and the vowel difference is a phonetic by product. At a subsequent stage, what was once a phonetic by product now becomes a phonological distinction and becomes a key signal of lexical contrast, leaving the original phonation differences to weaken or disappear. If a phonation difference disappears from the language, it then leaves the mutation of vowels to take up the contrastive function. This phonologization of a vowel quality difference leads to a split in the vowel system, doubling the number of vowel phonemes. To illustrate this idea, Table 4.29 is a hypothetical scenario based on the Rengao data shown in Table 4.28. It shows a possible future state of the language, with eight distinct vowel phonemes from what were previously four.

If register were lost, the four vowel phonemes shown in Table 4.28 could become eight; this would happen if (1) the phonation correlates of the two registers disappeared, and (2) the phonetic differences between vowels in each register – which were once a by product of the phonation difference – now take over the functional load in signalling contrast in the lexicon.

Table 4.29 *Hypothetical scenario based on the Rengao situation shown in Table 4.28*

Rengao vowel phonemes: hypothetical future scenario	
/i/	
/ɛ̇/	
/e/	
/ɛ/	
/u/	
/o ^u /	
/o/	
/ɔ/	

Table 4.30 *Vowel phonemes of Cambodian (AA)*

i	i:	ɪ	ɪ:	u	u:
e	e:	ə	ə:	o	o:
		ɛ̇:			ɔ̄:
		ɛ̄:		ɔ̄:	
a	a:			ɑ̄	ɑ̄:

Diphthongs: ie ae ao aə ea iə uə oa wə əy əi.

Adapted from Headley 1998: 24.

This exact scenario has taken place in MSEA languages, including languages of the Khmeric and Katuic families and Haroi (see Thurgood 2003b and references therein). An example is Khmer. Khmer has an unusually large number of vowel phonemes in its phonological system, as shown in Table 4.30 (based on Headley 1998: 24, in contrast to Table 4.14 above; see Huffman 1976).

Headley (1998: 23, after Ferlus 1992) explains this proliferation of vowels with reference to changes that have taken place since the time of Old Khmer, a language that had a two register system of the kind described above for Kuay, Kui, Sedang, and Rengao. The historical correspondences are summarized in Table 4.31.

In Table 4.31, the left column shows the Old Khmer vowel nucleus phonemes, which were pronounced differently depending on which register the syllable occurred with (as described for Rengao, above). The middle column shows the modern vowels that preserve aspects of their pronunciation with the old high register (clear phonation, voiceless initial consonant) and the right column shows the modern vowels that preserve aspects of their pronunciation with the old low register (breathy phonation, voiced initial

Table 4.31 *The relation between modern Khmer vowels (shown in Table 4.30) and the earlier two-register system in Old Khmer*

Old Khmer vowel nucleus phoneme	Modern Khmer vowel corresponding to Old Khmer vowel in high register	Modern Khmer vowel corresponding to Old Khmer vowel in low register
<i>two phonetic realizations, depending on whether syllable was high or low register</i>	<i>erstwhile voiceless initial in clear phonation syllable</i>	<i>erstwhile voiced initial in breathy phonation syllable</i>
a'	u/wə	a
a	ɔ:	a:
ā'	ea/oa	a
ā	ie	a:
i	i	ə
ī	i:	əy
ī	i	ə
ī	i:	əi
u	u	o
ū	u:	ɔ:
œ	ə:	aə
e	ɛ:	ɛ:
ē	ɛ:	ae
o	ɔ:	ao

Reproduced from Headley 1998: 24.

consonant). The old phonation differences are not preserved in Modern Khmer, but the two versions of each vowel are still kept distinct. Today, the phonetic features that distinguish these vowels are now purely matters of vowel quality (position in the vowel space, as well as status as steady state or moving vowel).

While most systems feature a two way contrast in registers, up to four contrasts are attested. The register system in Chong (from the Pearic sub branch of Austroasiatic; see Thongkum 1988, 1991) is particularly complex, with a four way contrast that reveals not only phonation type differences but changes in phonation type during the time course of the syllable's articulation. In an instrumental study of register in Takhian Thong Chong, Christian DiCanio writes:

The register contrast in Takhian Thong Chong includes a modal register, a modal tense register, a breathy modal register, and a breathy tense register. The hyphen here indicates that there is a movement from one phonation type to another over the vowel's duration. Thus, the breathy tense register consists of breathy voice following the release of the onset consonant, with a change in voice quality towards more tense or pressed phonation.

(DiCanio 2007: 461)

Table 4.32 *Takhian Thong Chong registers*

Word	Register	Gloss
lɔɔŋ	modal	'stride'
ceet	modal	'to sharpen wood'
tɔŋ	modal	'house'
pʰat	modal	'tail'
lɔɔn	modal-tense	'navel'
cɔɔt	modal-tense	'deer'
paj	modal-tense	'palm'
ceɔk	modal-tense	'pig'
raaj	breathy-modal	'ten'
paat	breathy-modal	'peel'
pət	breathy-modal	'to fan'
paaj	breathy-tense	'two'
cɔɔŋ	breathy-tense	'Chong'
kɔtaɔk	breathy-tense	'bean'
tuj	breathy-tense	'squash'
pʰst	breathy-tense	'rattan'

Reproduced from DiCanio 2007: 462.

Table 4.32 shows some examples of words in the four registers.

DiCanio's instrumental analysis of Chong reveals a system in which the phonation type contrasts that define phonological registers are associated with distinctions in fundamental frequency (pitch), as well as harmonics. DiCanio shows that pitch and harmonics are necessary elements in distinguishing between registers in this language. At the same time, as noted in the next section in the synchronic description of tone systems in Vietnamese and White Hmong, we know that pitch contrasts that define phonological tones are associated with distinctions in phonation type. Note the near equivalence of these descriptions.

4.5 Tone

An oft cited feature of MSEA languages is that many of them are tone languages. When asked what this means, most linguists would agree with Yip (2002: 1): 'A language is a "tone language" if the pitch of the word can change the meaning of the word.' But as linguists of MSEA languages since Henderson (1952, 1965, 1967) have insisted, it is wrong to think that pitch is the sole or defining feature of a tone system in MSEA (see Brunelle and Kirby 2015 and Sidwell 2015; see also Abramson and Thongkum 2009).⁹

⁹ Hyman's definition of tone is more carefully worded with reference to pitch: 'A language with tone is one in which an indication of pitch enters into the lexical realisation of at least some morphemes.' (Hyman 2001: 1368; 2006: 229).

It is important to recognize that pitch is frequently only one of the phonetic components of “tone” as a phonological category . . . A phonological tone is in our area very frequently a complex of other features besides pitch such as intensity, duration, voice quality, final glottal constriction and so on.

(Henderson 1967: 171).

From this perspective, while tone and phonation register are sometimes considered to be distinct phonological phenomena, they should be treated as instances of a single overarching type of system insofar as they each involve the use of suprasegmental laryngeal features for lexical contrast. Pitch contours, distinctions in phonation, and other glottalic effects are all produced in the larynx, by the vocal folds, and are all articulatorily independent of segmental speech sounds produced with the lips, teeth, and tongue (i.e., consonants). Tone and phonation are intimately bound, and not essentially distinct. For this reason, we recognize that the sound system of an MSEA ‘tone’ language, such as Vietnamese, is not of a different species from that of a classical MSEA ‘register’ or ‘phonation type’ language such as Kri (Enfield and Diffloth 2009). Most systems that are identified as one or the other (in phonological terms) actually display properties of both (in phonetic terms). While many specialists do not intend the term ‘tone’ to refer exclusively to pitch, this is how most linguists understand the term. The term *tonation* allows us to capture this (see Brunelle and Kirby 2016, Kirby and Brunelle 2017, Brunelle *et al.* 2018). Nevertheless, for convenience I will continue to use the terms *tone* and *register* to refer to the two prototype systems that are identified within this possibility space.

A small tone inventory is observed in Hakha Lai, a Kuki Chin (Tibeto Burman) language spoken in Myanmar and India. Hakha Lai has three tones: rising, falling, and low (level) (Hyman and VanBik 2002). This is shown in Figure 4.2.¹⁰

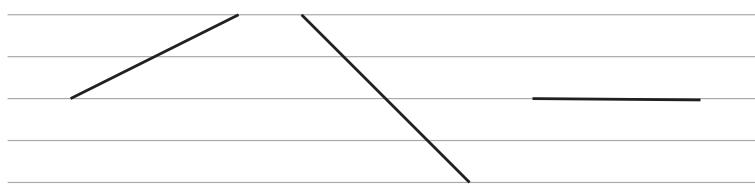


Figure 4.2 *Hakha Lai (TB) tones*
After Hyman and VanBik 2002: 2.

¹⁰ Figure 4.2 and subsequent figures in this section use a grid to depict tones. The faint horizontal lines in the background represent pitch levels. These lines can be labelled with the numbers 1 to 5, from low to high, following Chao (1930). Each bolded line represents a distinct tone in the language.

Hakha Lai words are mostly monosyllabic. Short stopped syllables (i.e., syllables with short vowels and checked finals) all have rising tone. Syllables with a long vowel and stop coda all have low tone. Reduced (i.e., monomoraic) syllables are toneless. Table 4.33 gives some examples.¹¹

Table 4.33 *Illustration of falling, rising and low tones on smooth syllables as realized after ka ‘my’ in Hakha Lai (TB)*

CVV	CVD	CVVD			
ka hmaà	‘my wound’	ka lùn	‘my heart’	ka tlaàŋ	‘my mountain’
ka zuù	‘my beer’	ka làw	‘my field’	ka raàl	‘my enemy’
ka keé	‘my leg’	ka hróm	‘my throat’	ka koóy	‘my friend’
ka ?oó	‘my voice’	ka tsál	‘my forehead’	ka tsáán	‘my time’
ka saa	‘my animal’	ka raŋ	‘my horse’	ka koom	‘my corn’
ka hnii	‘my skirt’	ka kal	‘my kidney’	ka boor	‘my bunch’

After Hyman and VanBik 2002: 2.

Another relatively simple example comes from Modern Standard Chinese. Figure 4.3 shows the tones: Tone 1 (high level, or 44 using the Chao tone number system), Tone 2 (high rising or 35), Tone 3 (low rising or 13), Tone 4 (high falling or 51).

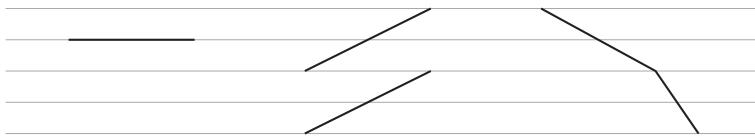


Figure 4.3 *Modern Standard Chinese (SN) tones*

In Mandarin Chinese, all words are specified as having one or another of these four tones. Thus, four different words have the same sequence of segments *ma*, differentiated only by their tone:

(4.7) Mandarin: minimal set of tone contrasts

mā [ma⁴⁴] ‘mother’

má [ma³⁵] ‘hemp’

mǎ [ma¹³] ‘horse’

mà [ma⁵¹] ‘scold’

Mandarin | SN

¹¹ The examples of tones in Table 4.33 are all presented after the toneless pronoun *ka* ‘my’. A rule in Hakha Lai converts an underlying rising tone to falling in phrase-initial position, and so if they were presented as stand-alone words, their underlying tones would not be observed. (See section 4.7 on tone sandhi for further discussion of Hakha Lai tones.)



Figure 4.4 *Cantonese (SN) tones*

Asterisks mark those tones that can also occur on checked syllables.

When we compare this to the situation in Cantonese, we see a couple of significant differences (Figure 4.4). First, the set of tones is larger, with six distinct pitch contours. Second, three of those contours (marked with asterisks in Figure 4.4) occur with both checked and unchecked syllables, while the other three tones only occur with unchecked syllables. (The three tones with checked syllables are sometimes regarded as pairs of distinct tones – checked and unchecked – leading to the oft heard claim that Cantonese has nine tones.) In Mandarin, the phonotactic rules are such that there are no checked syllables in the language.

- (4.8) Cantonese: minimal set of tone contrasts (Matthews and Yip 1994: 21)

yāu [jau⁵⁵] ‘worry’

yáu [jau³⁵] ‘paint’ (n.)

yau [jau³³] ‘thin’

yàuh [jau²¹] ‘oil’

yáuh [jau²³] ‘have’

yauh [jau²²] ‘again’

Cantonese | SN | Matthews and Yip 1994: 21

Northern Vietnamese, like Cantonese, has six tones, as shown in Figure 4.5 (the labels for the tones in Northern Vietnamese are, from left to right, upper before lower, *ngang*, *sắc*, *hỏi*, *huyền*, *ngã*, *nặng*):

- (4.9) Northern Vietnamese: minimal set of tone contrasts

ma [ma:³²] ‘ghost’

má [ma:³⁵] ‘mother’

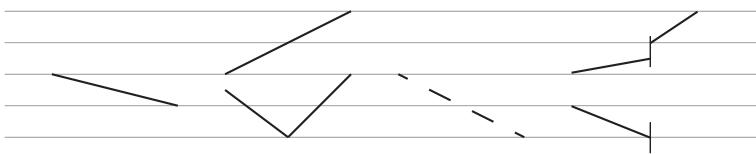
mả [ma:²¹³] ‘tomb’

mà [ma:³¹] ‘which’

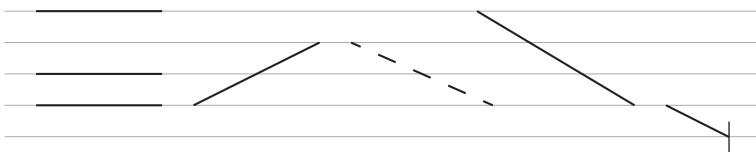
mᾶ [ma?²a³⁵] ‘horse’

ma [ma?²¹] ‘rice seedling’

Vietnamese | AA

Figure 4.5 *Northern Vietnamese (AA) tones*

The dashed line indicates breathy voice¹² and the vertical line indicates glottalization.

Figure 4.6 *White Hmong (HM) tones*

After Ratliff 1992.

The dashed line indicates breathy voice and the vertical line indicates glottalization.

As Figure 4.5 shows, some of the Northern Vietnamese tones are associated with non pitch laryngeal effects. One low falling tone can have breathy phonation, another features a glottal stop final with associated shortening of the rime, and a high rising tone features a glottal stop part way through the rime.

In White Hmong, similarly, the tone system features non pitch elements relating to phonation. Figure 4.6 illustrates the seven tones of White Hmong.

Ratliff (1992: 1) supplies the following minimal set for these seven tones (the orthography used by Ratliff features final consonants as tone markers – these are available in the orthography as there are so few consonants possible in final position in the language):¹³

(4.10) White Hmong: minimal set of tone contrasts

- tob* [to⁵⁵] ‘deep’
- to* [to³³] ‘to pierce’
- tos* [to22] ‘to wait’
- tov* [to24] ‘to add water’
- tog* [tɔ̄ 42] ‘to sink’
- toj* [to⁵²] ‘hill’
- tom* [to²¹] ‘there’

White Hmong | HM | Ratliff 1992: 1

¹² Marc Brunelle (pers. comm.) remarks: ‘Breathy phonation in *huyέn* is occasionally reported, but seems limited to citation forms’.

¹³ The *-m* tone has a syntactically conditioned variant *-d* (213), although *-d* can operate independently (Ratliff 1992: 120 2).

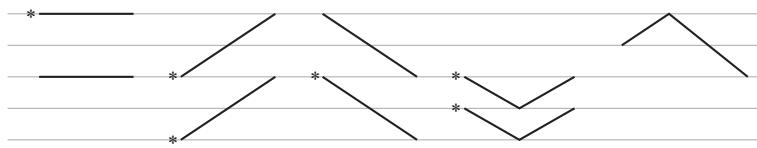


Figure 4.7 *Dong (TK) tones*

After Long and Zheng 1998: 31 2.

Asterisks mark those tones that can also occur on checked syllables.

Tones become largest in number in the Tai and Kam Sui speaking areas towards the south of China. The ‘record number’ according to Ramsey (1987: 244) is fifteen tones in a single system (nine on open syllables, six on checked), in Dong (corner of Hunan, Guizhou, and Guangxi: Long and Zheng 1998: 31 2), shown in Figure 4.7.

Minimal sets from Dong (Kam) are shown in (4.11).

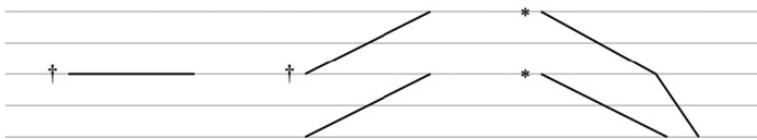
- | | |
|--|---------------------------------------|
| (4.11) <i>sau⁵⁵</i> ‘twist’ | <i>jk⁵⁵</i> ‘wet’ |
| <i>sau³⁵</i> ‘straw’ | <i>jk³⁵</i> ‘hard working’ |
| <i>sau²¹²</i> ‘rear’ (v.) | <i>jk²¹</i> ‘pitiful’ |
| <i>sau³²³</i> ‘steam’ (v.) | <i>jak³²³</i> ‘hungry’ |
| <i>sau¹³</i> ‘grass carp’ | <i>jak¹³</i> ‘fish pond’ |
| <i>sau³¹</i> ‘husband’ | <i>jak³¹</i> ‘rust’ |
| <i>sau⁵³</i> ‘soup’ | |
| <i>sau⁴⁵³</i> ‘egret’ | |
| <i>sau³³</i> ‘create’ | |

Dong | TK | Long and Zheng 1998: 31 2

This is claimed to be a fifteen tone system, even though the six ‘checked tones’ are not different in pitch contour from six of the non checked tones. While tradition in Chinese linguistics treats and labels tones that occur on checked rimes separately from those that occur on open rimes (as noted for Cantonese, above), a more parsimonious description of this system would treat the presence of a checked final consonant (p, t, k, or ?) as a segmental addition without relevance to the tone. That said, even if we interpret this as a nine tone system, Dong appears to have the largest number of tone distinctions in the MSEA area.

So far in this section we have treated tone systems as inventories. That is, we have listed the tones that a language has, implying that any syllable can have any tone. But there are often constraints in a language that lend texture to a tonal system. We can illustrate with the case of Lao.

The Lao set of tones resembles that observed in Mandarin, with a fifth tone in addition (a low falling tone). But as Figure 4.8 shows, there are limitations on the kinds of

Figure 4.8 *Lao (TK) tones.*

† also appears on [V_{short}+stop] syllable

* also appears on [V_{long}+stop] syllable

syllables that the tones may occur with. In Lao (unlike in Mandarin), some syllables have a stop consonant, such as p, t, or k, in final position. If there is no such stop in final position, a syllable can carry any of the five tones shown in Figure 4.8. Two example segmental sequences are /saw/ and /kha:/, each of which may have five distinct meanings depending on which of the five tones occurs, as shown in (4.12) and (4.13):

- (4.12) /saw/ occurs with all five tones in Lao:

saw1 /saw³³/ ‘rent’
saw2 /saw³⁵/ ‘stop’
saw3 /saw¹³/ ‘post’
saw4 /saw⁵¹/ ‘morning’
saw5 /saw³¹/ ‘sad’

Lao | TK

- (4.13) /kha:/ occurs with all five tones in Lao

khaa1 /kha:³³/ ‘galangal’
khaa2 /kha:³⁵/ ‘stuck’
khaa3 /kha:¹³/ ‘leg’
khaa4 /kha:⁵¹/ ‘trade’
khaa5 /kha:³¹/ ‘kill’

Lao | TK

But if the syllable’s segmental sequence features a stop in final position, not all tones may be realized. For example, with the two syllables /het/ and /khop/ the only tones that occur are tones 1 and 2:

- (4.14) /het/ occurs with only two tones in Lao:

hêt1 /het⁴⁴/ ‘do’
hêt2 /het³⁵/ ‘mushroom’
/het¹³/, /het⁵¹/, /het³¹/ are not possible words
Lao | TK

- (4.15) /khop/ occurs with only two tones in Lao:

khop1 /khop⁴⁴/ ‘meet’

khop2 /khop³⁵/ ‘bite’

/khop¹³/, /khop⁵¹/, /khop³¹/ are not possible words

Lao | TK

When the vowel is long, different restrictions apply. Examples (4.16) and (4.17) show /saap/ and /kha:t/, which may only occur with tones 4 and 5:

- (4.16) /sa:p/ occurs with only two tones in Lao:

saap4 /sa:p⁵¹/ ‘to know’

saap5 /sa:p³¹/ ‘dank’

/sa:p⁴⁴/, /sa:p³⁵/, /sa:p¹³/ are not possible words in Lao

Lao | TK

- (4.17) /kha:t/ occurs with only two tones in Lao:

khaat4 /kha:t⁵¹/ ‘to expect’

khaat5 /kha:t³¹/ ‘to be torn, or lacking’

/kha:t⁴⁴/, /kha:t³⁵/, /kha:t¹³/ are not possible words in Lao

Lao | TK

There are yet further gaps on the occurrence of certain tones with syllables of certain segmental sequences in Lao. These have to do with the initial consonant that occurs, in combination with the presence or absence of a stop in final position. Three principles (reflecting historical accident) capture the possible tones in Lao, defined by a combination of (1) class of syllable initial consonant and (2) presence or absence of a stop in syllable final position:

Principle 1. If C1 is a non aspirated stop (b , d , p , t , c , k , ?) or j (versus any other consonant initial) and C2 is not a stop, then Tones 2 and 5 do not occur (cf. *paa1* /pa:⁴⁴/ ‘forest’ versus *paa3* /pa:¹³/ ‘fish’ versus *paa4* /pa:⁵¹/ ‘elder aunt (Pa.eZ)’) (/pa:³⁵/, /pa:³¹/ are not possible words in Lao).

Principle 2. If C2 is a stop and the vowel is short, then tones 3, 4, and 5 do not occur (cf. *khop1* /khop⁴⁴/ ‘meet’ versus *khop2* /khop³⁵/ ‘bite’); in addition, if C1 is a non aspirated stop or j , Tone 1 does not occur (cf. *kop2* /kop³⁵/ ‘frog sp.’).

Principle 3. If C2 is a stop and the vowel is long, tones 1, 2, and 3 do not occur (cf. *saap4* /sa:p⁵¹/ ‘to know’ versus *saap5* /sa:p³¹/ ‘dank’); in addition, if C1 is a non aspirated stop or j , Tone 4 does not occur (cf. *daap5* /da:p³¹/ ‘sword’).

These gaps may appear puzzling in their complexity and apparent arbitrariness. As is often the case in language, a puzzling synchronic situation makes sense when it is interpreted and explained with reference to the diachronic processes that provided the causal/conditional context for its development. Recall from section 2.3.2 that historical comparative analysis of tone systems in Southwestern Tai languages makes reference to four parameters which historically determine the tone category of a word (Gedney 1989).

In the Lao case, first, the initial consonant of a syllable will belong to one of three classes, defined not by how it is pronounced in the modern language, but by how the corresponding segment was pronounced in Proto Tai. For example, the syllable initial voiceless aspirated stop /kh/ in Lao is a modern exponent of two distinct segments in Proto Tai: (1) a voiceless aspirated stop /kh/ and (2) a voiced stop /g/. Table 4.34 shows the three initial consonant classes, defined by their value in the proto language.

Table 4.34 *The three classes of syllable-initial consonant in Lao*

Class	Definition
High	Consonants in the modern language whose correspondents in the proto language have the property ‘voiceless, plus friction’
Middle	Consonants in the modern language whose correspondents in the proto language have the property ‘voiceless, minus friction’ ¹⁴
Low	Consonants in the modern language whose correspondents in the proto language have the property ‘voiced’

Second, a syllable will be either ‘live’ (i.e., with vocalic or sonorant final) or ‘dead’ (i.e., with stop final).

Third, if a syllable is ‘live’, it makes a difference whether the syllable is historically unmarked for tone (A) or historically marked as Tone 1 (B) or Tone 2 (C).

Fourth, if ‘dead’ (D), it makes a difference whether a syllable has a long or a short vowel.

This set of possibilities allows us to chart a historical comparative map of the Lao tone system, as we can do for any Tai language, using as coordinates (1) the proto initial consonant classes as rows, (2) the two syllable statuses (live versus dead) as column rubrics, (3) the three live syllable tone categories (A, B, C) as distinct columns, and (4) the two dead syllable tone categories as distinct columns. See Table 4.35.

¹⁴ Note that ‘voiceless, minus friction’ includes obstruents that were phonetically voiced, whether they are written as Li’s ‘voiced preglottalized’ ?b ?d or as implosives. Thanks to David Solnit for pointing this out.

Table 4.35 *Modern Lao (TK) examples of reflexes of traditional Tai tones*

Proto initial	Syllable status				
	'live'			'dead'	
	A	B	C	D _{SHORT}	D _{LONG}
voiceless, +friction	<i>khaa3</i> 'leg'	<i>khaa1</i> 'galangal'	<i>khaa5</i> 'slave'	<i>khat2</i> 'contrary'	<i>khaat5</i> 'torn'
	<i>kaa3</i> 'crow'	<i>kaa1</i> 'fish sp.'	<i>kaa4</i> 'dare'	<i>kat2</i> 'bite'	<i>kaat5</i> 'cabbage'
	<i>khaa2</i> 'stuck'	<i>khaa1</i> 'cost'	<i>khaa4</i> 'commerce'	<i>khat1</i> 'select'	<i>khaat4</i> 'hope'

By way of summary of the last two sections, an impression of the degree to which the phonetic features of pitch, phonation type, vowel quality, checked/open syllable ending, vowel length, and quality of initial consonants are intertwined should now be possible. Under a unified rubric of tonation, Austroasiatic languages tend towards the register type. ‘The Austroasiatic languages of Southeast Asia (including Khasic, but excluding Vietnamese) share a number of phonological features, such as sesquisyllabic words (also present in Gta? [which belongs to the Munda subgroup]), large vowel inventories (often including large numbers of diphthongs), and to a lesser extent register distinctions’ (Jenny *et al.* 2015: 137). By contrast, most non Austroasiatic languages in MSEA tend towards the tone type that is, where four or more distinctions in pitch and associated laryngeal features of a word can change the meaning of the word.

Perhaps the most important evidence in favour of viewing tone and register as mere focal points under a unified rubric of tonation comes from the processes of historical change that lead to the emergence and development of tonation systems. These processes are the topic of the next section.

4.6 Tonogenesis

Comparative linguists in the early twentieth century found Vietnamese to be a challenge for classification of languages into distinct families. Evidence for genealogical relations between two languages usually comes from comparison of the forms of basic vocabulary. Many Vietnamese words are similar to Chinese words, even a good number that refer to everyday objects. See Table 4.36.

Early analysts also saw that Vietnamese basic vocabulary items had similarities with Austroasiatic languages, suggesting that Vietnamese belonged in that family, but the presence of tones in Vietnamese was felt to be a challenge. As Gage (1987: 497) puts it, there is a ‘great typological gulf’ between the phonological systems of Vietnamese and other Austroasiatic languages (especially with respect to tone and monosyllabism).

Table 4.36 Vietnamese (AA) and Chinese (SN) lexicon

	Early Sino-Vietnamese	Modern Chinese
‘paper’	giáy	zhǐ
‘character’	chữ	zì
‘wife’	vợ	fù
‘husband’	chồng	zhàng
‘hat’	mũ	mào
‘garden’	vườn	yuán
‘gold’	vàng	huáng
‘to smelt’	rèn	liànl
‘knife’	dao	dāo
‘small cup’	chén	zhǎn

Data are from Alves 2016: 281–9.

In a seminal article on Vietnamese, Maspéro (1912) supposed that a language – and by extension, a language family – either had tone or did not. He assumed that a language could not acquire tone and so concluded against a relationship precisely because Vietnamese had tones and the other Austroasiatic languages did not. He supposed that Vietnamese tones could be inherited from Proto Tai (if Vietnamese were a Tai language), but he could not see how tones could arise *de novo* (if Vietnamese were an Austroasiatic language). While Maspéro was wrong to think that tone could only be acquired from an ancestor language, his work was important because he drew attention to the need to address tones in the historical comparative linguistics of the area.

Picking up on Maspéro four decades later, Haudricourt (1954) made a breakthrough by arguing that tones could be acquired other than by inheritance from an ancestor language, through internal processes of language innovation and change. On this assumption, he concluded that tones *could* have arisen in an Austroasiatic language, and that this happened in Vietnamese.

The historical investigation of Sinitic languages has shown that while they have all acquired tone from a common ancestor, new tones and new tone systems evolved and developed since earlier forms of the language. Norman (1988) summarizes the findings of historical linguists based on the historical documents known as Qieyun or Rhyming Tables (Karlgren 1915, Pulleyblank 1970, 1971, 1984). Analysis of these sources reveals that Middle Chinese, spoken around 600 CE, had four tonal categories: level, rising, departing, and entering. The entering category was composed of the syllables with stop finals.¹⁵

¹⁵ In our discussion of modern tone systems, above, we saw that in descriptions of some MSEA languages, especially Sinitic languages, tones on checked-final syllables are often treated as distinct from tones on open-final syllables, even if the pitch contours of the syllables are identical (see the cases of Cantonese and Dong in section 3.4).

The other three categories were distinguished from each other by their different pitch contours, that is, by their tones. As Norman (1988) notes, it is possible to reconstruct the historical correspondences between these early tones and their exponents in modern Sinitic languages, but it is not possible to know the actual phonetic values of the early tones. In other words, we know that the three tones were different from each other in pitch, but we do not know what they sounded like.

From this state of affairs, each of the tone categories split into two registers.¹⁶ This split was associated with the distinction, found in all tone categories, between voiced and voiceless consonants in initial position. The voicing distinction was associated with a difference in pitch of the syllable, as a phonetic by product of the initial voicing. At the same time that the initial voicing distinctions were lost, the associated pitch difference compensated by picking up the functional load of contrast, and thereby becoming phonemicised, splitting the tone system in two, and creating eight tones. The resulting structure shown in Table 4.37 is still attested in some varieties of Chinese, such as the Yue languages (Cantonese being an example).

Most Chinese languages have now merged some of the cells in Table 4.37, collapsing the distinctions represented in the table. Table 4.35, above, shows the results of essentially the same scenario in historical Tai. Despite the wide variety of modern tone systems in Sinitic languages, synchronic descriptions of their tones will often be mapped onto this underlying plan.

Table 4.37 *Eight traditional Chinese tone categories*

Initial tone class in ancestor language	Tone category			
	<i>píng</i> ‘level’	<i>shǎng</i> ‘rising’	<i>qù</i> ‘departing’	<i>rù</i> ‘entering’
<i>yīn</i> ‘upper’ (from old voiceless-initial syllables)	1. <i>yīnpíng</i> ‘upper level’	3. <i>yīnshǎng</i> ‘upper rising’	5. <i>yīnqù</i> ‘upper departing’	7. <i>yīnrù</i> ‘upper entering’
<i>yáng</i> ‘lower’ (from old voiced-initial syllables)	2. <i>yángpíng</i> ‘lower level’	4. <i>yángshǎng</i> ‘lower rising’	6. <i>yángqù</i> ‘lower departing’	8. <i>yángrù</i> ‘lower entering’

The categories are shown as a function of (1) four-way contrast in ‘tones’ in the ancestor language (columns) and (2) two-way contrast in ‘register’ arising from loss of voicing distinction in initial consonants (rows).

After Norman 1988: 54.

¹⁶ The word ‘register’ is used here in a different sense from that of phonation-type register, described in section 3.4. Here, it refers to the ‘high’ and ‘low’ versions of a tone that result from a historical split.

Let us return now to Haudricourt's investigation of the historical status of Vietnamese in relation to the mostly non tonal Austroasiatic languages that surround it. In his follow up on Maspéro's earlier work, Haudricourt argued that Vietnamese could be plotted in the same way as proposed for Chinese (Norman 1988: 55; see also Alves 2016). This is shown in Table 4.38.

Table 4.38 *Traditional Vietnamese tone categories*

Register	Tone category			
	A	B	C	D
'upper' (from old voiceless-initial syllables)	1. ngang	3. sá̄c	5. hó̄i	7. sá̄c
'lower' (from old voiced-initial syllables)	2. huyé̄n	4. nang	6. ngã̄	8. nang

The terms for Vietnamese tones are the ones used in modern Vietnamese – each label models the relevant tone.

Reproduced from Norman 1988: 55.

Table 4.38 suggests the same scenario for Vietnamese as that described for Sinitic languages. In this scenario, the ancestor language already has a tone system in place. Haudricourt's puzzle was to explain how that tone system emerged in the first place. His proposal, now widely accepted in spirit (cf. Thurgood 2002, for a reappraisal, and Alves 2016),¹⁷ is as follows: Category D corresponds to words that had final stop consonants (cf. the 'entering' category in Sinitic); category C corresponds to words which historically ended in s or h; words of category B were argued to derive from words with glottal finals; and remaining words were in the A category. Let us review step by step how this process of tonogenesis occurred (following Haudricourt 1954, Matisoff 1973b, Thurgood 2002; see also Alves 2016). (In what follows, I group categories B and D together under a single category B.)

At Stage 1 of the process, the language has no tones. Three distinct types of syllable can be identified, differing from each other with respect to how the syllable terminates: (A) with a vowel or nasal consonant in final position, (B) with a (glottal or oral) stop in final position, or (C) with a voiceless fricative in final position. See Table 4.39.

At this stage, the three types of syllables contrast in terms of their segmental make up. However, they also contrast in terms of their pitch, in an incidental way. The different syllable endings each have a distinct phonetic effect on the pitch of the syllable. The

¹⁷ Brunelle (pers. comm.) writes that the Thurgood reappraisal is pre-empted by Haudricourt's (1965) nailing down of the voice quality element.

Table 4.39 *Stage 1 of tonogenesis, pre-tonal*

A	B	C
CV-ø/C _{nasal}	CV-ʔ/p/t/k	CV-h/s

presence of the final stop in the (B) type syllables tends to result in a rising pitch, while the presence of the final voiceless fricative in the (C) type syllables tends to result in a falling pitch. At this stage, these are naturally caused tendencies in the realization of pitch, and they may hardly be noticeable. They are incidentally associated with phonetic features of the finals, but they do not function to signal lexical contrast.

In a second stage, the pitch contours that were only incidental now become phonologized. This is associated with the loss of the finals.¹⁸ The lexical distinctions remain. The original non contrastive pitch differences now take up the load of lexical contrast. This stage is illustrated in Table 4.40.

Table 4.40 *Stage 2 of tonogenesis, pitch contours being phonologized*

A	B	C
CV _{tone A}	CV _{tone B}	CV _{tone C}

At this same stage, we can explicitly distinguish syllable types in terms of the voicing of their initial segment, which again has associated regular non contrastive pitch differences, as noted already in the above accounts of historical tone splits in Sinitic and Tai languages. This is captured in Table 4.41.

Table 4.41 *Stage 2 of tonogenesis, showing syllable types distinguished by voicing of syllable-initial segment*

A	B	C
C _{voiceless} V _{tone A}	C _{voiceless} V _{tone B}	C _{voiceless} V _{tone C}
C _{voiced} V _{tone A}	C _{voiced} V _{tone B}	C _{voiced} V _{tone C}

¹⁸ David Solnit (pers. comm.) points out that syllables in the A column did not lose their final nasals. However, for convenience of presentation here, I leave out reference to those final nasals.

In the next stage of the process, shown in Table 4.42, the initial contrasts in voicing are lost (all initial voiced obstruents become voiceless), and the regular pitch differences associated with erstwhile voicing values also phonologize, now taking on the load of maintaining lexical contrast.

Table 4.42 *Stage 3 of tonogenesis, initial contrasts, lost (here, all Cs are voiceless) incidental pitch differences arising from previous ±voicing now make lexical contrast*

A	B	C
C-V _{tone A1}	C-V _{tone B1}	C-V _{tone C1}
C-V _{tone A2}	C-V _{tone B2}	C-V _{tone C2}

This hypothesis of a three stage process of tonogenesis for Vietnamese is supported by comparative data. ‘In comparing Vietnamese words with their Austroasiatic cognates, Haudricourt observed that Vietnamese words of category C (*hỏi ngã*) frequently correspond to cognate words ending in *h* or *s* in other languages’ (Norman 1988: 55). Table 4.43 shows some comparisons with Mon and Mnong.

Table 4.43 *Comparison of Vietnamese tone category C words with Mon, Mnong, and Kri cognates (all AA)*

	Vietnamese	Mon	Mnong	Kri
‘seven’	báy	tpah	poh	pajh
‘nose’	mũi	muh	mühl	muujh
‘root’	rẽ	rɔh	ries	leerh

The table shows the modern Vietnamese tones *hỏi* and *ngã* corresponding to words with final -s or -h in other Austroasiatic languages.

Reproduced from Norman 1988: 55 (Kri data added by Enfield).

The data from Khmer in Table 4.44 show further evidence of the same correspondence. In relation to category B (*sắc* and *nặng* tones), Haudricourt saw a correlation with words in other Austroasiatic languages that end in a glottal stop. Table 4.45 shows some correspondences with Khmu and Riang. And in relation to the Category A tones proposed for early Vietnamese, Gage (1987) supplies the examples in Table 4.46, showing relevant correspondences with Khmu.

Table 4.44 *Comparison of Vietnamese tone category C words with Khmer cognates (both AA)*

	Vietnamese	Khmer
‘scratch, draw’	ké	kes
‘toss (away)’	bó	poh
‘root’	rẽ	rs
‘hole’	lõ	luh

The table shows the modern Vietnamese tones *hói* and *ngã* corresponding to words with final -s or -h in other Austroasiatic languages. (Note also Kri *leerh* ‘root’.)

Reproduced from Gage 1987: 511.

Table 4.45 *Comparison of Vietnamese tone category B words with Khmu, Riang, and Kri cognates (all AA)*

	Vietnamese	Khmu	Riang	Kri
‘leaf’	lá	hla?	la?	sulaa?
‘rice’	gao	rənko?	ko?	-
‘fish’	cá	ka?	-	?akaa?
‘dog’	chó	so?	so?	coo?
‘louse’	chí	-	si?	ci?
‘endure’/‘hurt’	chịu	cu?	-	-

The table shows the modern Vietnamese tones *sá* and *nă* corresponding to words with final -? in other Austroasiatic languages. After Norman 1988: 56 and Gage 1987: 512 (Kri data added by Enfield).

Table 4.46 *Comparison of Vietnamese tone category A words with Khmu cognates (both AA)*

	Vietnamese	Khmu
‘bamboo’	tre	talaa (Northern Khmu)
‘thou’	mày	mee (2SG.M)

After Gage 1987: 512.

These findings can be summarized for Vietnamese using a similar template as that shown for Sinitic, above. This is shown in Table 4.47, from Thurgood (2002).

This account is now broadly applied to the historical linguistics of tone in MSEA, and can explain the origin of the original tones of Proto Chinese. As Norman (1988: 56) notes, Chinese likely derived its original four tone system (see top row of Table 4.37) by

Table 4.47 *The six Vietnamese tones, plotted as a function of two non-tonal parameters of syllables in the proto-language*

Finals	Open, nasal	Stopped	Voiceless fricatives
Initials			
proto-voiceless (high pitch)	*pa > pa 'ngang'	*pak > pák 'sác'	*pas > pà 'hòi'
proto-voiced (low pitch)	*ba > bà 'huyền'	*bak > pák 'nặng'	*bas > pã 'ngã'

Columns represent erstwhile distinctions in syllable-final position; rows represent erstwhile distinctions in syllable-initial position.

Reproduced from Thurgood 2002: 335, Figure 1.

means of similar syllable final elements as Vietnamese: level tone from the class of words with sonorant or zero final, rising tone from the class of words with glottal stop finals, departing tone from the class of words with final *s*, and entering tone from the class of words with stop finals (Norman 1988: 57). This set the initial conditions for the scenario described above in Table 4.37, when loss of initial voicing distinctions created a split of these four tones into two registers each. Ratliff applies this logic in her analysis of the development of tones in Hmong Mien languages, as shown in Table 4.48.

Table 4.48 *Haudricourt's model of tonogenesis applied to White Hmong*

Stage 1	Atonal	CV	CV?	CVh
Stage 2	Initial tonogenesis	CV	CV~	CV~
Stage 3	Pitch differences based on initial	CV voiceless- initial	CV~ voiceless- initial	CV~ voiceless- initial
Stage 4	White Hmong tones	CV voiced-initial [tɔ ⁵⁵] 'deep' [tɔ ⁵³] 'hill'	CV~ voiced-initial [tɔ ²⁴] 'mix' [tɔ ²²] 'wait'	CV~ voiced-initial [tɔ ³³] 'pierced' [tɔ ³¹] 'sink'

Adapted from Ratliff 2015: 249.

The logic of tonogenesis is laid bare in the historical tone boxes described in Chapter 2 (see Gedney 1972 on historical Tai). Table 4.49 shows tone boxes for Karen (TB) (Manson 2011: 4), with proto tones as columns, and proto initials as rows. Each box is a distinct modern tone category.

In support of these proposals, we can observe languages spoken today that contain precisely the starting conditions for the proposed historical phonological processes. Let us return to the case of Kri. As we have seen above, Kri is not a tone language, but it has a system of two registers that divide all syllables in the lexicon into one of two classes: heavy and light. Further, there are contrasts in what we call terminance (Enfield and Diffloth

Table 4.49 ‘Karen tone box’

	A	B	B'	C
Proto-aspirated	1 (III) water [<i>*h^ht</i>]	4 (VI) star leaf [<i>*p^han</i>] flower [<i>*p^hɔ</i>]	7 (Va) bone child right [<i>*la</i>] fingernail fire [<i>*mi</i>] spicy [<i>*m^he</i>] take [<i>*p^he</i>] give [<i>*k^ha</i>] bitter [<i>*p^hi</i>]	10 (VIII) sky iron pig skin/bark shoot [V] [<i>*k^hri</i>] [<i>*p^ho</i>] [<i>*p^hwe</i>] [<i>*hɛ</i>] [<i>*p^hi</i>] [<i>*p^hi</i>] [<i>*k^hu?</i>] [<i>*h^he?</i>][<i>*k^hu?</i>]
	branch [<i>*p^hɔ</i>]			
	chicken [<i>*s^han</i>]			
	sleep [<i>*m^hi</i>]			
	die [<i>*p^hi</i>]			
	2 (II)	5 (VIIa) silver [<i>*t^hɔn</i>] ginger [<i>*p^hɛŋ</i>]	8 (V) paddy blow/howl [<i>*ti</i>] egg cheek [<i>*pu</i>] liver [<i>*sm</i>] eat [<i>*p^ham</i>] left [<i>*se</i>] be at, exist [<i>*ta</i>]	11 (VIIa) alcohol wing heart call/shout near [<i>*pi</i>] [<i>*hɛ</i>] [<i>*k^ho</i>] [<i>*su</i>] [<i>*sa</i>] [<i>*ta</i>]
	rabbit [<i>*te</i>]			
	navel [<i>*te</i>]			
	spear [<i>*pan</i>]			
	white [<i>*p^hva</i>]			
Proto-voiced	3 (I) nest [<i>*bwe</i>]	6 (IV) sun stone [<i>*loŋ</i>]		12 (VII) monkey eye/face brain intestines rib deep [<i>*zo?</i>] [<i>*me?</i>] [<i>*no?</i>] [<i>*bre?</i>] [<i>*rr?</i>] [<i>*jɔ?</i>]
	tongue [<i>*ble</i>]			
	person [<i>*bra</i>]			
	name [<i>*min</i>]			
	drunk [<i>*mun</i>]			
	red [<i>*le</i>]			
	old [humans] hot [<i>*go</i>]			

After Manson 2011: 4; roman numeral tone labels are from Luce 1985

Table 4.50 *Three values for terminance in Kri (AA)*

Voiced	Checked	Voiceless
t ^o :j ‘tail’	t ^o :j ² ‘bowl’	t ^o :j ‘follow’
z ^o u: ‘turtle’	z ^o u: ² ‘pig basket’	z ^o u: ‘dry’
cama:l ‘shiny’	?uma:l ² ‘to hunt’	dal ‘to bounce’
kavər ‘stir’	kav ² ar ² ‘embrace’	tar ² ‘to run out of workspace’
care:w ‘green’	sare:w ² ‘to raise/feed’	ble:w ‘four-eyed turtle’
j ^a :m ‘sugar cane’	j ^a :m ² ‘to cry’	
ka:n ‘oversize’	ka:n ² ‘to hunt by night’	
b ^u :j ¹ ‘dust’	p ^u :j ² ‘tree sp. (<i>licuala</i> sp.)’	
caŋ ‘tree sp.’	caŋ ² ‘salty’	

After Enfield and Diffloth 2009: 16.

2009), by which all syllables can be divided, on another plane, into three classes: voiced, checked, and voiceless.¹⁹ Table 4.50 shows minimal sets for the terminance distinctions. The intersection of register (with two values) and terminance (with three values) yields six distinct syllable types in Kri, as shown in Table 4.51. Now see that this non-tonal system has precisely the same underlying logic as the fully fledged tonal system of Vietnamese, which is in the same subgroup of Austroasiatic as Kri. Table 4.52 is a simpler version of Table 4.47, above.

Table 4.51 *The six Kri (AA) syllable types, defined by intersections of terminance and register distinctions*

Terminance	voiced	checked	voiceless
Register			
light	LIGHT VOICED	LIGHT CHECKED	LIGHT VOICELESS
heavy	HEAVY VOICED	HEAVY CHECKED	HEAVY VOICELESS

After Enfield and Diffloth 2009: 39.

Table 4.52 *The six Vietnamese tones, plotted as a function of two non-tonal parameters of syllables in the proto-language*

Finals	voiced	checked	voiceless
Initials			
proto-voiceless	‘ngang’	‘sáć’	‘hỏi’
proto-voiced	‘huyền’	‘nặng’	‘ngã’

Reproduced from Thurgood 2002: 335, Figure 1.

¹⁹ Note that checked finals are also voiceless, technically speaking. Here, and in the table on Vietnamese below, ‘voiceless’ excludes checked.

Phonological systems that may be shifting from non tonal to tonal in nature can show synchronic dependencies (or at least, correlations) between certain pitch contours and certain other elements of a syllable. In Moklen (Austronesian), tones ‘can be predicted by looking at syllable type and consonant voicing’ (for details see Larish 1997: 139–40, drawing on Swastham 1982: 51–2 and Court 1971: 87). In general, syllables with unchecked finals (including those with nasals) have falling pitch patterns, with height of tone conditioned by voicing of the main syllable initial consonant and the presence or absence of a final nasal (Larish 1997). The associations (which are tendencies, not strict correlations) are summarized in Table 4.53.

Table 4.53 *Associations between syllable type and pitch in Moklen (AN)*

Pitch pattern	Syllable conditions	Examples
45	unreleased final -p, -t, k, -?	pa:t45 ‘four’ tu:k45 ‘to boil’
453	main-syllable-initial consonant unvoiced <i>or</i> final -h	te.han453 ‘post, mast’ pe.c'an453 ‘banana’ ba21.leh45 ‘to return’ bəh45 ‘do, make’
343	main-syllable-initial consonant voiced <i>and</i> minor-syllable-initial consonant unvoiced	c'bə.ba:y343 ‘curry’ ka.ba:ŋ343 ‘boat’
232	main-syllable-initial consonant voiced <i>and</i> minor-syllable-initial consonant voiced	ba.yəy232 ‘shirt’ da.ga:232 ‘basket’
35	major syllable checked <i>and</i> main-syllable-initial consonant approximant <i>and</i> vowel long <i>and</i> minor-syllable-initial consonant unvoiced	ka.ya:235 ‘crocodile’ pa.la:235 ‘wooden platform for grilling fish’
24	major syllable checked <i>and</i> main-syllable-initial consonant approximant <i>and</i> vowel long <i>and</i> minor-syllable-initial consonant voiced	da.la:k24 ‘blood’ ba.la:24 ‘hot charcoal’

After Larish 1997.

The ‘phonetically predictable’ nature of pitch patterns in Moklen, as partially outlined here with these data from Larish (1997), strongly resemble the kinds of conditions described for tonogenesis in languages such as Lao, mentioned above.

Table 4.54 *Moklen (AN) pitch-based minimal pairs*

Pair	Meanings
wa.ləy232	'eight'
wa.ləy453	'day'
buu.la:232	'dehusked rice'
buu.la:453	'to yell at (someone)'
kə.la:t21	'mushroom'
kə.la:t45	'hot' (temperature)
du.lu:k22	'kind of fish'
du.lu:k44	'to shine' (flashlight)

After Larish 1997: 140.

While the process appears not yet to have resulted in a fully fledged tone system, the existence of pitch based minimal pairs demonstrates that the process is under way. See Table 4.54.

Let us summarize the upshot of the complex set of issues covered in the above sections, which have cycled in a fugue like fashion through vowel distinctions, pitch distinctions, phonation distinctions, and the historical relations among these. Tone is about more than pitch. Phonological register is about more than phonation type. Vowel systems are about more than points in the vowel space. Syllables are about more than series of segments. These things are always interconnected in languages, but they are especially closely intertwined in the languages of MSEA, as we have seen in our discussion of the historical phonology of the area, as well as in the synchronic phonetics and phonology of Austroasiatic languages in particular.

We have concentrated on a well known account of tonogenesis in MSEA languages. It is worth emphasizing that this account does not represent the only path that tonogenesis might follow (see Homber et al. 1979: 49, Svantesson 1989, Hyslop 2009, Matisoff 1999: 86). Current research on the onset of tonogenesis suggests other possibilities. For example, in the Phnom Penh dialect of Khmer, what appears to be the onset of tonogenesis is observed in just a fragment of the language's phonological system. Words that begin with clusters consisting of voiceless unaspirated stops followed by [r] are being realized with voiceless aspirated stop initials, no [r], breathy voice on the vowel, and a low rising pitch on the rime. For example, [kru:] 'teacher' becomes [kʰy:¹³] (Guion and Wayland 2004, Kirby 2014a, 2014b).

To finish this section, I want to reiterate the point that the ideas of 'tone languages' and 'register languages' are idealizations. Each idea captures something real, but in some cases at least they are capturing different parts of a single, more general reality. The truth is that in every tone system, pitch distinctions are intertwined with other

laryngeal features, and conversely in every register system, phonation distinctions are intertwined with other laryngeal features. Whether we are considering register languages like Chong and Kri or tone languages like Vietnamese and Hmong, the systems feature an exhaustive classification of syllables in the language into a small number of categories, each category being defined by a bundle of laryngeal phonetic properties and their association with the articulation of prevocalic and postvocalic segments.

4.7 Tone Sandhi

In some grammatical contexts, a word can be pronounced in a different tone from the one that is lexically assigned to it (i.e., the tone that will be used when the word is given in citation form). This change of tone depending on combinatorial context is called tone sandhi, borrowing from the Sanskrit grammar term meaning ‘joining’, referring to the mutational effects of fusion in morphological derivations such as affixation and compounding.

The term *tone sandhi* is mostly used in the linguistics of East Asian and Southeast Asian languages. It can be understood in more or less strict senses. In what we might call a strict sense, tone sandhi might be taken to refer to completely arbitrary and automatic replacement of one tone by another in defined environments of co occurrence. An example is that of Xiamen, a Sinitic language of China, in which the word *hai* ‘ocean’ is lexically specified for a high falling tone (contour 53) but changes to a high level tone (contour 44) when used as a modifier in the phrase *hai⁴⁴ kih²⁴* ‘ocean front’ (see below). In a broader, less strict sense, tone sandhi refers to any form of alternation or change of a tone that is regularly or predictably associated with the position of a syllable relative to other syllables. An example is that of Hakha Lai, a Kuki Chin language of Myanmar and India, in which a rising tone becomes a falling tone when it occurs after a rising tone, a process which can be accounted for by an assimilation like rule against ‘jumping’ levels between the endpoint of one tone and the starting point of the subsequent tone (see below). Typologically, strict sandhi phenomena are rare, not only in global terms but also in MSEA. By contrast, sandhi phenomena in the broader sense are found in tone languages around the world (though the term *sandhi* tends not to be used outside of Sinitic and MSEA linguistics). In the linguistics of tone in African and Mesoamerican languages, for example, sandhi like tone phenomena are discussed in terms of notions such as tone alternation, coarticulation, and assimilation.

A well known and relatively simple example of tone sandhi is that of Mandarin. In Mandarin, when a word of Tone 3 precedes another word of Tone 3, then the first word is pronounced with Tone 2. Here are some examples:

nǐ ‘you’ + *hǎo* ‘good’ → *ní hǎo* ‘greetings’

Mandarin | SN

hěn ‘very’ + *jiǔ* ‘long (time)’ → *hén jiǔ* ‘very long (time)’

Mandarin | SN

shuǐ ‘water’ + *zhǔn* ‘standard’ → *shuǐ zhǔn* ‘spirit level’

Mandarin | SN

The example of Mandarin tone sandhi is simple in comparison to sandhi processes observed in other Sinitic languages. Languages of the Southern Min subgroup of Sinitic have complex systems of tone sandhi in which most if not all tones are affected in certain contexts. In the Xiamen ‘tone circle’ (Barrie 2006, Thomas 2008, Zhang *et al.* 2006), when a word with a certain tone is used in a specific morphosyntactic context, such as a defined position in a compound or phrase, it may be realized as the next tone in the circle. This is illustrated in Figure 4.9. The examples listed in Table 4.55 show how the tone circle works. These Sinitic examples of tone sandhi fit with a strict definition of the term: a purely phonological process in which the tone of a word changes as an arbitrary function of the tone of a neighbouring word.

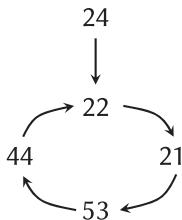


Figure 4.9 *The Xiamen (SN) tone circle*

After Chen 1987: 111.

Table 4.55 Examples of Xiamen (SN) tone sandhi following the tone circle pattern

Base form	Sandhi form		
<i>p'ang</i> ‘fragrant’ 44	<i>p'ang</i>	tsui 22	‘fragrant water’ (perfume) 53
<i>we</i> ‘shoes’ 24	<i>we</i>	tua 22	‘shoelaces’ 21
<i>pih</i> ‘ailment’ 22	<i>pih</i>	lang 21	‘ailing person’ 24
<i>ts'u</i> ‘house’ 21	<i>ts'u</i>	ting 53	‘roof top’ 53
<i>hai</i> ‘ocean’ 53	<i>hai</i>	kih 44	‘ocean front’ 24

After Chen 1987: 112.

White Hmong has a system of tonal alternation which is referred to as tone sandhi (Ratliff 1987) but it does not fit this strict definition. This is because its application is partly conditional on ‘grammatical category and the particular lexical items involved’ (Ratliff 1987: 71). In White Hmong tone sandhi, the phonetic environment of a word ‘is not sufficient to guarantee tone change’ (Ratliff 1987: 71). The observed tone sandhi phenomena take place ‘optionally’ and ‘unpredictably’ (Ratliff 1987: 73). The process is not purely phonological but is also morphological. ‘Inasmuch as tone sandhi serves to create new words in White Hmong, it constitutes one of the morphological functions of tone in the language’ (Ratliff 1987: 71). (See section 4.5 on related phenomena of ‘morphological tone’.) But while the White Hmong system is ‘arbitrary’ in its application, Ratliff argues that it is historically derived from a more regular form of motivated tone sandhi in an ancestor language: ‘compound formation by tone change’ (Ratliff 1987: 73, following Downer 1967). Compound formation by tone change ‘is a new use being made of an old, once more regular and thoroughgoing, phonetically motivated tone sandhi system’ (Ratliff 1987: 73).

White Hmong tone sandhi is of the progressive type, meaning that ‘the trigger word precedes the word that undergoes the tone change’ (Ratliff 1987: 73). The result is a ‘paradigmatic replacement of one tone by another’, but not in the circle like manner of Min, shown above. Rather, there is a neutralization and reduction of tones, from five to three (Ratliff 1987: 73). See Table 4.56.

Table 4.56 *Pattern of tone replacement in progressive tone sandhi in White Hmong*

-j	}	→ -g
-s		
-m		
-v		→ -Ø
-Ø		→ -s

1. A syllable undergoes tone replacement only when it follows a syllable with tone -b or tone -j.

2. Syllables with tones -b and -g do not undergo tone replacement.

After Ratliff 1987: 73.

Table 4.57 *Biao Min Yao (HM) tones*

Tone	Pitch value
1	44
2	21 (often breathy)
3	35
4	42?
5	24
7	54?

After Solnit 1985: 183ff.

An example from the other main branch of Hmong Mien comes from the Mienic language Biao Min Yao (Solnit 1985: 183ff.). The tones of Biao Min Yao are shown in Table 4.57.²⁰

Similarly to the case in White Hmong, tone sandhi in Biao Min Yao takes place only between syllables that represent morphemes in a relation of modifier head within a noun phrase (Solnit 1985: 184). The modifier, which occurs before the noun, undergoes sandhi according to the following patterns. First, tones 1, 3, 5, and 7 – being all of the tones that end with a relatively high pitch – become high falling (i.e., like Tone 7, but ‘slightly lengthened’ and usually without the glottal stop; Solnit 1985: 183). Here are some examples:

- (4.18) di⁴⁴ ‘cloth’ + bau³⁵ ‘grey’ → **di**⁵⁴ bau³⁵ ‘grey cloth’
 pau³⁵ ‘bake’ + cywən⁴⁴ ‘brick’ → **pau**⁵⁴ cywən⁴⁴ ‘baked brick’
 lu²⁴ ‘old’ + lwei⁴⁴ ‘clothes’ → **lu**⁵⁴ lwei⁴⁴ ‘old clothes’
 an^{54?} ‘duck’ + phe^{42?} ‘white’ → **an**⁵⁴ phe^{42?} ‘white duck’
 Biao Min Yao | HM (Solnit 1985: 183)

Tone 2 (Low falling) goes to a mid falling tone (like Tone 4, but without the breathiness):

- (4.19) nūŋ²¹ ‘ox’ + ji²¹ ‘flesh’ → **nūŋ**⁴² ji²¹ ‘beef (i.e., ox flesh)’
 Biao Min Yao | HM (Solnit 1985: 184)

The situation with Tone 4 is more complicated. If a syllable with Tone 4 has an aspirate initial (including voiceless sonorants except h), then there are no sandhi changes (or, the output of sandhi is Tone 4):

²⁰ The six tones are numbered according to their category in historical Hmong-Mien tone phonology, where four proto tones A, B, C, and D split into eight. Biao Min Yao Tone 1 A1, 2 A2, 3 B1, 4 B2/C2/D2, 5 C1, 7 D1 (Solnit 1985: 182).

- (4.20) phɔ⁴²² ‘radish’ + təi⁴²² ‘field’ → **phɔ**⁴²² təi⁴²² ‘radish field’
 Biao Min Yao | HM (Solnit 1985: 184)

If a syllable of Tone 4 has a voiced obstruent initial, then the output of tone sandhi is unpredictable. In some cases, there is no change:

- (4.21) blan⁴²² ‘spicy’ + lai⁴⁴ ‘vegetable’ → **blan**⁴²² lai⁴⁴ ‘spicy vegetable’
 Biao Min Yao | HM (Solnit 1985: 184)

In other cases the output is a high falling tone (the same output tone as that of tones 1/3/5/7):

- (4.22) gwai⁴²² ‘diligent’ + min²¹ ‘person’ → **gwai**⁵⁴ min²¹ ‘diligent person’
 Biao Min Yao | HM (Solnit 1985: 184)

See Solnit (1985) for further information about the conditions and functions of tone alternations in Biao Min Yao. For more on tone sandhi in a Hmong Mien language, see Ratliff (1992) and section 5.6.1 below.

Cases of tone sandhi in Tibeto Burman languages include Myebon Sumtu Chin (Watkins 2013a) and Hakha Lai (Hyman and VanBik 2004). Hakha Lai is a Kuki Chin language, spoken in Myanmar and Mizoram State, India. Hakha Lai words are largely monosyllabic. As noted above, the language has three tones: falling, rising, and low level. We can represent these tones in terms of their start and end points, where L low and H high. See Table 4.58.

These three tones contrast only on ‘smooth syllables’ (i.e., syllables with a vowel or sonorant final). As noted already, Hakha Lai syllables realize their underlying tone only when they are not in phrase initial position. In phrase initial position, a tonal alternation takes place: specifically, a rising tone becomes a falling tone. This means that ‘underlying falling and rising tones will be realized identically as a falling tone when occurring at the beginning of a phrase’, as shown in (4.23).

Table 4.58 *Hakha Lai (TB) tones, with values represented by start and end points*

Tone	Value	Description
1	HL	falling (represented with a grave accent)
2	LH	rising (represented with an acute accent)
3	LL	low level (unmarked)

H high, L low

- (4.23) ka hmaà ‘my wound’ → hmaà ‘wound’ (HL→HL)
 ka keé ‘my wound’ → keè ‘leg’ (LH→HL)
 Hakha Lai | TB (Hyman and VanBik 2004: 824)

Hakha Lai has three further rules of tonal alternation by which the tone of a syllable changes depending on the tone of an adjacent syllable. The sandhi rules show both progressive and regressive features. Here are the three rules (Hyman and VanBik 2004: 825):

1. Falling becomes low after non rising (i.e., falling or low: F F → F L, L F → L L)
2. Rising becomes falling after rising (R R → R F)
3. Rising becomes low before low (R L → L L)

These rules only target four of the nine logical possibilities of pairwise syllable combinations. Five of these possibilities do not undergo any tone change, as shown in Table 4.59. Table 4.60 shows the four combinations that do undergo change.

Table 4.59 *Five combinations of tone in Hakha Lai (TB) that do not undergo tone sandhi change*

	Input	Output
a	LH+HL	→ LH HL (unchanged)
b	HL+LH	→ HL LH (unchanged)
c	LL+LH	→ LL LH (unchanged)
d	HL+LL	→ HL LL (unchanged)
e	LL+LL	→ LL LL (unchanged)

Table 4.60 *The four combinations of tone in Hakha Lai (TB) that undergo tone sandhi change*

	Input	Output
f	HL+HL	→ HL LL (d in Table 4.59)
g	LH+LH	→ LH HL (a in Table 4.59)
h	LH+LL	→ LL LL (e in Table 4.59)
i	LL+HL	→ LL LL (e in Table 4.59)

A single rule can be given to account for these data. In the cases where no changes occur, ‘the first tone ends at the same level at which the second tone begins’, while in those where a change must occur, ‘the second tone begins at the opposite tone level with

which the first tone ends' (Hyman and VanBik 2004: 826). Here is the rule: 'The end tone of one syllable should be the same as the beginning tone of the next' (or: 'do not change tone levels between syllables'). Levels can change within syllables (low to high or high to low) but they cannot change between syllables. The rule captures 'a syllable contact phenomenon driven by an articulatory tendency to minimise tonal ups and downs (Hyman 1978: 261)' (Hyman and VanBik 2004: 826).

Let us consider one more example of tone sandhi, this time from Wuming Zhuang, a Tai Kadaï language spoken in southern China (Snyder and Lu 1997). Wuming Zhuang has six tones, shown in Table 4.61.

There are two basic patterns of tone alternation in Wuming Zhuang. First, Tone 5 (35) and Tone 1 (24) both have as their output Tone 3 (55). Second, Tone 2 (21) and Tone 6 (33) both have as their output Tone 4 (42). This is depicted in Table 4.62.

Table 4.61 *Wuming Zhuang*
(TK) tones

Tone	Pitch value
1	24
2	21
3	55
4	42
5	35
6	33

After Snyder and Lu 1997.

Table 4.62 *Two patterns of tone alternation in Wuming Zhuang* (TK)

Pattern 1		
Tone 5	35	→ 55 (Tone 3)
Tone 1	24	
Pattern 2		
Tone 2	21	→ 42 (Tone 4)
Tone 6	33	

After Snyder and Lu 1997.

The alternations can be understood in terms of a simple rule if the tones are analysed into component parts. Each tone in Wuming Zhuang can be analysed as having both (1) a tone register value (H versus L register in a two way split tone space) and (2) within each register, a contour value, described in terms of its beginning and end points (with four possibilities HH, LL, LH, and HL). Hence, we can represent the tones with a sequence of Ls and Hs, the first letter designating their register, and the next two designating their contour within that register. Using this system, Tone 5 (35) and Tone 1 (24) are of different registers (H versus L, respectively), but have the same contour values (both are LH). Table 4.63 gives the Wuming Zhuang tones in these re-described terms.²¹

Table 4.63 *Wuming Zhuang (TK) tones re-described in terms of register (first letter) and contour within that register*

Tone	Pitch value	Register/contour value
1	24	L.LH
2	21	L.HL
3	55	H.HH
4	42	H.HL
5	35	H.LH
6	33	H.LL

After Snyder and Lu 1997.

We can now restate the two sandhi alternation patterns as shown in Table 4.64.

Table 4.64 *Two patterns of tone alternation in Wuming Zhuang (TK) re-stated using a register/contour description of tone*

Pattern 1		
Tone 5	H.LH	→ H.HH (Tone 3)
Tone 1	L.LH	
Pattern 2		
Tone 2	L.HL	→ H.HL (Tone 4)
Tone 6	H.LL	

After Snyder and Lu 1997.

²¹ Note that the contours of tones 3 and 6 are notated in Snyder and Lu (1997) as L and H; I use LL and HH here for clarity of exposition in the present context.

The breakdown of each tone into register and start end of contour, respectively, in terms of H and L, allows a simple generalization about the output of the tone alternation in the two sandhi patterns, namely: *the output tone will have H register and the left edge of its contour will be H.*

Both tones 3 and 4 already fulfil both of these conditions, and so no tone alternation is observed (or, if you like, the output of the alternation is identical to the input). The other four tones are grouped into pairs, according to the value of the right edge of the contour: Tones 5 and 1 both have H at the right edge, while Tones 2 and 6 both have L. (See Snyder and Lu 1997 for further information.)

We have concentrated on examples of tone sandhi from Sino Tibetan and Hmong Mien languages, and a case from Tai Kadai. In descriptions of many tonal languages from the Tai and Austroasiatic families, there is little if any reference to tone sandhi. Occasionally, there is an explicit statement of the lack of sandhi, for example on Vietnamese: ‘There is no attested fully fledged tone sandhi although there is a degree of tonal coarticulation’ (Hà 2012: 8; cf. Brunelle 2009b, Brunelle *et al.* 2012; see also Brunelle *et al.* 2016 on incipient tone sandhi in Hanoi Vietnamese). This is an area for further investigation.

4.8 Intonation

The phonetic value of pitch in speech is not only recruited for distinguishing between words (in tone languages), but pitch can also be used in phonological and paralinguistic systems of intonation and prosody (Ladd 1996). While the phenomenon of lexical tone is a feature of only a subset of languages, the use of pitch in intonation and prosody is widely thought to be universal. For MSEA languages, this means that the phonetic element of pitch is being utilized for two distinct systems at the same time: tone and intonation. As Gussenhoven (2004: 45) states, ‘most tone languages will have some form of structural intonation’. In relation to the MSEA area, Matisoff (1994: 116) notes that ‘even in languages with elaborate omnisyllabic tone systems, intonation certainly exists as a phenomenon independent of tone’ (see Samarin 1952, Hyman and Monaka 2011, Brunelle *et al.* 2018).

This raises some obvious questions. Since a bit of speech cannot have two pitch values at the same time, do tone and intonation come into conflict? If they do come into conflict, which of the two systems wins out? Abramson and Svastikula (1983) posed these questions in connection with the relation between tone and intonation in Thai. A question that arises is whether the lexically fixed pitch contour of words in a tone language is affected when words are linked together in running speech. Sentential intonation determines pitch of phrases and sentences, including

phenomena such as downdrift, by which the average pitch of speech descends from beginning to end of an utterance. Through a controlled acoustic study of Thai phrases, Abramson and Svastikula concluded that when tone and intonation come into conflict in Thai, it is lexical tone that wins out. They found that the lexically assigned tones always remained ‘physically distinct’, thereby dominating the use of pitch in the system.

In Vietnamese, similarly, lexical tone distinctions are preserved in running speech, while ‘the pitch range of sentences can be shifted up or down to indicate communicative functions, without affecting patterns of contrasts between lexical tones’ (Brunelle and Kirby 2016: 199, citing Trần 1967; Đỗ *et al.* 1998, Nguyễn and Boulakia 1999, Brunelle *et al.* 2012). As Brunelle and Kirby (2016: 199–200) go on to note, similar shifts in overall range of pitch through utterances, leaving lexically specified relative tone shape intact, are described for Thai (Luksaneeyanawin 1998) and Kammu (Khmu) (Karlsson *et al.* 2007, House *et al.* 2009, Karlsson *et al.* 2010, Karlsson *et al.* 2012). For work on tone intonation interaction in Chinese, see Chao (1933), Shen (1989), Guo (1993), Cao (2002).

While pitch is generally constrained by lexically specified tone in MSEA, there is one area in the grammars of tone languages where the intonational use of pitch for pragmatic expression appears to dominate. This is in relation to sentence final particles. In Lao, sentence final particles may ‘differ from other word classes in not being specified for tone at all’ (Enfield 2007: 73): ‘Rather, they are perhaps “blank” with respect to lexically specified tone, permitting variation of pitch contour along naturally motivated, pragmatically expressive principles’ (Enfield 2007: 73). Along these lines, Hả has shown that in Vietnamese, intonation can override the lexical tone of backchannel particles and short utterances (Hả 2010, Hả and Grice 2010, Hả 2012, cited by Brunelle and Kirby 2016: 199). Also in Thai, ‘some final particles systematically preserve their lexical tones, while others are overridden’ by non lexical considerations (Pittayaporn 2007).

4.9 Prosodic Hierarchy and Phonological Words

The idea of a phonological/prosodic hierarchy refers to a nested set of levels or loci of phonological organization, each of which presupposes the other (Nespor and Vogel 1986). These levels are shown in Table 4.65.

The phonological/prosodic hierarchy was not first developed with reference to MSEA languages and it does not readily apply to them (see Brunelle *et al.* 2018). Schiering *et al.* (2010) look at the applicability of the hierarchy in Vietnamese. They propose that the only units needed are the syllable and the phonological phrase. ‘Neither the prosodic foot nor the prosodic word can be motivated for the (Vietnamese) language’ (Schiering

Table 4.65 *Elements of the phonological/prosodic hierarchy*

Level of organization	Definition
Phoneme	minimal unit of contrastive sound
Mora	minimal unit for syllable construction, maximally CV in structure; a monomoraic structure cannot take stress
Syllable and syllable weight	light syllables (one mora), heavy syllables (two morae), and superheavy syllables (three morae)
Foot	a phonological unit with one stressed syllable and any number of unstressed syllables
Phonological word	a unit consisting of at least one foot

After Nespor and Vogel 1986.

et al. 2010: 680). This echoes Thomas's view that on purely descriptive grounds 'there is no significant unit in Vietnamese intermediate between the syllable and the phonological phrase' (Thomas 1962: 521). By contrast, Pham (2008) argues that 'the prosodic word almost certainly exists in Vietnamese' (Pham 2008: 2). She gives evidence from encliticization that a prosodic word in Vietnamese can be made up of a complete syllable plus one or more phonologically dependent clitics (see also Nguyen and Ingram 2007). Schiering *et al.* do not consider the encliticization evidence.

An important element in resolving these debates concerns the status of stress. Again, there is no consensus here, as Brunelle (2017: 283) notes on the state of research on Vietnamese: 'Some authors analyze the language as stressless, others have proposed that it has unbounded final stress, bounded iambic stress, or even a morphosyntactically conditioned stress system.' The 'puzzling level of disagreement' is due partly to the isolating nature of the language and the fact that 'no morphophonological alternations or phonotactic distributions that could be used to diagnose word stress have ever been reported' and partly to confusion of distinct factors including word stress, focus, and phrase final lengthening. Brunelle's instrumental study of Southern Vietnamese shows that 'there is little evidence for word stress' and that 'reports of final stress can be reinterpreted as phrase final lengthening' (Brunelle 2017: 283). Further research is needed, both on Vietnamese and more broadly on the status of the phonological hierarchy in analysis of morphophonological systems of those MSEA languages that tend towards monosyllabism.

While there is only limited discussion of the concept of a phonological/prosodic hierarchy, and the need (or not) for all of its levels in describing a given MSEA language, grammarians of the area make ample use of the basic notions of phoneme (see the phoneme charts cited in sections 4.1–4.2), as well as syllable and stress. The notions of foot and phonological word are less frequently relied upon.

This said, a number of discussions of the notion of phonological word in MSEA languages can be found, more often in those languages with a higher ratio of syllables per word. In Chrau (Austroasiatic), the phonological word is defined as ‘a phonological stretch containing only one main (stressed) syllable’, which may, in addition, ‘also contain one, only one, unstressed syllable (presyllable)’ (Thomas 1967: 28). In Moklen and Moken (Austronesian), there is ‘an iambic word canon in which an unstressed presyllable is followed by a stressed major syllable within a metrical foot’ (Larish 1997: 127). In Lao (Tai Kadai), the phonological word ‘must have at least one foot’ – foot being defined as unit with ‘one stressed syllable’ and possibly ‘one or more fully dependent and unstressed syllables’ – and ‘if there is more than one foot, then the final one will take primary stress’ (Enfield 2020). And in Daa Chin (Tibeto Burman) the phonological word ‘is always constituted of a stressed syllable ... which may optionally be preceded by an unstressed syllabic constituent’ (So Hartmann 2009: 35).

While there has been little discussion of the notion of phonological words in MSEA languages, there has been even less discussion of the relationship between the phonological word and the grammatical word (Dixon and Aikhenvald 2002, 2020). Post (2009: 938) defines a grammatical word as ‘a grammatical unit between morpheme and syntactic phrase, whose constituents are a head plus immediate (local) dependents (prefixes, suffixes, or compound elements) in a fixed, continuous order in terms of a given semantic value.’ Enfield (2020) argues that the boundaries of the grammatical word in Lao do not always align directly with the boundaries of the phonological word. This can be explicated with reference to a distinction between three kinds of unstressed syllable in Lao, shown in Table 4.66.

Table 4.66 *Kinds of unstressed syllables in Lao (TK) distinguished by their identity on the implicational scale minor > light > unstressed*

	Minor syllable	Light syllable	Unstressed syllable
<i>pakaat5</i> ‘notice’	+	+	+
<i>pa^o-kham2</i> ‘goldfish’	–	+	+
<i>bak^o kham2</i> ‘Kham (m.nonresp)’	–	–	+
<i>pak2</i> ‘pierce’	–	–	–

The unstressed syllables are highlighted in bold.

After Enfield 2020: 181.

Neither clitics nor prefixes are phonological words in themselves, but depend phonologically on the stressed syllables to which they attach. The crucial difference is that a prefix is part of a grammatical word, while a clitic is a grammatical word in itself. Generally, clitics can also occur as stressed syllables and retain essentially the same

meaning. Compare the two occurrences of the negative morpheme in the following Lao example (4.24), as a proclitic *bō^o* ‘n’t’ in the question, and as a stand alone stressed syllable *bōòl* ‘no’ in the answer.

- (4.24) Q: *mùng2 bō^o hēn3 sùa3 bōò3*
 2SG.B NEG see tiger QPLR
 ‘Did you not see tigers?’
- A: *bōòl*
 NEG
 ‘No.’

Lao | TK (Enfield 2020: 189)

It is widely held that the languages of mainland Southeast Asia are almost wholly isolating, meaning that their morpheme word ratio approaches one (see discussion in Chapter 3). But which sense of ‘word’ is intended? Consider the following Lao example with eight monosyllabic morphemes:

- (4.25) *kuu3 ni^o ka^o si^o bō^o daj^o paj^o sùù4*
 1SG.B TPC TLNK IRR NEG ACHV DIR buy
 ‘As for me, I will also not get to go and buy (it).’

Lao | TK (Enfield 2020: 211)

If we count the grammatical words in this example, there are eight, and so we see the expected ratio of one morpheme per ‘word’. But if we count the phonological words, there are just two (a three syllable foot and a five syllable foot). Questions of morphosyntactic typology in MSEA will receive radically different answers depending on the notion of ‘word’ that is assumed. More work is needed on the phonological word in MSEA languages, and its relationship to the concepts of morpheme and grammatical word.

The areally consistent nature of the phonological word being iambic in structure where the word has more than one syllable arguably correlates with other aspects of the typological structure of MSEA languages. Donegan and Stampe (1983, 2002) make this argument based on the comparison between Munda languages and other Austroasiatic (i.e., Mon Khmer) languages. Table 4.67 shows a side by side comparison of features of Munda and non Munda Austroasiatic languages.

Donegan and Stampe argue that all of these differences are causally grounded in the distinction between two kinds of ‘word canon’: trochaic versus iambic. If this one feature changes in the language, perhaps from pressure due to language contact, then this change may cause the other developments to follow by means of internal processes. Similar arguments have been made to explain thoroughgoing shifts in linguistic profile in mainland Austronesian languages (Thurgood 1999, Grant and Sidwell 2005): through

Table 4.67 *Comparison of phonological and other grammatical features of Munda and Mon-Khmer (non-Munda) Austroasiatic languages*

	Munda	Mon-Khmer
Phrase accent	falling (initial)	rising (final)
Word order	variable SOV, AN, Postpositional	rigid SVO, NA, Prepositional
Syntax	case, verb agreement	analytic
Word canon	trochaic, dactylic	iambic, monosyllabic
Morphology	agglutinative, suffixing, polysynthetic	fusional, prefixing or isolating
Timing	isosyllabic, isomoric	isoaccentual
Syllable canon	(C)V(C)	unacc. (C)V, acc. (C)(C)V(G)(C)
Consonantism	stable, geminate clusters	shifting, tonogenetic, non-geminate clusters
Tone/register	level tone (Korku only)	contour tone/register
Vocalism	stable, monophthongal, harmonic	shifting, diphthongal, reductive

After Donegan and Stampe 2002: 111.

changes in stress and phonological word structure, Moklenic and Chamic languages underwent ‘far reaching phonological change’ (Larish 1997: 135).

Larish (1997: 127) notes that Moklen and Moken have converged towards the ‘areal norms’ of MSEA, not only in the morphosyntactic typology, but also – and perhaps most strikingly from the perspective of Austronesian – in phonological organization: ‘These languages also have a tendency towards monosyllabic word shape, prefixing morphology with no suffixation, isoaccentual (stress timed) rhythm, shifting and tonogenetic consonantism, contour tones or register, reductive vocalism in unaccented syllables, and shifting/diphthongal vocalism in accented syllables’ (Larish 1997: 127). These proposals are the topic of continuing research and have been challenged in certain ways (e.g., Ring and Anderson 2018 in relation to the Munda case, Pittayaporn 2005 in relation to the Moklenic case).

Pittayaporn (2005) examines evidence to test the claim by Larish (1999: 381) that Moken ‘may have adopted word final stress under Mon Khmer influence, and this single change could have served as a catalyst for a complete typological shift’. He reviews several domains in which Moken departs from insular Austronesian language typology, including vowel height contrasts, vowel length distinctions, diphthongs, syncopation (loss of an unstressed sound) and stress, and cluster resolution. Pittayaporn argues that the change in typological structure should not be simplistically attributed to a single source of Mon Khmer influence, but that they are due to a combination of factors, including language internal developments which may have been instigated and/or accelerated by language contact (Pittayaporn 2005: 204).

Word Formation

Languages of the world differ in the number of morphemes that are used, on average, in building words. While MSEA languages are often said to provide paradigm cases of the isolating morphological profile, it is important to understand what this means. If a language were to show the logical extreme of the isolating profile – with a morpheme to word ratio of exactly 1:1 – the language would have no word formation processes to speak of. MSEA languages can be regarded as isolating only in so far as they are further towards the isolating end of the spectrum than most languages of the world. This does not mean that they have no morphology. As we shall see in this chapter, MSEA languages have ample morphological resources for forming words.

More than half a century ago, Kratochvíl (1968) debunked the myth that all words in standard Chinese are monomorphemic monosyllables. He offers a five way distinction in ‘basic types of morphological construction’, shown in Table 5.1.

Compounds are the second most frequent of these word types in texts in conversation (Kratochvíl 1968: 64). The category of ‘compound’ belies a diversity of underlying

Table 5.1 *Kratochvíl’s five-way distinction in ‘basic types of morphological construction’ in Mandarin Chinese (SN)*

Word type	Token frequency rank	Example
Simple monomorphemic monosyllabic	1	huǒ ‘fire’
Compounds (polymorphemic polysyllabic)	2	huǒ-chē ‘train’ (fire-vehicle)
Monomorphemic polysyllabic	3	húdié ‘butterfly’
Polymorphemic polysyllabic	4	ré-du ‘temperature’ (hot-DEGREE)
Polymorphemic monosyllabic	5	maō-r ‘cat-DIM’

Column 1 lists the five types; Column 2 shows their rank relative to each other in terms of frequency in texts; Column 3 gives an illustrative example of each type.
After Kratochvíl 1968.

structures. Even in his ‘elementary outline’ of compounds in Chinese, Kratochvíl (1968: 73–9) is able to fill seven pages in explicating the possible structures, under four headings: coordinate compounds, subordinate compounds, reduplicated compounds, and stump compounds. This barely scratches the surface of the depth and breadth of compounding in Chinese.

The case of Mandarin is offered here in order to illustrate the degree to which a famously ‘isolating’ language can in fact show a diverse range of ways in which words can consist of multiple morphemes. Given the profile of MSEA languages more broadly, we should expect the same situation to pertain in similarly ‘isolating’ languages of MSEA. But while the languages of MSEA can be placed towards the isolating end of a continuum, it is a myth that the languages show the logical extreme of an idealized isolating/analytic typological profile.

5.1 Form Classes

In order to build words, phrases, and sentences, a speaker of any language must start with a set of form classes. The most important of these are the familiar word classes such as noun, verb, adjective, and preposition. There are many other types of word class in addition, for example ideophones, discourse particles, and numeral classifiers (see following sections of this chapter).

We can first make some generalizations about the major types of form class found in MSEA languages.

5.1.1 Noun Verb Distinction Is Usually Clear

MSEA languages typically show a clear distinction between the noun and verb categories. There is of course some flexibility, with certain words occurring as both nouns and verbs, with related meanings. Consider, for example, the Lao word *ñam2*. When used as a noun, *ñam2* means (roughly) ‘salad’. When used as a verb, it means ‘to make salad out of’ something. Another example is *mok2*. As a noun, *mok2* is ‘a dish made in a package by steaming’, while as a verb it means ‘to steam something in a package’. These two words are members of a set of words in Lao that show regular alternation, in a similar way to the set of English words that can be used either as nouns denoting tools or as verbs for denoting actions done by those tools: *spoon*, *shovel*, *hammer*. But in the main, in MSEA languages, nouns are used as nouns and verbs are used as verbs, with clear differences in grammatical behaviour. Members of a verb class will take an array of markings, including aspectual modal distinctions such as perfect, experiential, and negation. Members of a noun class will show other grammatical behaviours, including

their use in numeral classifier constructions. See chapters 6 and 7 for overviews of the nominal and verbal domains in MSEA languages.

5.1.2 ‘Adjectives’ Are Verbs, or Are Verb Like

In some languages, an adjective class will be entirely distinct from both the noun and verb classes. But in many languages the adjective class will either be noun like or verb like, meaning that adjectives will resemble either nouns or verbs in terms of their grammatical behaviours. In English, adjectives happen to be distinct in their grammatical behaviour from both nouns and verbs. In a noun like adjective language (such as Latin), the adjective ‘old’ might be rendered noun like as, roughly, ‘an old one’, while in a verb like adjective language, such as Sedang (Smith 1979: 84), or any other MSEA language, it might be rendered verb like as, roughly, ‘to be old’. Adjectives in MSEA languages are clear examples of the verb like type, meaning that they show many of the same basic grammatical behaviours that verbs do (Comrie 2007: 41).

Post (2008) investigates the status of the adjective concept in Thai. He highlights the analyst’s role in determining word class categories, through making useful generalizations about the data rather than discovering an objectively privileged taxonomic organization in the language. His data show that adjectives may be understood not as a separate word class in Thai, but as a semantically defined subset of verbs (‘property concept’ words; Dixon 2010). In Thai, a verb, such as *dəən* ‘walk’, acts as a predicate without any marking (Post 2008: 343):

- (5.1) *khon níi dəən*
 CLF:PERSON PRX walk
 ‘This person walks.’ (active verb, intransitive predicate)
 Thai | TK | Post 2008: 343

By contrast, if a noun, such as *phrá?* ‘monk’, is to be used as a predication, it must appear as the complement of a copula verb (Post 2008: 343):

- (5.2) *khon níi pen phrá?*
 CLF:PERSON PRX ACOP monk
 ‘This person is a monk.’ (concrete noun, copular complement)
 Thai | TK | Post 2008: 343

The following two examples show that the two form classes do not overlap in behaviour:

- (5.3) **khon níi phrá?*
 CLF:PERSON PRX monk
 Thai | TK | Post 2008: 343

- (5.4) **khon* *níi pen dəən*
 CLF:PERSON PRX ACOP walk
 Thai | TK | Post 2008: 343

Now consider the behaviour of the property concept word *dii* ‘good’. The following examples show that *dii* ‘good’ follows the same pattern of grammatical behaviour as the verb *dəən* ‘walk’, and not that of the noun *phrá?* ‘monk’:

- (5.5) *khon* *níi dii*
 CLF:PERSON PRX good
 ‘This person is good.’ (verb like property term, intransitive predicate)
 Thai | TK | Post 2008: 343
- (5.6) **khon* *níi pen dii*
 CLF:PERSON PRX ACOP good
 Thai | TK | Post 2008: 343

Another grammatical property – the use of the relativizer *thîi* – similarly distinguishes nouns from verbs, where *dəən* ‘walk’ is joined to a nominal head using the relativizer, while a noun must, again, be hosted by a verb, here the copula *pen*:

- (5.7) *khon* *thîi dəən*
 CLF:PERSON REL walk
 ‘the person who walks’ (active verb, intransitive predicate)
 Thai | TK | Post 2008: 344
- (5.8) **khon* *thîi pen dəən*
 CLF:PERSON REL ACOP walk
 Thai | TK | Post 2008: 344
- (5.9) *khon* *thîi pen phrá?*
 CLF:PERSON REL ACOP monk
 ‘the person who is a monk’ (concrete noun, copular complement)
 Thai | TK | Post 2008: 344
- (5.10) **khon* *thîi phrá?*
 CLF:PERSON REL monk
 Thai | TK | Post 2008: 344

Again, the property concept word *dii* ‘good’ shows the same pattern of grammatical behaviour as the verb *dəən* ‘walk’, in contrast from the noun *phrá?* ‘monk’:

- (5.11) *khon thii dii*

CLF:PERSON REL good

‘the person who is good’ (verb like property term, intransitive predicate)

Thai | TK | Post 2008: 344

- (5.12) **khon thii pen dii*

CLF:PERSON REL ACOP good

Thai | TK | Post 2008: 344

Post (2008) describes further grammatical tests, including negation and aspect marking, showing that property concept words in Thai can be grouped together with words denoting actions rather than words denoting things.

Upon closer examination, subclasses of verbs in Thai can be clearly distinguished from each other, and this can form the basis of an argument that adjectives are indeed a distinct class from (or at least a subclass of) verbs in Thai. A property concept like *suuŋ* ‘tall’ can enter into a comparative construction, while processes/actions like *khit/dəən* cannot (Post 2008: 345):

- (5.13) *khon nii suuŋ kwaa khon nân*

[CLF:PERSON PRX]_{NP1} [tall]_x [more]CMPR [CLF:PERSON DST]_{NP2}

‘This person is taller than that person.’ (verb like property term)

Thai | TK | Post 2008: 345

- (5.14) **khon nii khit/dəən kwaa khon nân*

[CLF:PERSON PRX]_{NP1} [think/walk]_x [more]CMPR [CLF:PERSON DST]_{NP2}

*‘This person thinks/walks more than that person (does)’ (stative/active verbs)

Thai | TK | Post 2008: 345

Post gives further evidence showing how property concept words such as *dii* ‘good’ can be distinguished at a finer level of grain from action/event words such as *dəən* ‘walk’: this evidence includes distinct patterns of nominalization (Post 2008: 349), adverbialization (p. 353), and a specific pattern of reduplication with intensifying meaning (p. 364).

The case of Thai adjectives illustrates the typical situation in MSEA languages. Property concept words are like typical verbs in most respects, and in a few respects they will be distinct from those verbs. Depending on the details in the case of any specific language, and depending on one’s inclination and perhaps theoretical persuasion, a researcher may either conclude that there is no adjective class in the language (adjectives being verbs) or that adjectives form a distinct word class but have a lot in common with verbs.

Similar statements can be found for languages across MSEA. In Vietnam Mien, ‘adjectives share many properties of verbs’ (Doan and Mai 1992: 89). For example, both adjectives and verbs take direct negation with the preposed negator *m᷑j¹*:

- (5.15) *zie¹ m᷑j¹ j᷑n⁴*
 [1SG] NEG eat
 ‘[I do] not eat.’

Vietnam Mien | HM | Doan and Mai 1992: 84

- (5.16) *m᷑j¹ log⁴*
 NEG good
 ‘not good’

Vietnam Mien | HM | Doan and Mai 1992: 91

Both adjectives and verbs can be directly modified by adverbs such as the preposed marker *c᷑j⁵*:

- (5.17) *muoj² miŋ² zie¹ c᷑j⁵ miŋ²*
 2SG go 1SG also go
 ‘You go, I also go’.

Vietnam Mien | HM | Doan and Mai 1992: 83

- (5.18) *c᷑j⁵ zw᷑j⁴*
 also beautiful
 ‘also beautiful’

Vietnam Mien | HM | Doan and Mai 1992: 92

The issue of whether there really are adjectives in MSEA languages (and many other languages of the world) remains controversial, and can only be addressed seriously with detailed analysis of the full facts for each language. Such depth of analysis goes beyond our scope here, and indeed beyond the usual scope of the kinds of reference grammar works that we rely on in this book. In any case, we can safely say that adjectives are distinctly verb like in this part of the world.

5.1.3 ‘Adpositions’ Are Often Nouns or Verbs

A third generalization about word classes in the MSEA area is a noted lack of a distinct word class that is widely observed elsewhere in the world, namely the class of prepositions. In many MSEA languages, meanings of English prepositions such as *to*, *with*, *by*, *from*, *on*, *under*, *through*, and *behind* – members of a distinct form class in English and many other languages – are expressed either by verbs or nouns in specific grammatical constructions, as highlighted (in bold) in the following examples:

- (5.19) *Nws cog zaub ntsuab rau hauv vaj lawm*
 3SG plant vegetable green be.at interior garden already
 ‘She planted green leafy vegetables in her garden.’
 Hmong | HM | Clark 1989: 188
- (5.20) *kñom caq mcul knoy sac*
 1SG insert needle interior flesh
 ‘I stuck a needle in her skin.’
 Khmer | AA | Clark 1989: 188
- (5.21) *Chị ngồi trên ghế*
 sister sit top chair
 ‘You sit on the chair.’
 Vietnamese | AA | Clark 1989: 189
- (5.22) *léih deui hàih hái jēung téi hahbihn*
 2SG pair shoes be.at CLF table underside
 ‘Your shoes are under the table.’
 Cantonese | SN | Matthews and Yip 1994: 117

While verbs and nouns carry a significant load in performing preposition like functions, such as adding a peripheral argument to a core clause (e.g., a goal of motion, or a location of an event or action), this does not mean that MSEA languages have no true prepositions. For example, in Lao, the word *duaj4* neither a verb nor a noun functions as a dedicated preposition like particle which can introduce a comitative or instrumental argument into a peripheral position in the clause. Here is an example:

- (5.23) *man2 kin3 khaw5 duaj4 mùù2*
 3SG.B eat rice with hand
 ‘He eats/ate rice with his hands.’
 Lao | TK

This said, the meaning conveyed in (5.23) is more idiomatically conveyed by a multi verb construction, in which the verb *saj4* ‘to use’ takes as its complement a noun referring to the entity used as an instrument:

- (5.24) *man2 saj4 mùù2 kin3 khaw5*
 3SG.B use hand eat rice
 ‘He eats/ate rice with his hands.’
 Lao | TK

Other MSEA languages similarly have few true prepositions, often having just one preposition with a semantically general locative meaning. In Kri, for example, the preposition *taa* takes locations as its complement; the word does not have other grammatical functions such as noun or verb:

- (5.25) *hang vât phàsiì taa naaq*
 3SG spend tax LOC DEM.EXT
 ‘He paid tax there.’
 Kri | AA | Enfield notes (050729b:02.32)

- (5.26) *lôôh taa kuraaq cawq naaq*
 exit LOC path 1PL DEM.EXT
 ‘(It will) emerge at our path there.’
 Kri | AA | Enfield notes (050729b:14.31)

5.1.4 Verbs Can Act as ‘Adverbs’ without Derivation

MSEA languages tend not to have large classes of adverbs comparable to derived English words such as *slowly*, *carefully*, *accidentally*, and *apparently*. A frequent way to modify verbs in MSEA languages is simply to use another verb as its direct modifier, without any formal derivation. See Chapter 7 for examples, under the rubric of multi-verb constructions.

5.2 Compounding

A compound consists of two open class words joined together to form a single word. In English, for example, *blackbird* is a compound consisting of the two words *black* and *bird*. This is distinct from the phrase *black bird* in two important ways. The first concerns how it is pronounced. The compound *blackbird* takes stress only on the first syllable, while in the phrase *black bird*, the two component words take roughly equal stress. The second difference concerns meaning. The meaning of the phrase *black bird* is simply a combination of the meaning of the two component words – a *black bird* is a bird that is black – while a compound often has a meaning that, in the typical manner of morphological derivation, cannot be deduced from the sum of the parts: a *blackbird* is not any bird that is black (and indeed not all blackbird species are black).

Kratochvíl (1968: 73ff.) provides a nuanced analysis of compounding in Mandarin Chinese. His analysis is useful as a likely model for MSEA languages in general. He identifies numerous categories of compound. For example, coordinate compounds

resemble the structure of syntactic coordinate constructions. The meanings of the root forms and of the whole derived word are similar. For example:

- (5.27) *chéng* ‘town wall, walled town’ + *shì* ‘market, market town’ → *chéngshì*
‘town’

Mandarin | SN | Kratochvíl 1968: 75

- (5.28) *péng* ‘friend’ + *yǒu* ‘friend’ → *péngyou* ‘friend’¹

Mandarin | SN | Kratochvíl 1968: 75

- (5.29) *jiāo* ‘to establish contact’ + *tōng* ‘to go through’ → *jiāotōng*
‘communication’

Mandarin | SN | Kratochvíl 1968: 75

Subordinate compounds are of several types, including attribute head, head referent, head modifier, head measure, and reduplicated, as in the following examples, respectively (see Kratochvíl 1968: 76ff. for details):

- (5.30) *jūn shì*
armed.force affairs
‘military affairs’

Mandarin | SN | Kratochvíl 1968: 76

- (5.31) *bào chóu*
to.recompense enmity
‘to avenge’

Mandarin | SN | Kratochvíl 1968: 76

- (5.32) *shuō míng*
to.say clear
‘to explain’

Mandarin | SN | Kratochvíl 1968: 77

- (5.33) *zhǐ zhāng*
paper a.sheet.of
‘paper [in sheets]’

Mandarin | SN | Kratochvíl 1968: 77

¹ The roots *péng* and *yǒu* do not occur in contemporary Mandarin as stand-alone words, but they do occur in other multimorphemic expressions in which they contribute the meaning ‘friend’. The compound *péngyou* is the everyday word for ‘friend’ in Mandarin.

- (5.34) *rén~rén*
 man~RDP
 ‘everybody, each person’
 Mandarin | SN | Kratochvíl 1968: 78

Jenny, Weber, and Weymuth give an overview of compounding in Austroasiatic languages (Jenny *et al.* 2015: 38). They note that, as with English examples like *blackbird*, compounds are distinguished from phrases mainly by stress pattern and semantics (Jenny *et al.* 2015: 38, although they note that Khasi has overt morphological markers to distinguish compounds from phrases; Jenny *et al.* 2015: 41).

Compounding is sometimes signalled by tone changes to a stem. Here are some illustrative examples from White Hmong:

- (5.35) *zaub ntsim*
 vegetable peppery
 ‘peppery vegetable’
 White Hmong | HM | Ratliff 1992: 59
- (5.36) *zaub ntsig*
 vegetable peppery
 ‘peppery vegetable’ [as a particular plant or a pickled dish]
 White Hmong | HM | Ratliff 1992: 59
- (5.37) *muaj txiaj*
 have money
 ‘have money’
 White Hmong | HM | Ratliff 1992: 60
- (5.38) *muaj txiag*
 have money
 ‘rich’
 White Hmong | HM | Ratliff 1992: 61 | Downer 1967: 594

Compounds often consist of a simple juxtaposition of noun and noun. The resultant meaning is not usually predictable from the sum of the parts, though of course it will be motivated by the meanings of the parts. The following examples are from across the MSEA language families:

- (5.39) *muh mat*
 nose eye
 ‘face’
 Koho | AA | Jenny *et al.* 2015: 39

- (5.40) *céhen có:n*
 thing tree
 ‘bird’
 Car | AA | Jenny *et al.* 2015: 39
- (5.41) *me: thi*
 eye water
 ‘tear’
 Pwo Karen | TB | Kato 2003: 639
- (5.42) *pê gî*
 bee fluid
 ‘honey’
 Lahu | TB | Matisoff 1973a: 54
- (5.43) *siaw²⁴ ka:n³³*
 wing fish
 ‘fins’
 Hainan Cham | AN | Thurgood *et al.* 2014: 105
- (5.44) *vw̄m² k'ut⁵*
 water hole
 ‘well’
 Vietnamese Mien | HM | Doan and Mai 1992: 52
- (5.45) *tin¹ fiuŋ⁴*
 foot hand
 ‘techniques, skills’
 Zhuang | TK | Luo 2008: 325

Another common structure for compounds in MSEA languages is the combination of noun and verb/adjective, as illustrated in the following:

- (5.46) *t̄q:k c̄a:*
 water eat
 ‘drinking water’
 Chong | AA | Jenny *et al.* 2015: 40
- (5.47) *ne:ok cap trɔj*
 man catch fish
 ‘fisherman’
 Khmer | AA | Jenny *et al.* 2015: 40

- (5.48) *ne²¹ ee⁵³*
 mother new
 ‘stepmother’
 Aizhai Miao | HM | Yu 2010: 365

- (5.49) *vet³¹ lui³³*
 ear long
 ‘rabbit’
 Judu Gelao | TK | Kang 2009: 35

There are also many compounds consisting of verb and verb. As Jenny *et al.* (2015) note, verbal compounds ‘are not always clearly distinguishable from multiple verb predicates or serial verb constructions’ (Jenny *et al.* 2015: 40; cf. Chapter 7 below). These might combine verbs with two related meanings, as in the following examples:

- (5.50) *mua bán*
 buy sell
 ‘trade’
 Vietnamese | AA | Jenny *et al.* 2015: 40 | Brunelle 2015
- (5.51) *sph ràn*
 sell buy
 ‘trade’
 Mon | AA | Jenny *et al.* 2015: 40

Or they might combine verbs that are synonyms, or near synonyms, as these examples show:

- (5.52) *c^hoŋ c^ha:*
 eat eat
 ‘eat’
 Bunong | AA | Jenny *et al.* 2015: 40 | Phillips 1973: 131
- (5.53) *luə̄i biə̄ŋ*
 be.lazy be.lazy
 ‘be lazy’
 Vietnamese | AA | Jenny *et al.* 2015: 40 | Brunelle 2015: 916

In this example from Phan Rang Cham, the compounding of two verbs yields a noun:

- (5.54) *bay məñum*

eat drink

‘a feast’

Phan Rang Cham | AN | Thurgood 2005: 502

Certain quite specific compounds are shared across the area. For example, many MSEA languages lack a word meaning ‘parents’, but instead use a noun noun compound of ‘mother’ and ‘father’ for this meaning. The relative order of the two parts of the compound varies. Compare the first and second group of examples here:

- (5.55) *cha me*

father mother

‘parents’

Vietnamese | AA | Jenny *et al.* 2015: 39 | Brunelle 2015

- (5.56) *phɔɔ mē*

father mother

‘parents’

Thai | TK | Iwasaki and Horie 2005: 37

- (5.57) *mí bá*

mother father

‘parents’

Burmese | TB | Hnin Tun 2014: 272

- (5.58) *me? ba:p*

mother father

‘parents’

Koho | AA | Jenny *et al.* 2015: 39

- (5.59) *?a?i ?a?am*

mother father

‘parents’

Pacoh | AA | Jenny *et al.* 2015: 39

- (5.60) *now pa*

mother father

‘parents’

Sedang | AA | Jenny *et al.* 2015: 39

The examples shown above in this section only scratch the surface of the phenomena of compounding in MSEA languages. Like any process of morphological derivation, the

details of compounding are textured and complex. There is a great array of underlying formal structures and an even greater array of underlying semantic relations. The process of compounding shows variation from language to language in terms of its formal properties, its productivity, and its underlying semantic principles, if any can be detected. One of the defining characteristics of compounds is that their semantics are idiosyncratic. Most of the work of describing compounds, then, is a matter of careful lexicography, and so will tend to receive only superficial or tentative treatment in grammatical descriptions.

5.3 Psycho-collocations

Many languages of MSEA feature a type of lexicalized compound known as a psycho collocation (after Matisoff 1986). An example is Pwo Karen *θà khō* ‘worry’, literally ‘heart hot’ (Kato 2019: 139). Psycho collocations have two main defining properties. The first is semantic. Their meanings denote states and properties in the interconnected domains of cognition, emotion, and personality: concepts like happy, sad, surprised, frightened, impulsive, kind, and arrogant. The second property concerns their morphemic make up. A psycho collocation features a noun that explicitly refers to a part of the body, as an imagined locus of psychological activity. As Matisoff (1986: 8) notes, in English words like *happy*, *angry*, and *scared*, ‘there is no overt clue [that the words refer] to a psychological or mental phenomenon’, while in East and Southeast Asia ‘such concepts are expressed by collocations, one of whose constituents refers explicitly to the psyche’. This locus of the psyche is almost always an internal organ (usually the heart or liver), but it may also be another part of the body, such as eye, ear, tooth, face, mouth, nose, or bone.

Psycho collocations are not unique to MSEA. In languages all over the world, psychological predicates often feature explicit reference to a part of the body as an imagined seat of psychological activity (see, for example, Enfield and Wierzbicka 2002 on Ewe, Japanese, Hebrew, Kuot, Oneida, Russian, English, Koromu, Kaytetye, as well as Southeast Asian languages Lao and Chinese; Sharifian *et al.* 2008 on Kuuk Thaayorre, Basque, Persian, Northeastern Neo Aramaic, Old English, English, Dutch, and Tunisian Arabic, as well as South/East Asian languages Indonesian, Malay, Chinese, Japanese, and Korean). That said, MSEA stands out as an area in which these kinds of expression appear to be highly elaborated.

Studies of psycho collocations subsequent to Matisoff’s (1986) survey include Clark (1996) on the MSEA area, Jaisser (1990) on White Hmong, and VanBik (1998) on Hakha Lai. Most recently, a reference volume on the mainland Southeast Asian

language area (Vittrant and Watkins 2019) supplies data on psycho collocations in a range of MSEA languages. In the rest of this section, we draw on the data presented there.

Table 5.2 shows some examples from Bangladesh Khumi (Kuki Chin and Tibeto Burman), mostly featuring the word for heart, but also including eye and ear. Table 5.3 supplies some examples from another Tibeto Burman language, Pwo Karen. These mostly refer to the heart but also the ear and face. And in Yongning Na (Mosuo),

Table 5.2 *Example psycho-collocations in Bangladesh Khumi (Kuki-Chin, TB)*

Heart		
<i>p'liwng⁴ bi⁴</i>	'heart hot'	'unhappy, sad, upset'
<i>p'liwng⁴ hây³</i>	'heart good'	'generous'
<i>p'liwng⁴ a'¹dü¹</i>	'heart equal'	'harmonious'
<i>p'liwng⁴ a'¹pha¹</i>	'heart sufficient'	'happy, contented, satisfied'
<i>p'liwng⁴ khay²</i>	'heart keep'	'hold a grudge'
<i>p'liwng⁴ dång⁴</i>	'heart think.about'	'pay attention, think'
<i>p'liwng⁴ a'¹ma¹</i>	'heart lost'	'sad, upset'
<i>p'liwng⁴ a'¹phöng⁴</i>	'heart suspect'	'worry'
<i>p'liwng⁴ kha¹</i>	'heart strong'	'feel sure, certain, secure'
Eye		
<i>möy³ a'¹dåy²</i>	'eye MID.lure'	'attracted to, interested in'
Ear		
<i>k'no⁴ döy⁶-lä³</i>	'ear die-NEG'	'lie awake, unable to fall asleep'

After Peterson 2019: 21.

Table 5.3 *Example psycho-collocations in Pwo Karen (TB)*

Heart		
<i>θà-kho</i>	'heart to.be.hot'	'to worry'
<i>θà-xwí</i>	'heart to.be.light'	'to be glad'
<i>θà-làn</i>	'heart to.ascend'	'to be angry'
<i>θà-mé</i>	'heart to.sprout'	'to fear'
<i>ká-θà</i>	'to.be.difficult heart'	'to be shy'
Eye		
<i>ná-káin</i>	'ear to.be.bent'	'to be dishonest'
Ear		
<i>mé-kho</i>	'face to.be.hot'	'to be ashamed'

After Kato 2019: 139.

a Tibeto Burman language of southwestern China, psycho collocations also mostly feature the heart as the locus, but also bone, as shown in Table 5.4.

Table 5.4 *Example psycho-collocations in Yongning Na (Mosuo) (TB)*

Heart		
<i>nu³¹-mi¹³-fi³³</i>	‘heart happy’	‘be happy’
<i>nu³¹-mi¹³-kwɔ³³-huu³³</i>	‘heart LOC go’	‘be pleased’
<i>nu³³-gɔ³³</i>	‘heart hurt’	‘be pitiable’
<i>nu³¹-mi¹³-kwɔ³³k^huu¹³</i>	‘heart LOC place(v.)’	‘memorize’
<i>nu³³-t^hswe³³</i>	‘heart insert’	‘trust’
Bone		
<i>ʒ³³-mə³³-t^hu³³</i>	‘bone NEG arrive’	‘have no strength’

After Lidz 2019: 240.

Southern Min (Sinitic) psycho collocations reference a wide range of body parts, as shown by examples featuring ear channel, heart, face, mouth, heart, nose, bone, tongue, and tooth in Table 5.5. In the Paraok variety of Wa (Austroasiatic), psycho collocations refer exclusively to the heart. Table 5.6 gives some examples. Vietnamese psycho collocations feature reference to *dạ* ‘stomach, belly, heart’, *mặt* ‘face’, and *lòng* ‘entrails, bosom’. Table 5.7 shows examples with *lòng* (note two orderings are possible).

Table 5.5 *Example psycho-collocations in Southern Min (SN)*

<i>hi⁷-khang¹ khin¹</i>	‘ear:channel light’	‘be gullible’
<i>sim¹-sek⁴</i>	‘heart fit’	‘interesting’
<i>bin⁷-sek⁸</i>	‘face ripe’	‘be familiar with’
<i>chhui³-ta¹</i>	‘mouth dry’	‘thirsty’
<i>sim¹-sng¹</i>	‘heart sour’	‘sad’
<i>phi⁷ sng¹</i>	‘nose aching’	‘mean, stingy’
<i>kek⁴-kut⁴</i>	‘provoke bone’	‘to be contrary’
<i>toa⁷-chih⁸</i>	‘big tongue’	‘to list, stutter’, ‘someone who stutters’
<i>thih⁴-khi²</i>	‘iron tooth’	‘to be uncompromising’

After Chappell 2019: 193.

Table 5.6 *Example psycho-collocations in Wa (Paraok) (AA)*

<i>pat r^hɔm</i>	‘heart is salty’	‘angry’
<i>Paoh r^hɔm</i>	‘heart is hot’	‘upset, irritated’
<i>tjip r^hɔm</i>	‘heart is big’	‘bold’
<i>tyk r^hɔm</i>	‘heart is asthmatic’	‘sad’
<i>l^haoŋ r^hɔm</i>	‘heart is tall’	‘arrogant’
<i>n^hiəŋ r^hɔm</i>	‘heart is hard, stiff’	‘stubborn’
<i>b^hay r^hɔm</i>	‘heart is wide open’	‘generous’
<i>t^hɔ r^hɔm</i>	‘heart is shallow’	‘intolerant’

After Watkins 2019: 443.

Table 5.7 Example psycho-collocations in Vietnamese (AA) featuring reference to *lòng* 'entrails, bosom'

<i>đau lòng</i>	'be.aching <i>lòng</i> '	'be heartbroken'
<i>nóng lòng</i>	'be.hot <i>lòng</i> '	'be impatient, be anxious'
<i>vui lòng</i>	'be.joyful <i>lòng</i> '	'be happy'
<i>lòng tham</i>	' <i>lòng</i> be.greedy'	'avarice, miserliness'

After Do-Hurinville and Dao 2019: 396 7.

In Thai (Southwestern Tai), the most common psycho noun in Thai is *cay* 'heart'. There are two syntactic types of simple bimorphemic psycho collocations, where '[V *cay*] expresses a temporary state of mind, whereas [*cay* V] expresses a more permanent trait of character' (Jenny 2019b: 568; see Enfield 2007: 264 on the same situation in closely related Lao). Some modifiers occur with *cay* in both orders. See examples in Table 5.8.

Table 5.8 Example psycho-collocations in Thai (TK) featuring reference to *cay* 'heart'

<i>cay</i> 'heart' + V		
<i>cay-yen</i>	'heart cool'	'patient'
<i>cay-dam</i>	'heart black'	'unkind, stingy'
<i>cay-kwâay</i>	'heart wide'	'generous'
<i>cay-khén</i>	'heart hard'	'unyielding'
V + <i>cay</i> 'heart'		
<i>?ün-cay</i>	'warm heart'	'appeased'
<i>cèp-cay</i>	'hurt heart'	'feel hurt'
<i>sǐə-cay</i>	'waste heart'	'sad, unhappy'
<i>plèek-cay</i>	'strange heart'	'astonished, surprised'
Pairs of opposite orders		
<i>dii-cay</i>	'good heart'	'happy'
<i>cay-dii</i>	'heart good'	'kind'
<i>nóɔy-cay</i>	'little heart'	'sad, disappointed'
<i>cay-nóɔy</i>	'heart little'	'easily offended'
<i>róɔn-cay</i>	'hot heart'	'anxious'
<i>cay-róɔn</i>	'heart hot'	'impatient'

After Jenny 2019b: 568.

Psycho collocations in Hmong (Mong Leng) mostly feature reference to the liver, but also the stomach and the heart, as shown in Table 5.9, which shows the same two orderings illustrated in Table 5.8 for Thai (see also Jaisser 1990 on White Hmong).

Table 5.9 *Example psycho-collocations in Hmong (Mong Leng, Hmong-Mien) mostly featuring reference to the liver*

V-N psychological states or sensations		
<i>zoo-sab</i>	‘good liver’	‘be happy’
<i>chim-sab</i>	‘angry liver’	‘be angry’
<i>khua-sab</i>	‘wistful liver’	‘be wistful, nostalgic’
<i>txhawj-xeeb</i>	‘worry heart’	‘be worried’
N-V psychological or ethical properties		
<i>sab-zoo</i>	‘liver good’	‘be good hearted’
<i>sab-phem</i>	‘liver ugly’	‘be evil hearted’
<i>sab-luj-sab-dlaav</i>	‘liver big liver wide’	‘be generous, brave’
<i>plaab-plawv-luj</i>	‘stomach heart large’	‘be intelligent’

After Mortensen 2019: 618 9.

Psycho collocations in Colloquial Eastern Cham (Austronesian) also feature reference to the liver, but also the body, head, and belly. Some examples are shown in Table 5.10.

Table 5.10 *Example psycho-collocations in Colloquial Eastern Cham (AN) featuring reference to the liver, body, head, and belly*

<i>sam taj</i>	‘beautiful liver’	‘good hearted’
<i>cha? taj</i>	‘ugly liver’	‘evil, bad’ (of a person)
<i>prɔj taj</i>	‘large liver’	‘reckless, foolhardy’
<i>prɔj rup</i>	‘large body’	‘self-important’
<i>prɔj kɔ?</i>	‘large head’	‘arrogant’
<i>prɔj tjan</i>	‘large belly’	‘generous’

After Brunelle and Hǎn 2019: 533.

In some MSEA languages, psycho collocations do not feature a physical body part but use a term that is translated into English as ‘mind’. In three prominent cases, this involves a borrowing from Pali *citta* ‘mind’. Examples from Mon (Austroasiatic), Khmer (Austroasiatic), and Burmese (Tibeto Burman) are shown in Tables 5.11, 5.12, and 5.13, respectively.

Table 5.11 Example psycho-collocations in Mon (AA)

cvt ‘mind’-PM		
usually (but not always) expressing a permanent trait		
cvt kle?	‘heart short’	‘short tempered’
cvt klōp	‘heart long’	‘tolerant’
cvt prsh	‘heart quick’	‘impulsive’
PM-cvt ‘mind’		
usually (but not always) expressing transient feelings		
mip cvt	‘happy heart’	‘happy’
?on cvt	‘little heart’	‘sad’
pɔn cvt	‘full heart’	‘agree’

The noun *cvt* ‘mind’ (from Pali *citta*) does not refer to the anatomical heart.

After Jenny 2019a: 286.

Table 5.12 Example psycho-collocations in Khmer (AA)

ceut kdav	‘heart is hot’	‘disappointed’
ceut klang	‘strong heart’	‘upset, stressed’
ceut tuliaj	‘spacious heart’	‘generous’
khaw:k ceut	‘empty heart’	‘disappointed’
traceak ceut	‘cool heart’	‘calm, serene’

The noun *ceut* ‘mind’ (from Pali *citta*) does not refer to the anatomical heart.

After Haiman 2019: 335.

Table 5.13 Example psycho-collocations in Burmese (TB)

sei?‑kouN ²	‘mind run.out’	‘be fed up with’
sei?‑paɔ?	‘mind explode’	‘become angry’
sei?‑to?	‘mind be.short’	‘to be short-tempered’
sei?‑jo?	‘mind be.complex’	‘be confused’

The noun *sei?* ‘mind’ (from Pali *citta*) does not refer to the anatomical heart.

After Vittrant 2019: 75.

5.4 Reduplication

Reduplication is a linguistic process used in most if not all languages of the world. In an overview of reduplication in Austroasiatic languages, Jenny *et al.* (2015: 42) identify

two formal parameters for distinguishing between sub types. One parameter concerns whether the reduplication is full or partial. The other concerns whether there is stem alternation. These formal parameters cross cut various semantic functions. The distinction works for the other languages of this area as well.

In its simplest form, reduplication is simply repetition of a specified form, where this repetition conveys a predictable meaning. For example, in Thai, in one pattern of reduplication, when a noun is reduplicated, the output is a plural form of that noun:

- (5.61) *dek~dek*
 child~RDP
 ‘children’
 Thai | TK

In Khmer, reduplication of a noun can signal a form of semantic generalization:

- (5.62) *srej~srej*
 woman~RDP
 ‘women in general’
 Khmer | AA | Alves 2015: 550 | Huffman 1970: 298

In Xia’ao Zhuang, reduplication of a numeral classifier means ‘each and every’:

- (5.63) *nə:n⁴²~nə:n⁴²*
 QW~RDP
 ‘each and every (one)’
 Xia’ao Zhuang | TK | Wei 2012: 111

Reduplication can signal a family of meanings relating to attenuation of the intensity of an action or distribution of an ongoing action over time. In Taiwanese, reduplicating a verb ‘suggests that the action is performed for a short unspecified amount of time’ (Lin 2015: 214):

- (5.64) *góá beh seng khòaⁿ~khòaⁿ*
 1SG want first see~RDP
 ‘I want to read this book a little (before I buy it)’
 Taiwanese | SN | Lin 2015: 214

In Phan Rang Cham, full reduplication signals ‘some sort of distributive meaning’ (Thurgood 2005: 494), as shown in this example which includes vowel mutation in the reduplicated syllable:

- (5.65) *sənɪŋ~senəŋ*
 to.think~RDP

‘to meditate’

Phan Rang Cham | AN | Thurgood 2005: 495, from Blood 1967: 17

And in Vietnam Mien, reduplication of an adjective serves to attenuate the intensity of the predicated property:

- (5.66) *pe²~pe²*
 RDP~white
 ‘whitish’

Vietnam Mien | HM | Doan and Mai 1992: 51

This semantic function of attenuation is also shown in the following examples from Muong and Vietnamese:

- (5.67) *haw³⁵~haw³⁵*
 want~RDP
 ‘to want somewhat’

Muong | AA | Alves 2015: 550 | Hoàng V. H. 1998: 168

- (5.68) *dēp^{11?}~dēp^{11?}*
 pretty~RDP
 ‘somewhat pretty’

Vietnamese | AA | Alves 2015: 550 | Thompson 1965: 152

The following examples of reduplication from Katu, Vietnamese, and Wa show the function of marking continuation or distribution of an action over time:

- (5.69) *ta:p~ta:p*
 clap~RDP
 ‘to clap repeatedly’

Katu | AA | Alves 2015: 539 | Nguyễn and Nguyễn 1998: 90

- (5.70) *kɔj³³~kɔj³³*
 look~RDP
 ‘to keep looking’

Vietnamese | AA | Alves 2015: 539 | Thompson 1965: 152

- (5.71) *du~dik*
 RDP~to.pedal
 ‘to pedal repeatedly and randomly’

Wa | AA | Alves 2015: 539 | Zhou and Yan 2006: 390

Reduplication in Eastern Kayah Li is not a word formation mechanism but has the meaning ‘also, too, either’. This form of reduplication is unusual in that it is ‘a simple matter of copying whatever syllable happens to be clause final, regardless of either form class or syntactic function’ (Solnit 1997: 52):

- (5.72) *vē ma ʔe kā phé thé ja~ja*
 1s be.so eat COM simply pig flesh~RDP
 ‘I only ate pork, too (as did he).’
 Eastern Kayah Li | TB | Solnit 1997: 52

- (5.73) *vē cwá to~to*
 1s go NEG~RDP
 ‘I won’t go either.’
 Eastern Kayah Li | TB | Solnit 1997: 52

Full reduplication in Austroasiatic languages often functions to modify predictions, for example with attenuation or intensification, such as in the following examples:

- (5.74) *đó đó*
 red~RDP
 ‘reddish’
 Vietnamese | AA | Jenny *et al.* 2015: 42

- (5.75) *kwaŋ~kwaŋ*
 big~RDP
 ‘really big’
 Bunong | AA | Jenny *et al.* 2015: 42

- (5.76) *jal³³~jal³³*
 long~RDP
 ‘rather long’
 Ruc | AA | Alves 2015: 550 | Nguyễn 1993: 90

Another widespread function of full reduplication with no stem alternation in Austroasiatic languages is ‘plurality or non specificity’, both broadly construed. See the following examples:

- (5.77) *ne:w~ne:w co:k ce:n cā:k tɔŋ*
 child~RDP run come from home
 ‘The children run from their home’
 Chong | AA | Jenny *et al.* 2015: 43

- (5.78) *he⁵⁵~he⁵⁵*
 day~RDP
 ‘every day, always’
 Bugan | AA | Jenny *et al.* 2015: 43

- (5.79) *her~her*
 fly~RDP
 ‘continue to fly’
 Khasi | AA | Jenny *et al.* 2015: 43

- (5.80) *pìm l̥v d̥eh h̥vm~h̥vm, ɻuə̥ h̥ù? p̥ateh p̥uh*
 manner INTER 3 speak~RDP 1SG NEG believe NEG
 ‘However he speaks, I don’t believe him’
 Mon | AA | Jenny *et al.* 2015: 43

- (5.81) *mənih m̥ùə̥~m̥ùə̥*
 human one~RDP
 ‘someone’
 Mon | AA | Jenny *et al.* 2015: 43

Reduplication with stem alternation is also widespread. Here are some examples of reduplication with stem alternation in a selection of Austroasiatic languages (the Vietnamese examples showing stem alternations in tone, Khmer showing alternations in onset, and Sedang, Bru, and Mang showing alternations in vowel and/or rhyme):

- (5.82) *d̥o~d̥ó*
 RDP~red
 ‘reddish’
 Vietnamese | AA | Jenny *et al.* 2015: 44

- (5.83) *n̥hè~n̥hə̥*
 RDP~light
 ‘rather light’
 Vietnamese | AA | Jenny *et al.* 2015: 44

- (5.84) *m̥è:n~l̥è:n*
 really~RDP
 ‘really’
 Khmer | AA | Jenny *et al.* 2015: 44

- (5.85) *slò:t~bò:t*
 gentle~RDP
 ‘nice, kind agreeable’
 Khmer | AA | Jenny *et al.* 2015: 44
- (5.86) *me ?ni~?nɔ ku~ka mi~mot tuy me*
 so RED~pull.out RED~bite RED~enter inside so
 ‘(The snake) came out (of the water) killed (the girl) and entered into that (water).’
 Sedang | AA | Jenny *et al.* 2015: 44
- (5.87) *tu ta?*
 RDP~to.do/make
 ‘to do/make in general’
 Bru | AA | Alves 2015: 550 | Nguyễn *et al.* 1986: 40
- (5.88) *əa²² eac³⁵*
 to.eat~RDP
 ‘to eat in general’
 Mang | AA | Alves 2015: 550 | Nguyễn *et al.* 2008: 164.

In Jahai, the possibilities for stem alternation are numerous, with a single stem potentially undergoing several distinct alternations to mark several distinct meanings:

- (5.89) *cip* ‘go’
cip~cip CONTINUATIVE full reduplication, no change
cip~cip DURATIVE full reduplication, with vowel alternation in first syllable
cp~cip IMPERFECTIVE full reduplication, with vowel deletion in first syllable
ca~cip RECIPROCAL partial reduplication, with rime alternation in first syllable
 Jahai | AA | Jenny *et al.* 2015: 44 5 | Burenhult 2005: 94f

In some cases that resemble reduplication in MSEA languages, there are no attested corresponding base forms from which the apparent reduplications are derived. Just as we know in English what *cranberry* means without having a meaning for *cran*, what looks like a reduplication may not have a known source stem in the language. Here are two examples:

- (5.90) *mic~mec*
 ‘ant’
 Mlabri | AA | Jenny *et al.* 2015: 45

- (5.91) *le:p~la:p*
 ‘butterfly’
 Kui | AA | Jenny *et al.* 2015: 45

Kruspe (2004) describes a complex range of processes of derivational reduplication in Semelai. For example, in a form of partial reduplication called light syllable reduplication, everything in the stem up to the final coda is copied and affixed before the stem. Light syllable reduplication mostly applies to verbal roots with an intensification function. The following examples show light syllable reduplication of verbs with the semantic effect of ‘intensification’:

- (5.92) *ki brkas* *?ilɔ~?ilɔk*
 3_A tie.together INTNS~be.good
 She tied (them) together really well.
 Semelai | AA | Kruspe 2004: 149

- (5.93) *kɔhn pon dkʰe~dkʰes*
 3_S then INTNS~be.close
 She was really close.
 Semelai | AA | Kruspe 2004: 149

- (5.94) *rpo~rpɔk tʰi*
 INTNS~clap hand
 really clapping (their) hands
 Semelai | AA | Kruspe 2004: 149

- (5.95) *kna?* *ki cŋe~cŋew*, *ki jyɔk* *sma?*
 happen 3_A INTNS~look.down 3_A observe person
 It happened (that) (when) he looked down, he saw someone.
 Semelai | AA | Kruspe 2004: 149

The following examples show light syllable reduplication of nouns with the semantic effect of plurality or non specificity:

- (5.96) *rlmɔ~rlmɔl*
 RDP~male human
 ‘various men’
 Semelai | AA | Kruspe 2004: 220

- (5.97) *jadi? svn, dom, kraba~krabat de ca<r>lɔn*
 happen SC AFF RDP~official 3PL.A back<CAUS>
 ‘And so if it happened see, indeed, the various officials turned their backs
 (on him).’
 Semelai | AA | Kruspe 2004: 220

The second broad type of derivational reduplication in Semelai is coda copy. The coda consonant of the final syllable of a bisyllabic stem is attached to the coda position of the stem’s open penultimate syllable. When the stem is monosyllabic, its onset is also reduplicated to form a penultimate syllable, while the V slot is filled by an epenthetic rule. The functions of coda copy are diverse. The following examples illustrate coda copy reduplication of verbs with the effect of making the verb intransitive and conveying an imperfective aspect:

- (5.98) *ki gɔ? dlɔy*
 3A fell tree
 ‘He felled the tree.’
 Semelai | AA | Kruspe 2004: 111

- (5.99) *kəhn g?~gɔ?*
 3S IMPERF~fell
 ‘He was felling.’
 Semelai | AA | Kruspe 2004: 111

- (5.100) *gŋ goy*
 RDP~to.carry
 ‘to be carrying’
 Semelai | AA | Kruspe 2004: 111

- (5.101) *mc mɔc*
 RDP~to.dive.into
 ‘to be diving’
 Semelai | AA | Kruspe 2004: 112

Coda copy reduplication in Semelai is also productive in word class derivation. For example, a noun can be changed into a stative verb. In one pattern, nouns of body parts and clothing are converted to a verb meaning ‘to have N’:

- (5.102) *ps pvs*
 RDP~tail
 ‘to have a tail’
 Semelai | AA | Kruspe 2004: 151

- (5.103) *th tuh*
 RDP~breast
 ‘to have breasts’
 Semelai | AA | Kruspe 2004: 151

- (5.104) *sk suk*
 RDP~hair/fur/features
 ‘to have hair, fur, feathers’
 Semelai | AA | Kruspe 2004: 151

In another pattern, which involves copying of the coda as an infix (thus stretching the limits of the concept of ‘reduplication’), a noun for a thing that can be utilized is converted into a verb that means ‘to be utilizing N’:

- (5.105) *b<h>nih*
 <RDP>seed
 ‘to sow rice, of women’
 Semelai | AA | Kruspe 2004: 151

- (5.106) *c<?>ru?*
 <RDP>tray.used.to.winnow.rice
 ‘to winnow rice with a tray’
 Semelai | AA | Kruspe 2004: 151

- (5.107) *sa<m>tvm*
 <RDP>right.hand
 ‘to use right hand’
 Semelai | AA | Kruspe 2004: 151

- (5.108) *sa<l>wel*
 <RDP>left.hand
 ‘to use left hand’
 Semelai | AA | Kruspe 2004: 151

Coda copy can turn an adjective into a noun, as in the following examples (again involving infixation of the coda):

- (5.109) *j<ŋ>lɔŋ*
 <RDP>to.be.long
 ‘length’
 Semelai | AA | Kruspe 2004: 223

- (5.110) *s<c>dəc*
 <RDP>to.be.cool
 ‘coolness’
 Semelai | AA | Kruspe 2004: 223
- (5.111) *r<l>mɔl*
 <RDP>to.be.male
 ‘male, man’
 Semelai | AA | Kruspe 2004: 223
- (5.112) *k<r>dɔr*
 <RDP>to.be.female
 ‘female, woman’
 Semelai | AA | Kruspe 2004: 223

These examples only scratch the surface of the complexities of Semelai morphological processes. See Kruspe (2004) for further details.

Let us now turn to the case of morphological processes in Vietnamese. This is worth doing for two reasons. First, thanks to Thompson’s excellent grammar of Vietnamese, we have an especially rich description of a single language system to draw on. Second, because Vietnamese has been cited since Sapir (1921) as a canonical isolating language, we can take the opportunity here to temper the myth that there is effectively nothing in the way of productive morphology in this language. As we are about to see, this is not the case.

Thompson (1987: 150ff.) gives a detailed account of what he calls *specializing derivatives* in Vietnamese. Above, I have described these sorts of derivations as involving reduplication, or partial reduplication, with possible stem alternation. Kruspe’s approach, in her 2004 description of Semelai, is to give formulae for constructing the appropriate output of these derivational patterns. Thompson conceives of the matter differently. He suggests that Vietnamese has ‘chameleon affixes’, which ‘have some kind of consistent phonetic resemblance to their bases’ (Thompson 1987: 139). Thompson’s idea is that a morpheme is added, but the phonological form of the morpheme is not fixed in advance; rather, it surfaces in the same form as its immediate environment, in the same way that a chameleon changes colour to match its surroundings. Here we consider four types of chameleon affix, laid out in Table 5.14.²

Thompson (1987: 150ff.) goes in detail through a vast number of patterns and regularities of formation of these ‘affixes’ or reduplicative structures. Let us now consider just a sample of the great variety of types of reduplication that he identifies.

² Thompson (1987: 139 and *passim*) also discusses a fifth type: vocalic affixes.

Table 5.14 *Four types of chameleon affix in Vietnamese (AA)*

Type	Definition
Perfect	'have exactly the same phonetic shape as the bases with which they occur'; i.e., full reduplication without stem alternation.
Tonal	'differ from their bases only in tone'; i.e., full reduplication, with alternation of the stem only in the tone
Riming	retain the rime of the stem; i.e., reduplication in which the initial of the stem is changed
Alliterative	retain the initial of the base; i.e., reduplication in which the rime (everything but the initial) is changed

After Thompson 1987: 139–40.

A form of ‘specializing derivative’ is productive in Vietnamese. Thompson distinguishes four types, all involving ‘perfect chameleon affixes’ (1987: 151–3). The first type is ‘distributive’. These are based on nouns. Here is an example:

- (5.113) *nhà~nhà*
 RDP~house
 ‘all the houses’
 Vietnamese | AA | Thompson 1987: 151

The second type is ‘iterative’. These are based on verbs, as in the following example:

- (5.114) *làm~làm*
 RDP~work
 ‘continually working, unfinished’
 Vietnamese | AA | Thompson 1987: 152

Third, there are ‘attenuatives’, based on adjectives:

- (5.115) *đẹp~đẹp*
 RDP~pretty
 ‘somewhat pretty’
 Vietnamese | AA | Thompson 1987: 152

Thompson’s fourth category is ‘intensives’. These are formed from a range of bases, as illustrated in the following examples:

- (5.116) *mau~mau*
 RDP~rapidly
 ‘very rapidly’
 Vietnamese | AA | Thompson 1987: 153

- (5.117) *vụn~vụn*
 RDP~in.small.pieces
 ‘in very small pieces’
 Vietnamese | AA | Thompson 1987: 153

- (5.118) *dời~đời*
 RDP~life,period.of.existence
 ‘eternity’
 Vietnamese | AA | Thompson 1987: 153

A set of ‘emphatics’ are formed ‘with extremely diverse affixes’ (1987: 154ff.). The following are six of the types that Thompson identifies (with the stem alternation in bold face, to indicate which part of the stem has been replaced with new material):

A. Emphatics with tonal affixes:

- (5.119) *bú~bu*
 RDP~be.big
 ‘be very big’
 Vietnamese | AA | Thompson 1987: 156

- (5.120) *xóp~xôp*
 be.spongy~RDP
 ‘be very spongy’
 Vietnamese | AA | Thompson 1987: 156

B. Emphatics with vocalic affixes; mostly ‘suffixing’ (i.e., infixing in the second syllable):

- (5.121) *hởc~hác*
 hole,hollow~RDP
 ‘be emaciated, gaunt’
 Vietnamese | AA | Thompson 1987: 157

- (5.122) *rõi rãi*
 be.unoccupied~RDP
 ‘have leisure time’
 Vietnamese | AA | Thompson 1987: 157

- (5.123) *vùng vẫng*
 shake~RDP
 ‘be shaking with anger’
 Vietnamese | AA | Thompson 1987: 157

- (5.124) *múp míp*
 be.fat~RDP
 ‘be fat, plump’
 Vietnamese | AA | Thompson 1987: 157

C. Emphatics with riming prefixes; most common are *b* and *l*:

- (5.125) *bàu-nhậu*
 RDP~be.wrinkled
 ‘be very wrinkled’
 Vietnamese | AA | Thompson 1987: 158

- (5.126) *bối-rối*
 RDP~be.confused
 ‘be troubled, perplexed, upset’
 Vietnamese | AA | Thompson 1987: 158

- (5.127) *hở-mò*
 RDP~groping.for
 ‘to grope feverishly’
 Vietnamese | AA | Thompson 1987: 158

- (5.128) *tò-mò*
 RDP~groping.for
 ‘to be curious, want to know all’
 Vietnamese | AA | Thompson 1987: 158

D. Emphatics with riming suffixes:

- (5.129) *tham-lam*
 be.greedy~RDP
 ‘be greedy, covetous’
 Vietnamese | AA | Thompson 1987: 159

- (5.130) *tạp-nhập*
 be.mixed,miscellaneous~RDP
 ‘be mixed’
 Vietnamese | AA | Thompson 1987: 159

- (5.131) *cười~nhuời*
 to.laugh~RDP
 ‘to tease, joke’
 Vietnamese | AA | Thompson 1987: 159

E. Emphatics with alliterative prefixes:

- (5.132) *rắc~rối*
 RDP~be.mixed.up,tangled
 ‘be complicated, intricate’
 Vietnamese | AA | Thompson 1987: 159

- (5.133) *lai~láng*
 RDP~abound
 ‘overflow’
 Vietnamese | AA | Thompson 1987: 159

- (5.134) *dẫn~đo*
 RDP~to.measure
 ‘weigh both sides of a question’
 Vietnamese | AA | Thompson 1987: 159

F. Emphatics with alliterative suffixes. These are ‘by far the most numerous and diverse’:

- (5.135) *nôm~na*
 demotic.script~RDP
 ‘popular language’
 Vietnamese | AA | Thompson 1987: 160

- (5.136) *xấu~xa*
 be.bad,ugly~RDP
 ‘be wicked’
 Vietnamese | AA | Thompson 1987: 160

- (5.137) *đất~dai*
 land~RDP
 ‘territory’
 Vietnamese | AA | Thompson 1987: 161

- (5.138) *rầm~rập*
 make.a.heavy.noise~RDP
 ‘be noisy’
 Vietnamese | AA | Thompson 1987: 163

Thompson provides over 300 examples, exemplifying dozens of patterns of stem alternating reduplication ('suffixes', in his terminology). In terms of their productivity, these run the gamut from relatively productive (e.g., *a* with more than twenty examples given) to ad hoc (many cases have only one or two forms, or where the 'source' morpheme has 'severely limited distribution'). Thompson notes some analytical problems (1987: 168), such as the occurrence of unanalysable cases (i.e., where the two syllable structure is itself a 'single morpheme'), the occurrence of compounds (where both syllables have independent occurrence with similar/related meanings e.g., *béo bór* 'be advantageous'; *béo* 'be fat', *bór* 'be profitable'), and so called phrases (i.e., where the two elements are from different parts of speech in a construction).

Thompson suggests that this system has its provenance in a historical stage of the language in which the morphology was more transparently productive:

It may be argued that at an earlier stage of the language there was an elaborate morphological system, involving many types of chameleon affixes with more or less consistent meanings. These may or may not have been related historically to a variety of onomatopoeic and other single morpheme forms of reduplicative type.

(Thompson 1987: 178)

While we have concentrated so far in this section on Austroasiatic languages, we can note that these types of stem alternating reduplications are also found in MSEA languages of other families that are often thought to lack productive morphological processes. An example is Lao. Table 5.15 shows examples of a form of expressive reduplication in Lao which takes a noun N and yields an expression meaning 'Ns and things associated with them'. The reduplicated syllable is postposed, with stem alternation in which a back vowel is changed into a front vowel at the same height. When the source stem does not feature a back

Table 5.15 Expressive reduplication in Lao (TK) with stem alternation, back vowel changed to front vowel

Original Form		Reduplicated Form
<i>cɔɔk5</i>	'drinking glass'	<i>cɔɔk5 ceek5</i>
<i>khuq1</i>	'bucket'	<i>khuq1 khitq1</i>
<i>pum4</i>	'book'	<i>pum4 pəm4</i>
<i>toq2</i>	'table'	<i>toq2 teq2</i>
<i>kɔɔng4</i>	'camera'	<i>kɔɔng4 keeng4</i>
<i>pɔɔ3</i>	'hessian'	<i>pɔɔ3 peɛ3</i>
<i>kua3</i>	'salt'	<i>kua3 kia3</i>
<i>hua4</i>	'fence'	<i>hua4 hia4</i>

Table 5.16 *Expressive reduplication in Lao (TK), front vowel changed to schwa*

Original Form		Reduplicated Form
<i>tip2</i>	‘rattan rice container’	<i>tip2 tɔp2</i>
<i>piik5</i>	‘wing’	<i>piik5 pɔək5</i>
<i>pet2</i>	‘duck’	<i>pet2 pat2</i>
<i>pheet4</i>	‘sex’	<i>pheet4 phəət4</i>
<i>sek1</i>	‘cheque’	<i>sek1 sək1</i>
<i>deek5</i>	‘eat (vulgar)’	<i>deek5 dək5</i>

vowel, a schwa may be used in the reduplicated syllable, as shown in Table 5.16. For further types of expressive reduplication in Lao, see Enfield (2007). .

5.5 Affixation

Many MSEA languages lack affixation in the usual sense of the term, as is expected given their isolating/analytic typological profile. Languages of the Austroasiatic language family are a notable exception. This exception is significant given that, as pointed out in Chapter 1, Austroasiatic languages make up nearly half of all languages found in the core MSEA area. In this section, I will draw mostly on the excellent overviews of Austroasiatic morphology presented by Jenny *et al.* (2015) and Alves (2015), to give a taste of the forms and functions of affixation in the area. Note that there is a general lack of suffixes, with the exception of Tibeto Burman languages on the western fringes of MSEA, along the Myanmar/India border (see DeLancey 2011), a fact that is presumably associated with the area’s predominance of an iambic word canon. Given a general correlation between phonological dependence and semantic dependence, an iambic word canon means that phonologically dependent morphemes will generally precede the morphemes they depend on.

Austroasiatic affixes are mostly prefixes and infixes. To different degrees across the Austroasiatic family, forms of affixation that appear to have been productive in ancestor languages lose their productivity in the course of historical development. At one end of the scale, morphology is clearly visible and productive in Aslian languages such as Semelai (Kruspe 2004). It is visible but less productive, and to an extent frozen, in languages including Mon and Khmer (Jenny *et al.* 2015: 46). And at the other end of the scale, it is hardly visible at all in languages including Vietnamese, Chong, and Bunong (Jenny *et al.* 2015: 45). Sagart (1999) suggests that Old Chinese ‘often assumed to be a classic isolating language’ was ‘typologically more like modern languages such as Khmer, Gyarong or Atayal, than like Middle and modern Chinese’. While in Middle and Modern Chinese, ‘only faint traces of the old morphology remain’, evidence ‘from

word families, modern dialects and related words in neighboring languages' suggests that a variety of derivational affixes were used in Old Chinese (Sagart 1999; see also Baxter and Sagart 1998).

In a minority of cases, MSEA affixation may be argued to be inflectional. This example from Bru shows a prefix signalling repetitive manner of action:

- (5.139) *si niəm*
 ASP to.weep
 ‘to keep on weeping’
 Bru | AA | Alves 2015: 539 | Hoàng and Tạ 1998: 7

But most forms of affixation have derivational functions. In this section, we survey a range of examples of affixation, illustrating some of the numerous functions and forms of affixation found in MSEA.

5.5.1 Existential/Locative Marking and Impersonal Predicate

In languages of the Bahnaric and Katuic sub branches of Austroasiatic, a nasal prefix marks existential/locative on pronouns. Here are two examples:

- (5.140) *ʔn naj*
 EXST there
 ‘this here is (something)’
 Bru | AA | Alves 2015: 542 | Hoang *et al.* 1986a: 39

- (5.141) *ʔn dɔ̄*
 LOC this
 ‘at this place’
 Koho | AA | Alves 2015: 542 | Hoang *et al.* 1986b: 53 4

5.5.2 Lexical Category Derivation

In many Austroasiatic languages, prefixes and infixes mark distinctions in lexical categories, for example deriving nouns from verbs or changing the valency of a verb (for example from intransitive to transitive). Here are cases of noun deriving infixation from three branches of Austroasiatic, most of which derive clearly from a Proto Austroasiatic <n> or <rn> infix (see, for example, Sidwell 2008: 542 3):

- (5.142) *t<ŋ>bɔh*
 <NMLZ>to.beat
 ‘act of beating’
 Jahai | AA | Alves 2015: 543 | Burenhult 2002: 72
- (5.143) *kh<an>é:t*
 <NMLZ>to.comb
 ‘comb’
 Kasong | AA | Alves 2015: 543 | Premsrirat 2011: 143
- (5.144) *k<rn>ooq*
 <NMLZ>to.dwell, remain
 ‘dwelling, house’
 Kri | AA | Enfield and Diffloth 2009: 47
- (5.145) *k<rn>eep*
 <NMLZ>to.pince
 ‘pincers, tongs’
 Kri | AA | Enfield and Diffloth 2009: 47
- (5.146) *p<hn>vas*
 <NMLZ>enter.holy.orders
 ‘holy orders’
 Old Khmer | AA | Jenny *et al.* 2015: 46
- (5.147) *k<hn>ɿ:əp*
 <NMLZ>squeeze, apply.pincers
 ‘pincers’
 Modern Khmer | AA | Jenny *et al.* 2015: 46
- (5.148) *katip* ‘cork’ → *kəntip* ‘cork, n.’
 Pacoh | AA | Jenny *et al.* 2015: 47
- (5.149) *cɪə* ‘dig’ → *həniə* ‘shove hoe’
 Sedang | AA | Jenny *et al.* 2015: 47
- (5.150) *blɔ* ‘to wear in the ear’ → *bənɔ* ‘earring’
 Koho | AA | Jenny *et al.* 2015: 47
- (5.151) *cʰ<ŋ>ret*
 <NMLZ>comb
 ‘comb, n.’
 Mlabri | AA | Jenny *et al.* 2015: 47

- (5.152) *ceh* ‘descend’ → *hneh* ‘pier’ (<sneh~cneh>)
 Mon | AA | Jenny *et al.* 2015: 47
- (5.153) *sɔ:l* ‘illuminate’ → *səmpɔ:l* ‘torch’
 Koho | Jenny *et al.* 2015: 48
- (5.154) *tɔ:m* ‘sing’ → *tr.nɔ:m* ‘song’
 Kammu | AA | Jenny *et al.* 2015: 48

Similar derivations are made by prefixes in some Austroasiatic languages, as shown in the following examples:

- (5.155) *rɔŋ key*
 NMLZ to.rest.one’s.head
 ‘pillow’
 Buxing | AA | Alves 2015: 544 | Gao 2004: 62
- (5.156) *k bay*
 NMLZ to.screen, to.shade/cover.sth
 ‘screen, movable curtain’
 Khmer | AA | Jenny *et al.* 2015: 46
- (5.157) *rñ tñn*
 NMLZ sit
 ‘chair’
 Kammu | AA | Jenny *et al.* 2015: 49
- (5.158) *?amínuvə* ‘possess’ → *ta?amí:nu* ‘owner’
 Car | AA | Jenny *et al.* 2015: 49

The last four examples were fairly transparent in terms of their semantics. The following three examples from Car show derived meanings that are related to the meanings of their source stems, but that are not predictable from them:

- (5.159) *ta viú:j*
 NMLZ be.hot
 ‘sun’
 Car | AA | Jenny *et al.* 2015: 49
- (5.160) *ŋé:?* ‘be hard’ → *ta ŋé:?* ‘bone’
 Car | AA | Jenny *et al.* 2015: 49

- (5.161) *ta ri:la*
 NMLZ be.crawling
 ‘animal, beast’
 Car | AA | Jenny *et al.* 2015: 49

In other MSEA languages, nominalization of this kind is achieved via compounding. For example, in Thai:

- (5.162) *thii juu*
 place to.dwell, remain, be.somewhere
 ‘address, place of living’
 Thai | TK

And here are two examples from Austroasiatic languages of affixation deriving a verb from a noun:

- (5.163) *və te:*
 VBLZ the.sound.of.one.calling.pigs
 ‘to call pigs’
 Koho | AA | Alves 2015: 544 | Hoang *et al.* 1986b: 54

- (5.164) *p<a>laajh*
 <VBLZ>an.arm.span
 ‘to measure something by arm spans’
 Kri | AA | Alves 2015: 544 | Enfield and Diffloth 2009: 46

5.5.3 Reciprocal/Collective Marking

In many Austroasiatic languages, a prefix marks a collective or reciprocal derivation on verbs, as shown in the following examples:

- (5.165) *ta ?waj*
 RCPL to.reside
 ‘to live together’
 Rengao | AA | Alves 2015: 549 | Gregerson 1979: 108

- (5.166) *tam pəm*
 RCPL to.hit
 ‘to fight’
 Chrav | AA | Alves 2015: 549 | Thomas 1971: 154

- (5.167) *kər ɻe:h*
 RCPL to.scold
 ‘to abuse each other’
 Palaung | AA | Alves 2015: 549 | Milne 1921: 52

- (5.168) *mə tə cuɻ dej pɻ*
 1DU RCPL obey with each.other
 ‘We two obey each other’
 Sedang | AA | Jenny *et al.* 2015: 53

The next example from Bru shows the same derivational meaning marked by an infix:

- (5.169) *sa<r>baw*
 <RCPL>to.call
 ‘to call each other’
 Bru | AA | Alves 2015: 549 | Hoàng and Tạ 1998: 82

In many other languages of MSEA, collective or reciprocal marking involves a stand alone marker in a dedicated syntactic construction. See section 7.4.3.

5.5.4 Causative

In many languages around the world, a causative form of a verb can be derived by a morphological process which changes the meaning of a clause from ‘X Vs’ (e.g., ‘X dies’ or ‘X eats rice’) to, roughly, ‘Y causes X to V’ (e.g., ‘Y kills X’ or ‘Y feeds X rice’). Causative affixes in MSEA languages mostly apply to intransitive verbs, but there are some exceptions (see examples 5.176, 5.177, 5.181 below).

In many Austroasiatic languages, a causative form is derived using a verbal prefix. The precise form of the prefix varies across the languages, but there are strong formal commonalities: an initial bilabial plosive or bilabial nasal is followed by a vowel, usually a low front or central vowel. Here are some examples:

- (5.170) *pa ɻem*
 CAUS to.sleep
 ‘to cause to sleep’
 Mlabri | AA | Alves 2015: 550 | Rischel 1995: 304, 357

- (5.171) *pa juu:n³⁵³*
 CAUS to.arise
 ‘to wake up’
 Thavung | AA | Alves 2015: 550 | Premsrirat 1998a: 197 → Premsrirat 1999: 107

- (5.172) *caj* ‘sick’ → *məjaj* ‘cause to be sick’
 Sedang | AA | Jenny *et al.* 2015: 51
- (5.173) *bə sɔŋ*
 CAUS straight
 ‘straighten’
 Koho | AA | Jenny *et al.* 2015: 51
- (5.174) *ma ho:c*
 CAUS to.die
 ‘to kill’
 Chong | AA | Alves 2015: 550 | Premsrirat 2011: 143
- (5.175) *pin jap*
 CAUS die
 ‘kill’
 Khasi | AA | Jenny *et al.* 2015: 51
- (5.176) *pń.máh*
 CAUS eat
 ‘feed’
 Kammu | AA | Jenny *et al.* 2015: 51
- (5.177) *pə loj*
 CAUS abandon
 ‘cause to abandon’
 Sedang | AA | Jenny *et al.* 2015: 51

In some Austroasiatic languages, the causative derivation involves an infix: for example, an infixed *m* in Kui, an infixed *u* in Old Mon, and an infixed *a* in Kri:

- (5.178) *tə<m>kpw*
 <CAUS>sit
 ‘cause to sit’
 Kui | AA | Jenny *et al.* 2015: 52
- (5.179) *g<u>lon*
 <CAUS>many, much
 ‘increase’
 Old Mon | AA | Jenny *et al.* 2015: 52

- (5.180) *s<ə>lô̂j*
 <CAUS>to.float.in.water
 ‘to float something on water’
 Kri | AA

- (5.181) *p<ə>raang*
 <CAUS>cross
 ‘take across’
 Kri | AA

In another kind of derivation, the following two examples show prefixes deriving intransitive verbs from transitive verbs, where the erstwhile undergoer O becomes the verb’s subject S:

- (5.182) *ha toh*
 INTR to.tear
 ‘to be torn’
 Katu | AA | Alves 2015: 543 | Nguyễn and Nguyễn 1998: 76

- (5.183) *gə̂ hə̂l*
 INTR to.cut
 ‘to be cut’
 Sre | AA | Alves 2015: 544 | Manley 1972: 46

- (5.184) *ti he?*
 ANTICAUS tear
 ‘torn’
 Pacoh | AA | Jenny *et al.* 2015: 52

Relatedly, Pacoh has a prefix that marks an otherwise volition verb as non volition or accidental:

- (5.185) *ta pəŋ*
 NVOLT shoot
 ‘shoot accidentally’
 Pacoh | AA | Jenny *et al.* 2015: 54

In many other MSEA languages, this valency alternation, in which the O argument of a transitive verb can be the S argument of an intransitive verb, is not explicitly marked at all. For example, in Lao:

- (5.186) *man2 tat2 cia4 lèèw4*
 3SG.B cut paper PRF
 ‘She has cut the paper.’
 Lao | TK

- (5.187) *cia4 tat2 lèèw4*
 paper cut PRF
 ‘The paper has been cut.’
 Lao | TK

While valency changing derivations are achieved by affixation in many Austroasiatic languages and thus in the lion’s share of MSEA languages in many other languages of the area, valency increasing derivations are done peripherastically, in ways that are comparable to the English strategy (e.g., *The cup broke* → *I caused the cup to break*, *He swept the house* → *I made him sweep the house*). MSEA languages do not, however, mark subordinate verbs as non finite, as in the English examples just given. Instead, MSEA languages realize this strategy via serial verb constructions (see Chapter 7, below).

5.5.5 Further Morphological Derivations

We close this section by noting some further functions of derivational morphology in Austroasiatic, in order to emphasize the breadth and array of morphology found in languages of the family that accounts for nearly half the languages in core MSEA.

Sedang features an ‘adversative’ prefix, as shown in the following example:

- (5.188) *?q saw ?eh lə kəde ?q*
 1SG fear 2SG ADVRS kill 1SG
 ‘I’m afraid you will kill me.’
 Sedang | AA | Jenny *et al.* 2015: 54

In Danau, a ‘gerundive’ prefix (glossed ‘NFIN’) forms an abstract noun from a verb:

- (5.189) *tə k̩i pin tʰəx bēj nì nə pə jɔx èkùtʰokæ̥e̥ enənej*
 COND go kill life other PROX TOP NFIN do sin same
 ‘If (you) go and kill another, it is the same as committing sin.’
 Danau | AA | Jenny *et al.* 2015: 54

Koho has a nasal prefix that is added to demonstratives and other nominals to derive interrogatives. The prefix consists of a nasal stop whose place of articulation agrees with that of the stem initial consonant. Here are two examples:

- (5.190) *m be*
 INTRG that, as
 ‘how’
 Koho | AA | Jenny *et al.* 2015: 55

- (5.191) *n caw*
 INTRG person
 ‘who, which’
 Koho | AA | Jenny *et al.* 2015: 55

Similarly, in Mon, prefixation on demonstrative/interrogative stems derives pronouns and locative or manner adverbs. Here are some examples (illustrating accompanying morphophonological change in the prefix):

- (5.192) *nɔ?* ‘this’ → *?i? nɔ? ~ nɔ?* ‘this one (pronoun)’
 → *?ə nɔ?* ‘here (locative adverb)’
 → *hə nɔ?* ‘like this (manner adverb)’
 Mon | AA | Jenny *et al.* 2015: 56

In Sedang, prefixation on numerals can derive measure words (using the prefix *tə*) and ordinals (using the prefix *mə*), as shown in these examples:

- (5.193) *paj* ‘three’ → *təpaj* ‘three fingers’ width’
 Sedang | AA | Jenny *et al.* 2015: 56
- (5.194) *paj* ‘three’ → *məpaj* ‘third’
 Sedang | AA | Jenny *et al.* 2015: 56

Finally, in Bugan, a number of lexical pairs suggest a range of lexical semantic alternations corresponding to alternations in vowel, tone, and initial consonant alternation, as illustrated in the following three examples, respectively:

- (5.195) *ŋɔ³³* cf. *ŋa³³*
 ‘peppery’ ‘salted’
 Bugan | AA | Jenny *et al.* 2015: 56
- (5.196) *tsan²⁴* cf. *tsan³³*
 ‘to smell’ ‘smelly’
 Bugan | AA | Jenny *et al.* 2015: 56
- (5.197) *nap⁵⁵* cf. *ŋap⁵⁵*
 ‘be closed’ ‘to close one’s eyes, to shut up’
 Bugan | AA | Jenny *et al.* 2015: 56

5.6 Tone in Word Formation

It is not uncommon for speakers of tone languages to use pitch and other features of tonation in morphological and syntactic processes. This is widely observed in African languages. However, descriptions of MSEA languages have not widely reported functions of tone in morphology and syntax. This may mean that there is not much to observe. Or it may mean that with further work we will discover some patterns and functions that have not yet been noticed. Some existing work is suggestive of this second possibility. For example, we saw in the description of reduplication ('chameleon affixation') in Vietnamese (see section 5.4) that tones are implicated in morphological processes: they can be targeted in reduplicative stem alternations.

The most important work in this area relating to MSEA languages is Ratliff (1992), showing the ways in which tone in White Hmong can play meaningful roles in the language beyond simply serving to mark contrast between lexical items. Three of these roles of tone are: (1) tone sandhi in compound formation, (2) defining lexical classes, and (3) as a component in the morphology of ideophones.

5.6.1 Tone Sandhi in Compound Formation

When compounds are formed, often there are phonetic/phonological markers of the relationship between the elements that are compounded, as noted for the role of stress in English compounds. In Hmong, the tone of an element in the compound can change due to sandhi rules. Ratliff notes that the sandhi form of a word, with its altered tone, can become dissociated from the trigger of the sandhi, and can become an independent word with a distinct tone from its original tone. The historical (Proto Hmongic) and the sandhi forms may coexist with slight meaning differences, as shown in Table 5.17.

Table 5.17 *Doublets in White Hmong (HM) arising from Sandhi effects*

Original (when stand-alone)		Sandhi-affected (when in compound)	
<i>caj</i>	root [edible]	<i>cag</i>	root [inedible]
<i>hauv</i>	source; base; summit	<i>hau</i>	head

After Ratliff 1992: 78–9.

Some other MSEA languages show intriguing pairs of words with related meanings, differing only in tone. It is possible that these could have emerged from similar historical processes.

5.6.2 Regular Grammatical Distinctions

There are cases in Hmong where tone appears to act as if it were an affix that marks regular distinctions in grammatical meaning or form class. In West Hmongic (of which White Hmong is a member) and East Hmongic branches of Hmong Mien, the distinction between dual and plural in the second person pronouns is made solely by tone: tones that developed from the Proto Hmong Mien A1 category signal ‘dual’, those from A2 signal ‘plural’ (see Ratliff 1992: 22). Here are the examples in White Hmong:

- (5.198) 2DU 2PL
neb (A1) *nej* (A2)
 White Hmong | HM | Ratliff 1992: 99 (Heimbach 1979, Bertrais Charrier 1964, Mottin 1978)

White Hmong has a set of spatio temporal words, almost all of which have the *m* tone (low falling, checked). They are used as prepositions, and appear to be cognate with, and/or derived from, corresponding spatio temporal nouns. Ratliff (1992) refers to them as denominational prepositions.³ Here are some examples:

- (5.199) *nras* → *nram*
 ‘plain’ ‘down’
 White Hmong | HM | Ratliff 1992: 105
- (5.200) *nrau* → *nraum*
 ‘place beyond’ ‘outside’
 White Hmong | HM | Ratliff 1992: 105
- (5.201) *zaus* → *zaum*
 ‘time’ ‘time when’
 White Hmong | HM | Ratliff 1992: 106

³ Although Ratliff calls them ‘denominational prepositions’ here, in earlier work (Ratliff 1990) she calls them ‘spatial deictics’, because they do not actually function as prepositions. For discussion, see Jarkey (2015: 40–3).

- (5.202) *thaum* ‘time when’
 White Hmong | HM | Ratliff 1992: 106

In addition, a change from the *m* tone of the above denominal prepositions to a *d* tone derives demonstratives (Ratliff 1992: 12). Here are some examples:

- (5.203) *nyob nrad*
 be down
 ‘down there’
 White Hmong | HM | Ratliff 1992: 113 (cf. 5.199)
- (5.204) *nyob sab nraud*
 be side outside
 ‘outside; on the other side’
 White Hmong | HM | Ratliff 1992: 113 (cf. 5.200)
- (5.205) *paug txheej thaud*
 many generation time
 ‘in the old days’
 White Hmong | HM | Ratliff 1992: 114 (cf. 5.202)

A number of male/female noun pairs are distinguished by tone alone, where the *g* tone (low falling, breathy) occurs in masculine nouns, and the *m* tone (low falling, checked) occurs in feminine nouns. Here are some examples:

- (5.206) *ntxawm* ‘youngest daughter’
 White Hmong | HM | Ratliff 1992: 125
- (5.207) *ntxawg* ‘youngest son’
 White Hmong | HM | Ratliff 1992: 125
- (5.208) *poj ntsuam* ‘widow’
 White Hmong | HM | Ratliff 1992: 126
- (5.209) *ntsuang* ‘widow(er); orphan’
 White Hmong | HM | Ratliff 1992: 126
- (5.210) *muam* ‘sister (man speaking); female peer relation’
 White Hmong | HM | Ratliff 1992: 126
- (5.211) *nraug* ‘young unmarried man’
 White Hmong | HM | Ratliff 1992: 126

5.6.3 Forming Ideophones

In her detailed description and analysis of the system of ideophones in White Hmong, Ratliff (1992) describes a range of ways in which the ideophones of the language many of which are disyllabic feature alternations between the two syllables not just in consonants and vowels, but also in tones. For example, an ideophone *duj dig* is derived from a stem *dig* ‘to be blind’. The ideophone is formed by prefixing a copy of the stem, changing the vowel to *u*, and changing the tone from *g* (mid falling, breathy) to *j* (high falling). Here are some examples of the ideophone in its grammatical context:

- (5.212) *ua duj dig*
 make IDEO blind
 ‘to feel one’s way with feet and hands like a blind person’
 White Hmong | HM | Ratliff 1992: 143

- (5.213) *maub duj dig*
 grope IDEO blind
 ‘to go along, feeling one’s way like a blind person’
 White Hmong | HM | Ratliff 1992: 143 | B *dig*

Another example involves the stem *teev*, a noun meaning ‘a drop (of liquid)’. Again, the ideophone is formed by prefixing a copy of the stem, changing the vowel to *u*, and changing the tone from *v* (mid rising) to *j* (high falling):

- (5.214) *dej nrog tuj teev*
 water drip IDEO drop
 ‘the water drips drop by drop’
 White Hmong | HM | Ratliff 1992: 144 | B *teev*

A final example is based on the syllable *qees*, a one syllable ideophone which, when used in combination with the verb *hais* ‘speak’, has the meaning ‘continue talking’:

- (5.215) *hais qees*
 speak IDEO
 ‘to continue talking’
 White Hmong | HM | Ratliff 1992: 146

The same derivation just described for the adjective *dig* ‘blind’ and the noun *teev* ‘drop’ can also be applied to the one syllable ideophone *qees* as stem:

- (5.216) *hais quj qees*
speak IDEO IDEO
'to speak loudly and deliberately'
White Hmong | HM | Ratliff 1992: 146

These examples only scratch the surface of the workings of a highly complex and nuanced system. For a richly detailed description, both synchronic and diachronic, see Ratliff (1992).

Reference and Nominal Syntax

All languages provide ways for speakers to refer to the people, places, and things that they want to talk about. These resources include not only nouns and other nominal lexical forms, but also strategies for productively creating more specific reference, using strategies such as simple modifiers (adjectives or equivalent; e.g., ‘small house’), relative clauses (e.g., ‘the house I built last year’), and possessive marking (e.g., ‘John’s house’). In section 6.1, we survey the main strategies found in MSEA for these three basic ways of modifying nouns in noun phrases.

Relatedly, all languages provide ways for speakers to quantify the things they are referring to (e.g., ‘three houses’). MSEA languages rely widely on forms of nominal classification for this and related functions (surveyed in section 6.2). And all languages will also provide highly generalized forms of nominal reference for those situations in which a referent is already known (pronouns; e.g., ‘it’) and where the physical or discourse context provides the information needed for identifying the referent (demonstratives; e.g., ‘*that* house’). Pronoun systems and demonstrative systems are discussed in sections 6.3 and 6.4.

We note here that zero anaphora is an area wide strategy for dealing with reference to noun phrases that have an established identity in discourse. A few examples illustrate the phenomenon (which is also illustrated in many other examples throughout this book). In the following cases, from Saek (TK), Kri (AA), Cham (AN), Iu Mien (HM), Mandarin (SN), and Kucong (TB) respectively, entities are first introduced into discourse with noun phrases and are then omitted from subsequent argument positions where their reference may be understood from context. The omitted references are functionally equivalent to pronouns in English and many other languages. In these examples, initial nominal mentions are given in square brackets and the subsequent zero anaphors are marked as ‘Ø’ and subscripted for coreference accordingly:

- (6.1) Ø_i thang4 Ø_j lè Ø_k bo phòò4 lè Ø_l tam1 Ø_k loo5
 arrive and NEG enough and pound more
 ‘(When you_i) arrived at (the house_j) and (if tea_k) wasn’t enough, (they_l)
 pounded more (tea_k)’

Saek | TK | Weijian Meng field record 20190513 08.216

- (6.2) \emptyset_i tzrôôh [suù] J_j tzrôôh [kdiij] J_j tzrôôh [kaùrq] J_j \emptyset_i piñq \emptyset_j
 meet game meet deer meet pig shoot
 ‘ \emptyset_i (We) encountered [game J_j], \emptyset_i encountered [deer] J_j , \emptyset_i encountered [pigs] J_j and \emptyset_i shot \emptyset_j .’

Kri | AA | Lutz Hughes 2016: 36

- (6.3) [$t^haj^{21}kaj^{33}$] η aw p^{24} pu $^{33}ma^{24}$ sien 21
 v.brother make NEG.have money
 maj $^{21}zaj^{33}$ \emptyset_i na:w p^{33} tsə p^{24} $t^huən^{21}$ $p^{iu^{55}}$ p $^hiaj^{21}$ \emptyset_j
 every day go mountain gather firewood sell
 ‘[The younger brother] J_i ; hadn’t been able to make money. Every day \emptyset_i went to the mountains, gathered [firewood] J_j and sold \emptyset_j .’

Cham | AN | Thurgood *et al.* 2014: 270 1

- (6.4) yie mbuo Iu Mienh naaic ninh mbuo gorngv [ninh mbuo] $_i$ yiem
 1 PL Iu Mien TOP 3 PL say 3 PL be.in
 Naamh Ging koiv ngaanc
 Nanking sea shore
 Yiem Naamh Ging koiv ngaanc se dorngx longx haic
 be.in Nanking sea shore TOP place be.good very
 maaih dorngx zoux liangx ndeic camv yaac \emptyset_i nquien youh
 have place make field:for:farming many and happy
 \emptyset_i mv zuqc zaangc mienv
 need.not venerate spirit

‘As for we Iu Mien (people), they say [we] $_i$ used to live in the coastal areas of Nanking. In the coastal area of Nanking the place was very good. There were many places to make farming fields and \emptyset_i were happy, and \emptyset_i did not have to venerate spirits.’

Iu Mien | HM | Arisawa 2016: 833

- (6.5) [$wō$] $_i$ shàng xīngqi gěi [Dàhuá Fàndiàn] $_j$ xiě le yi fēng
 I last week to Dahuá Hotel write PFV one CL
 $xìn$, \emptyset_i qǐng [$tāmen$] $_j$ bǎoliú yí jiān
 letter invite they reserve one CL
 $wūzi$, dào xiànzài \emptyset_j hái méi tōngzhī wǒ.
 room to now still not inform I
 ‘Last week [I] $_i$ wrote a letter to the [Dahuá Hotel] J_j , and \emptyset_i requested (that) they reserve a room. Up to now, \emptyset_j still haven’t informed me.’

Mandarin | SN | Li and Thompson 1981: 661

- (6.6) [*a⁵⁵ηaŋ³³*]_i *tshy³¹* *tɔ³³* *la³¹* *vɛ³³* *tsha³¹**tsj³³* *lɔ³³ηɛ³³* Ø_i
 konjac wash exit come LNK grater INSTR
tsha³¹ *ɛi⁵⁵.* *xie³³*
 grate fine PTC
 ‘First, wash the [konjac]_i, then with a grater grate [it]_i finely.’

Kucong | TB | Chang 2009: 279

Note that Kucong has a different basic clausal constituent order from the other languages cited here. The Kucong basic order is verb final, like most Tibeto Burman languages, while the languages shown in examples (6.1–6.5) have verb medial order.

6.1 Nominal Modification

As previewed in section 3.5, both logical orderings of head and modifier—head initial and head final—are found in MSEA. Sino Tibetan languages tend to be strongly head final in noun phrase structure, while Austroasiatic languages tend to be strongly head initial. Other language families of the area show a degree of mixed orientation, with some variation within each family (though with Tai and Hmong Mien languages being mostly head initial). When a language shows a mix of strategies, this is often attributed to language contact effects.

6.1.1 Simple Head Modifier Relations

Simple head modifier constructions may feature an adjective or adjective-like verb as modifier of a head noun, with no overt marking of the relation between the two phrasal constituents. Here are some examples with head initial structure:

- (6.7) *knkɔn kɛt*
 child be.small
 ‘a little child’
 Semelai | AA | Kruspe 2004: 203

- (6.8) *iĕr măq*
 chicken large
 ‘a large chicken’
 Chravu | AA | Thomas 1971: 139

- (6.9) *nja¹ lo⁴*
 river big
 ‘a big river’
 Mulao | TK | Wang and Zheng 1993: 66

- (6.10) *ci² la:n³*
 flag red
 ‘a red flag’
 Mulao | TK | Wang and Zheng 1993: 66
- (6.11) *me nyuam yaus*
 child young
 ‘young children’
 White Hmong | HM | Jarkey 2015: 31
- (6.12) *tsiaj quas*
 animal wild
 ‘wild animals’
 White Hmong | HM | Jarkey 2015: 32
- (6.13) *ə⁴⁴ dzəŋ³⁵*
 clothes red
 ‘red clothes.’
 Aizhai Miao | HM | Yu 2010: 267
- (6.14) *le³⁵ ta⁴⁴*
 rice hard
 ‘hard steamed rice’
 Aizhai Miao | HM | Yu 2010: 267

The languages just cited also show head initial order when the modifier is a demonstrative (such as ‘this’ or ‘yon’). Again, the modifier does not take any overt marker linking it to the head:

- (6.15) *knkɔn ke*
 offspring that
 ‘that child’
 Semelai | AA | Kruspe 2004: 211
- (6.16) *iér heq*
 chicken this
 ‘this chicken’
 Chrau | AA | Thomas 1971: 139
- (6.17) *ti⁶fɔŋ¹ na:i⁶*
 place this
 ‘this place, here’
 Mulao | TK | Wang and Zheng 1993: 66

- (6.18) *tus neeg no*
 CLF person this
 ‘this person’
 White Hmong | HM | Jarkey 2015: 38

- (6.19) *a^{44/21} pzui⁴⁴ nəŋ⁴⁴*
 one family this
 ‘this family’
 Aizhai Miao | HM | Yu 2010: 140

By contrast with languages of the families just illustrated, speakers of Sinitic languages position modifiers before their heads:

- (6.20) *tiānrán yánsè*
 natural colour
 ‘natural colours’
 Mandarin | SN | Li and Thompson 1981: 117

- (6.21) *hóng huā*
 red flower
 ‘a red flower’
 Mandarin | SN | Li and Thompson 1981: 118

When demonstratives are used as modifiers of head nouns in Sinitic languages, a linking marker – usually a classifier appropriate to the noun being modified – typically appears between the demonstrative and the head noun:

- (6.22) *gəɔ̄ pɔŋ^[1 3] sɿ^[51]*
 this CLF book
 ‘this book’
 Shanghaiese | SN | Zee and Xu 2003: 138

- (6.23) *chit tiāu lō'*
 this CLF road
 ‘this road’
 Taiwanese | Sinitic | Lin 2015: 122

The linker in this kind of construction need not be a classifier. A dedicated linking particle *de* (in Mandarin) is often used between non demonstrative simple modifiers and head nouns in Sinitic languages. For example:

- (6.24) *bái sè de qiáng*
 white colour PTC wall
 ‘white wall’
 Mandarin | SN

This is mimicked by non Sinitic languages whose speakers have been in intense contact with Sinitic language speakers. An example is Thailand Mien, whose speakers are recent arrivals in Thailand, having arrived within the last two centuries from China. In Thailand Mien, a Sinitic linking particle *tei*²⁴ links the preposed demonstrative *na:i*⁵⁵ ‘this’ to the nominal head *kwa*³³ ‘melon’:

- (6.25) *na:i*⁵⁵ *tei*²⁴ *kwa*³³ *na:i*³¹ *m*⁵⁵ *khu*⁵⁵
 these PCL melon PTC NEG tasty
 ‘These melons, they are not tasty.’
 Thailand Mien | HM | Liu 2012: 253

Head initial languages of MSEA tend not to include this kind of linker, which explicitly connects a modifier to its head, but there is a pattern suggestive of this in Jahai. The use of a third person singular pronoun immediately after a head noun acts as a kind of definite article in Jahai:

- (6.26) *ʔap ʔo?*
 tiger 3SG
 ‘the tiger’
 Jahai | AA | Burenhult 2005: 142

When a demonstrative, such as *ton* ‘that’, is used, the pronoun/article is retained, linking the demonstrative modifier to its head:

- (6.27) *mawẽ? ʔo? ton*
 gibbon 3SG that
 ‘that gibbon’
 Jahai | AA | Burenhult 2005: 142

Sometimes a language appears to allow both orders of head and modifier in simple noun phrases: head initial and head final. In Mulao, for example, the term for ‘saddle’ (for horses) consists of a head noun *a:n'* ‘saddle’ and a modifier *ma*⁴ ‘horse’, but these two elements can be ordered in both ways, head initial and head final:

- (6.28) *a:n' ma*⁴
 saddle horse
 ‘saddle’
 Mulao | TK | Wang and Zheng 1993: 32

- (6.29) *ma⁴ a:n¹*
 horse saddle
 ‘saddle’
 Mulao | TK | Wang and Zheng 1993: 32

The first example shows the native Tai head initial ordering, as observed, for example, in the Lao term for saddle: *qaan1 maa4* [saddle horse]. The second example shows the Sinitic head final ordering, as observed in Mandarin: *mǎ ’ān* [horse saddle]. It seems likely that the term with Sinitic ordering is adopted as a unit, and need not be taken as evidence that Mulao has a productive pattern of head final ordering in simple noun phrases.

Here is another case of a noun phrase likely calqued as a unit into Mulao. As in many other MSEA languages, the term for ‘train’ in Mulao literally means ‘fire vehicle’. Notice that the ordering of head and modifier in the Mulao expression follows the Sinitic pattern (cf. Mandarin *huǒ chē* [fire vehicle]), while the morphemes used are a Sinitic borrowing (*tshja¹*; cf. Mandarin *chē*) and a native Tai word (*f¹*; cf. Thai *fai* ‘fire’):

- (6.30) *f¹ tshja¹*
 fire vehicle
 ‘train’
 Mulao | TK | Wang and Zheng 1993: 31

Finally, there are cases in Mulao in which two orderings occur within a single complex construction. In the next two examples, a Sinitic phrase, in which the modifier precedes the head, is used as a postposed modifier of a native Tai head:

- (6.31) *fa⁵ tɔŋ¹ jep⁸*
 leaf dumpling leaf
 ‘reed leaves for making pyramid shaped glutinous rice dumpling’
 Mulao | TK | Wang and Zheng 1993: 31
 (cf. Mandarin *zòng yè*)

- (6.32) *pya:n⁶ tshe⁵ sjen⁴*
 thread machine thread
 ‘thread used for sewing machine’
 Mulao | TK | Wang and Zheng 1993: 31
 (cf. Mandarin *jī xiàn*)

In the Mulao case, alternative orderings of head and modifier do not convey differences in underlying meaning. Rather, they index different sources of structure

in the language: a native structure, inherited from ancestor languages, versus a borrowed structure.

In a different kind of case, a language may show alternative orderings of head and modifier, where this ordering either conveys a meaning difference or is actually not what it seems. In Eastern Kayah Li, two orders of nominal head and modifier are observed (Solnit 1997: 187–8). When a nominal (e.g., the noun ‘chicken’) acts as a modifier of a noun (e.g., ‘egg’), it appears before the noun it modifies. Here is an example, with ‘chicken’ modifying ‘egg’:

- (6.33) *chā də*
 chicken egg
 ‘chicken egg’
 Eastern Kayah Li | TB | Solnit 1997: 188

By contrast, when a verb or adjective modifies a noun, it appears *after* the noun it modifies:

- (6.34) *rì du*
 snake big
 ‘big snake’
 Eastern Kayah Li | TB | Solnit 1997: 187

These two structures can be combined in Eastern Kayah Li, such that a nominal head may have a nominal modifier occurring before it, and a verbal/adjectival modifier occurring after it:

- (6.35) *phremɔ́ hʌ je*
 woman skirt tattered
 ‘tattered woman’s skirt’
 Eastern Kayah Li | TB | Solnit 1997: 187

Another kind of case is illustrated in the following two examples from Lao, which show what appear to be opposite head modifier orderings, with only slightly different meanings:

- (6.36) *phak2 tom4*
 greens boil
 ‘boiled greens’
 Lao | TK

- (6.37) *tom⁴ phak²*
 boil greens
 ‘(a dish made from) boiled greens’
 Lao | TK

These examples might appear to suggest that both head initial and head final structures are possible in Lao noun phrases. But Lao is strictly head initial in its noun phrase structure. In fact, the second example shows the usual order. What looks in example (6.37) to be an adjectival/verbal modifier *tom⁴*, glossed as the verb ‘to boil’ is actually the head of the phrase. It is one of a set of verbs of food preparation that can also be used as nouns referring to dishes that are prepared using that action/process (see Enfield 2006b). So, the appearance that the language allows both head initial and head final structures is illusory. Both examples (6.36) and (6.37) are head initial.

In Cantonese, Matthews and Yip (1994: 49) note that ‘some compound nouns like food and animal terms have their head nouns in the initial position, which contrasts with the usual head final ordering found throughout Sinitic’. The examples of apparently head initial ordering in Cantonese nominal structures suggest that the pattern is only found in specified semantic domains. A first set of examples relate to foods (as in the Lao case just discussed):

- (6.38) *yùh sāang*
 fish raw
 ‘raw fish’
 Cantonese | SN | Matthews and Yip 1994: 49

- (6.39) *choi gōn*
 vegetable dry
 ‘dry vegetable’
 Cantonese | SN | Matthews and Yip 1994: 49

- (6.40) *ngàuh yuhk gōn*
 cow meat dry
 ‘beef jerky’
 Cantonese | SN | Matthews and Yip 1994: 49

A second set of examples relates to domestic animals, where the sex of the animal in question is expressed in a postposed modifier rather than the preposed modifier that is expected in a Sinitic language (compare Mandarin *gōng jī* for ‘rooster’):

- (6.41) *gāi gūng*
 chicken male
 ‘rooster’
 Cantonese | SN | Matthews and Yip 1994: 49
- (6.42) *gāi lá*
 chicken female
 ‘chicken’
 Cantonese | SN | Matthews and Yip 1994: 49
- (6.43) *jyū gūng*
 pig male
 ‘male pig’
 Cantonese | SN | Matthews and Yip 1994: 49
- (6.44) *jyū lá*
 pig female
 ‘female pig’
 Cantonese | SN | Matthews and Yip 1994: 49

6.1.2 Relative Clause Constructions

Relative clauses allow speakers to modify head nouns in more complex and specific ways than by adding simple modifiers (described in the last section). A relative clause is a complex modifier that is derived from a full clause or sentence. To illustrate, suppose that we planted two fields of rice: one is upstream, the other is downstream. Now suppose I want to refer to just the first one. When I say ‘rice that we planted upstream’, this structure is a noun phrase, consisting of a head noun ‘rice’ modified by a relative clause. One way of describing the modifying element ‘that we planted upstream’ is to say that it is derived from the sentence ‘We planted rice upstream’: the target referent element is moved (‘rice’ becomes head of the new noun phrase), and a relativizer ‘that’ is added, linking the head to the modifying phrase.

Most if not all languages provide ways for their speakers to use relative clause constructions for modifying head nouns in noun phrases. Languages will differ in a number of ways as to how this is done, including (1) the relative order of head and modifier in the phrase and (2) the presence or absence of an explicit relativizing particle. (Note that in many languages the use of an explicit relativizing particle is

optional, as it is in English: compare ‘rice that we planted upstream’ and ‘rice we planted upstream’, both of which are fine.)¹

In most languages of MSEA – with the exception of Sinitic languages, and languages that have been in intensive contact with them – relative clauses have the properties listed in Table 6.1.

Table 6.1 *Dominant properties of relativization in non-Sinitic MSEA languages*

-
1. The relative clause appears after its head.
 2. A marker of relativization may be used but is often optional, in which case the relative clause is simply used directly as a modifier.
 3. The set of grammatical roles played by the nominal which is modified by the relative clause is somewhat unrestricted.
 4. The head is positioned externally to the relative clause.
-

Data from White Hmong can be used to illustrate these points. In White Hmong, relative clauses are typically introduced by the relativizer *uas*. Here is an example:²

- (6.45) ... *ces nws mus nrog tus thab.han uas zov ntug dej ntawm*
 then 3SG go with CLF soldier that watch edge water there
 tham
 talk
 ‘... then he went to talk with a soldier who was watching the riverbank.’
 White Hmong | HM | example from Fuller 1985: 230, cited by Riddle
 1993: 58

In this example, the nominal head refers to a soldier. The basis for the construction is something like ‘The soldier was watching the riverbank’. Note that the soldier’s

¹ A distinction is sometimes made between ‘restricted’ and ‘non-restricted’ relative clauses. Restricted relative clauses serve to identify a sub-group of the modified noun’s referent. For example, in the phrase *mushrooms which are safe to eat*, the relative clause *which are safe to eat* picks out a restricted subset of all mushrooms, implying that not all mushrooms are safe. In the phrase *field mushrooms, which are safe to eat* (where an all-important pause is marked here by a comma), the relative clause does not pick out a subset of the referent of *field mushrooms*, but it serves to add extra information about them; namely, that they are (all) safe to eat. Many of the world’s languages allow these two readings of a relative clause construction, without any specific marking of the difference.

² In the examples throughout the rest of this section, the head of the relative clause is marked with underlining.

grammatical role in this construction is transitive subject, the argument in A function in the clause. The verb ‘watching’ has two arguments, A being the actor, and O being the undergoer.

In many languages, it is also possible to relativize on the transitive *object*, the argument in O function, as in ‘the riverbank that the soldier was watching’. In examining patterns of accessibility to relativization in languages of the world, linguists make reference to an *accessibility hierarchy*, identified by Keenan and Comrie (1977), as shown in Figure 6.1.

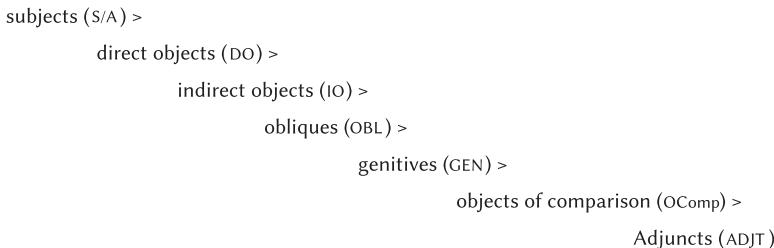


Figure 6.1 *Accessibility hierarchy for the construction of relative clauses in languages of the world*
After Keenan and Comrie 1977. The hierarchy predicts that if speakers of a language can form a relative clause that modifies an argument of any type on this hierarchy, then it will allow the same for any argument above that type.

The generalization represented by the hierarchy in Figure 6.1 is that if a language can relativize on a nominal in one of these roles, it can relativize on nominals with any role higher in the hierarchy. MSEA languages are liberal in this regard, allowing relativization on virtually any kind of argument in a source clause or sentence.

For example, in White Hmong, the relativized argument can be of any type on the accessibility hierarchy, including adjuncts (Riddle 1993). Here are examples showing relativization on arguments running all the way down the hierarchy from A (illustrated in the previous example) to adjuncts:

(6.46) Direct object relativized:

Ib gho uas tus kwv tij Hmoob Thaib tau piav...
 one thing that CLF kinsman Hmong Thai ATT tell
 ‘one thing that the Hmong kinsman from Thailand told [us]’
 White Hmong | HM | example from Neng 1987: 37, cited by Riddle
 1993: 58

- (6.47) Indirect object relativized:

Tus poj.niam uas kuv muab cov txhuv rau...
 CLF woman that 1SG give GRP rice to
 ‘the woman that I gave rice to’
 White Hmong | HM | Riddle 1993: 58

- (6.48) Oblique relativized:

...tus neeg uas nej yuav nrog sib tham
 CLF person that 2PL will with RECIP talk
 ‘the person with whom you want to talk’

White Hmong | HM | example from Thoj 1981: 31, cited by Riddle 1993: 59

- (6.49) Possessor relativized:

...cov hluas uas niam thiab txiv txom.nyem...
 GRP young that mother and father poor
 ‘young people whose parents are poor ...’

White Hmong | HM | example from Thoj 1981: 62, cited by Riddle 1993: 59

- (6.50) Object of comparison relativized:

tus txiv.neeg uas kuv siab dua
 CLF man that 1SG tall more.than
 ‘The man that I am taller than’
 White Hmong | HM | Riddle 1993: 59

- (6.51) Adjunct relativized:

...nyog ntawm ntug dej uas hla rau sab thaib.teb yooj.yim
 be at at edge water that cross to side Thailand easy
 heev very

‘... at the edge of the river where it’s very easy to cross to the Thai side’
 White Hmong | HM | example from Fuller 1985: 228, cited by Riddle 1993: 59

White Hmong allows a resumptive pronoun to be used to indicate the argument that has been relativized upon (compare the constraint against doing this in English: we can say *the woman who spoke Hmong* but not **the woman who she spoke Hmong*):

- (6.52) *Yog tus Xib.Hwb uas nws nyob X*
 be CLF teacher that 3SG live X
 ‘It’s the teacher that lives at X.’ (literally: ‘It’s the teacher that he lives at X.’)
 White Hmong | HM | example from Mottin 1978: 139, cited by Riddle 1993: 60

Moving now to an example from an Austroasiatic language, in Jahai, relative clause modifiers ‘occur at the right hand periphery of NPs’ and ‘are introduced by means of the relative marking proclitic *k* , which is attached to the first constituent of the relativised element’ (Burenhult 2005: 138). In the following examples, we see different elements relativized, including subjects, objects, and indirect objects:

- (6.53) S relativized:
bdil k pcah
 gun REL to.break
 ‘the gun that broke’
 Jahai | AA | Burenhult 2005: 139
- (6.54) A relativized:
hawēn k gej hobī?
 pig REL to.eat root.crop
 ‘the pig that ate the root crop’
 Jahai | AA | Burenhult 2005: 139
- (6.55) DO relativized:
slaj k wih rh rɔh
 swidden REL 3D IMPF to.clear
 ‘the swidden that they were clearing’
 Jahai | AA | Burenhult 2005: 138
- (6.56) IO relativized:
tmkal k je? ɻek tmakɔw
 man REL 1S to.give tobacco
 ‘the man whom I gave tobacco’
 Jahai | AA | Burenhult 2005: 139

Semelai, closely related to Jahai, shows a similar structure in relative clause constructions. In Semelai, the relativizer is the proclitic *mə* , which ‘is often used to name a category’ (Kruspe 2004: 201), as shown here:

- (6.57) *mə̄ gdo*
 REL be.old
 ‘the elderly’
 Semelai | AA | Kruspe 2004: 211

The following examples show *mə̄* used as a relativizer:

- (6.58) *tuŋwan mə̄ pŋt̪ haŋ dɔl*
 grandmother REL stay AT house
 ‘The grandmother who stayed at home.’
 Semelai | AA | Kruspe 2004: 211
- (6.59) *knkɔn ke, mə̄ tʰɔy ke*
 offspring that REL be.big that
 ‘that child, that (one) who is big’
 Semelai | AA | Kruspe 2004: 211

In Sinitic languages, by contrast with the examples we have seen above, relative clauses appear before their head nouns. Sinitic relative clauses are linked to their head nouns by an overt marker. In Mandarin, for example, this marker is *de*, as shown in the following examples (relativizing on arguments in A, O, and IO roles, respectively):

- (6.60) *qí zìxíngchē de rén děi xiǎoxīn.*
 ride bicycle NOM person must careful
 ‘People who ride bicycles must be careful.’
 Mandarin | SN | Li and Thompson 1981: 116
- (6.61) *Zhāngsān mǎi de qìchē hěn guì.*
 Zhangsan buy NOM car very expensive
 ‘The car that Zhangsan bought was very expensive.’
 Mandarin | SN | Li and Thompson 1981: 116
- (6.62) *nà shi wǒ gěi nǐ de shū.*
 that be 1SG give 2SG NOM book
 ‘That’s that book I gave you.’
 Mandarin | SN | Li and Thompson 1981: 117

Cantonese shows the same pattern, using a marker *ge*, a classifier with general meaning (Matthews and Yip 1994: 109). A resumptive pronoun is obligatory when an indirect object or object of preposition/coverb/comparison is relativized. It is optional when a direct object or possessor is relativized, and not used when a subject is relativized (Matthews and Yip 1994: 110–11).

- (6.63) A relativized:

sīk ngóh ge yàhn
 know me that people
 ‘people who know me’

Cantonese | SN | Matthews and Yip 1994: 110

- (6.64) O relativized:

ngóh sīk ge yàhn
 I know that people
 ‘people who I know’

Cantonese | SN | Matthews and Yip 1994: 110

- (6.65) O relativized:

ngóh chéng (kéuihdeih) sihk faahn ge pàhngyáuh
 I invite (them) eat food that friends
 ‘friends that I invite for dinner’

Cantonese | SN | Matthews and Yip 1994: 110

- (6.66) Possessor relativized:

(kéuihdeih) tiuh kwàhn hóu dyún ge sailouh léui
 their CL dress very short that little girls
 ‘the little girls whose dresses are very short’

Cantonese | SN | Matthews and Yip 1994: 111

- (6.67) Indirect object relativized:

ngóh sung fā bēi kéuihdeih ge behngyàhn
 I send flower to them that patients
 ‘the patients I sent flowers to’

Cantonese | SN | Matthews and Yip 1994: 111

- (6.68) Object of coverb relativized:

ngóh tùhng kéuihdeih kīnggái ge hohksāang
 I with them chat that students
 ‘the students that I chat with’

Cantonese | SN | Matthews and Yip 1994: 111

- (6.69) Object of comparison relativized:

ngóh jaahn chín dō gwo kéuihdeih ge yàhn
 I earn money more than them that people
 ‘The people who I make more money than’

Cantonese | SN | Matthews and Yip 1994: 111

Tibeto Burman languages show a similar pattern to Sinitic, with the relative clause preceding the head noun. Here are two examples from Daai Chin (western Myanmar) (So Hartmann 2009: 134).

- (6.70) *Angyan üng kah nih yah ei meh sun khyük betü kti.*
 effort INSTR S.AGR:1DU/PL get AO meat DEM disappear again NON.FUT
 ‘The meat that we got with great effort has disappeared again.’
 Daai Chin | TB | So Hartmann 2009: 134

- (6.71) *Khuui k'um a ve kti ah pii e sun noh*
 cave inside LOC is NON.FUT POSS:3SG friend PL DEM ERG
ah nih hei pyen lo mjoh u.
 S.AGR:3DU/PL DIR:forward speak ASP EVID PL
 ‘Her friend who was inside the cave scolded [her], it is told.’
 Daai Chin | TB | So Hartmann 2009: 135

Similar examples from Hakka Lai (also in western Myanmar) show head final structures in which the relativized arguments can be of varied grammatical roles. Relativization in Hakka Lai ‘requires a particular ablaut grade and an invariant relative clause particle’ (Peterson 2003). There are two relativizers – the post posed linkers *tuu* and *mii* which do not appear to differ in meaning, ‘except in terms of their potential relativization targets’, with *tuu* having a lower frequency in texts (Peterson 2003: 422). Here are some examples:

- (6.72) S relativized
in pii ?a it mii lawthlawpaa ka hmu?
 house LOC 3SS sleep₁ REL farmer 1ss see₂
 ‘I saw the farmer who slept in the house.’
 Hakka Lai | TB | Peterson 2003: 421

- (6.73) A relativized
thil ?a bat tuu lawthlawpaa ka hmu?
 thing 3SS hang₁ REL farmer 1ss see₂
 ‘I saw the farmer who hung up the clothes (*lit. the things*).’
 Hakka Lai | TB | Peterson 2003: 422

- (6.74) O relativized
lawthlawpaa ni? ?a ba? mii thil ka hmu?
 farmer ERG 3SS hang₂ REL thing 1ss see₂
 ‘I saw the clothes the farmer hung up.’
 Hakka Lai | TB | Peterson 2003: 422

- (6.75) Oblique relativized

lawthlawpaa ni? thil ?a ba? naak thiŋkuŋ ka hmu?
 farmer ERG thing 3ss hang₂ REL tree 1ss see₂
 'I saw the tree the farmer hung the clothes up on.'

Hakka Lai | TB | Peterson 2003: 422

- (6.76) Oblique instrument relativized

lawthlawpaa ni? ya ?a tan naak naam ka hmu?
 farmer ERG fish 3ss cut₂ REL knife 1ss see₂
 'I saw the knife the farmer cut the fish with.'

Hakka Lai | TB | Peterson 2003: 422

Note that Hakka Lai shows another possible ordering of constituents in relative clauses. Post head relative clauses are possible when there is an adverbial phrase:

- (6.77) S relativized

?in ?ii lawthlawpaa ?a it mii ka hmu?
 house LOC farmer 3ss sleep₁ REL 1ss see₂
 'I saw the farmer who slept in the house.'

Hakka Lai | TB | Peterson 2003: 422

- (6.78) A relativized

nikum ?ii lawthlawpaa thil ?a bat mii ka hmu?
 last.year LOC farmer thing 3ss hang₁ REL 1ss see₂
 'I saw the farmer who hung up the clothes last year.'

Hakka Lai | TB | Peterson 2003: 422

- (6.79) O relativized

nikum ?ii lawthlawpaa ni? thil ?a ba? mii ka hmu?
 last.year LOC farmer ERG thing 3ss hang₂ REL 1ss see₂
 'I saw the clothes the farmer hung up last year.'

Hakka Lai | TB | Peterson 2003: 422

We saw above that White Hmong has post head relative clauses. However, in the Hmong Mien language family more broadly, there is variation in head modifier order in these constructions. The following examples from Thailand Mien and Aizhai Miao show pre head relative clauses, with an explicit relativizing marker (*ŋet*³³ in Thailand Mien, *nay*⁴⁴ in Aizhai Miao), following the Sinitic pattern:

- (6.80) *bwa³³ tshau⁵⁵ bwa³³ kan⁵³ lom³³ ηei³³ bjau³³*
 1PL.INCL fry 1PL.INCL self catch PTC fish
 ‘We fried the fish we caught by ourselves.’

Thailand Mien | HM | Liu 2012: 256

- (6.81) *je³³ tsu⁹⁵⁵ met⁵³ pun³³ je³³ ηet³³ lui^{33/53} xou²⁴*
 1SG wear 2SG give 1SG PTC clothes
 ‘I wear the clothes you gave me.’

Thailand Mien | HM | Liu 2012: 256

- (6.82) *məŋ³¹ dzəŋ³⁵ naŋ⁴⁴ ɔ⁴⁴ qha⁴⁴ za⁴⁴*
 2SG wash PTC clothes dry ASP
 ‘The clothes you washed have dried.’

Aizhai Miao | HM | Yu 2010: 269

- (6.83) *bui⁴⁴ phu²² naŋ⁴⁴ tu³⁵ teu⁵³ le²¹ se²¹ tu³⁵*
 3SG say PTC word NEG mess count word
 ‘What he said is nothing serious.’

Aizhai Miao | HM | Yu 2010: 269

- (6.84) *pui⁵³ ka⁵³ ta⁵³ le⁵³ teho⁴⁴ naŋ⁴⁴ nziu⁴⁴*
 1PL fry self catch PTC fish
 ‘We fried the fish we caught by ourselves.’

Aizhai Miao | HM | Yu 2010: 269 70

Another example of a Hmong Mien language that mixes relative clause strategies (head + relative clause natively, relative clause + head under Chinese influence) is Xong (Sposato 2012). This language even shows the unusual order relativizer + relative clause + head:

- (6.85) *max nonx hlit miex*
 REL eat rice person
 ‘the person who’s eating rice’

Xong | HM | Sposato 2012: 51

Similarly, in languages of the Tai Kadai family whose speakers have had intensive contact with Sinitic speakers, we see variation in head modifier order in relative clause constructions. Sometimes this variation is observed within a single language. Pre head relative clauses occur in Tai languages spoken in China. These structures, which go against the general head initial patterning of noun phrases in Tai languages, may have

been caused by contact with Sinitic languages, suggested by the fact that the morphological marker used in these head final structures often resembles a Sinitic marker.

In an example from Xia'ao Zhuang, a head final relative clause construction is marked by the relativizer *ti*³³, which resembles the pan Sinitic *de* and cognate forms (cf. Mandarin *de*, Cantonese *ge* or more formally *dik*; Matthews and Yip 1994: 110–11):

- (6.86) *mə:ŋ²³¹ sui¹³ ti³³ kai³³ pu³¹ ni¹³ tai¹³ ni:ŋ³³ la:i⁴²*
 2SG buy REL CLF cloth this too small much
 ‘The top that you bought is too small.’

Xia'ao Zhuang | TK | Wei 2012: 156

In Jingxi Zhuang, different types coexist. In a first type, considered to be the main, native type, a head noun is followed by a modifying relative clause, without any linking marker (in some cases also being marked by a classifier placed before the head noun, as in the first example below):

- (6.87) *yo5 ləu4 kən2 lok8 me6 khai3 phak7 pai2 kwa5*
 1SG know CL son granny sell vegetable year last
?dei3 eŋ5wa2 ta3jo2 lu2 ei5 lo3
 get Tsing Hua university admit PP
 ‘I know that the son of the granny who sells vegetables was
 admitted to Tsing Hua University last year.’

Jingxi Zhuang | TK | Huang 2010: 24

- (6.88) *kən2 kin1 ma:k7khai5tam5 mi2 tsəi6 wan2wa2 ma2 a1?*
 person eat haw NEG COP yesterday come Q:PP
 ‘Was it yesterday that the person who ate the haw came?’

Jingxi Zhuang | TK | Huang 2010: 25

In a second type, a recent innovation on the Sinitic model, the relative clause precedes the head noun and is linked to it by the relativizer *ti5*:

- (6.89) *te5 no:ŋ4 ho:i3 yo5 tak7 ti5 wu2 kja2 ne:u2*
 CL girl give me pour PTC CLF tea one
 ‘the cup of tea that the girl poured for me’

Jingxi Zhuang | TK | Huang 2010: 25

Similarly, in Hainan Cham, which has also been in intensive contact with Sinitic languages, different orderings of modifier and head in relative clause constructions are observed. The following example has a verb object phrase as modifier post posed to the nominal head, in a head initial pattern. There is no marker of relativization:

- (6.90) *za:ŋp³³* Ø *li⁵⁵* *ŋanp³³* *za:jp³³* *p^hi⁵⁵*
 person Ø sell vegetables come CMPL
 ‘The person selling vegetables came.’
 Hainan Cham | AN | OZ 1983a: 38 | Thurgood *et al.* 2014: 245

Thurgood *et al.* (2014) state that these ‘posthead relative clauses are inherited’, that is, they are native to Hainan Cham as an Austronesian language. The following two examples illustrate prehead relative clauses which ‘with their distinctive, relative clause final *sa³³* have evolved under the influence of Mandarin’:

- (6.91) Ø *li⁵⁵* *ŋanp³³* *sa³³* *za:ŋp³³* *za:jp³³* *p^hi⁵⁵*
 Ø sell vegetables NOM person come CMPL
 ‘The person who sells vegetables came.’
 Hainan Cham | AN | Ouyang and Zheng 1983: 38 | Thurgood *et al.* 2014: 245
- (6.92) Ø *di⁵⁵* *nan³³* *sa³³* *mo³³* *si²¹* *maj³³* *sa³³*
 Ø lie.down that NOM cow be female NOM
 ‘The (yellow) cow lying down is female.’
 Hainan Cham | AN | Zheng 1997: 73 | Thurgood *et al.* 2014: 245

Thurgood *et al.*’s further comments are worth noting: ‘This is of interest not just as an example of language contact but as an instance of contact producing a typologically marked configuration: for SVO languages, HC prehead relative clauses are typologically rare’ (Thurgood *et al.* 2014: 245). The same point applies to the situations that pertain in the Tai and Hmong Mien languages, discussed above.

6.1.3 Possessive and Part Whole Constructions

All languages provide ways of expressing a relationship of possession between two entities, by modifying a head noun in terms of who or what it belongs to (see Seiler 1983, Heine 1997, Aikhenvald and Dixon 2013, and references therein on the linguistic expression of possession and ownership in the world’s languages). In English, one way to do this is to use special possessive pronouns, as in *my house* or *her brother*. Another is to use the possessive suffix *'s*, as in *Minh's dog* and *that man's hat*. Yet another is to use a phrase built around the word *of*, as in *the wings of a butterfly* or *the leaves of this tree*. MSEA languages do not have dedicated possessive pronouns comparable to English *my*, *his*, or *its*. The most common way to form a possessive construction in MSEA languages is simply to juxtapose the possessor (Pr) and the

possessee (Pe) in a phrase, where the possessee is the head of the phrase, and where there is no explicit marking of the possessive relationship.³

Chamic languages spoken in Vietnam show the MSEA pattern Pe Pr. Here are examples from Northern Roglai and Phan Rang Cham:

- (6.93) *ga?* *sa:k*

roof house

‘the roof of the house’

Northern Roglai | AN | Lee 1966: 65 | Thurgood *et al.* 2014: 101

- (6.94) *sa:k* *hã*

house 2P

‘your house’

Northern Roglai | AN | Lee 1966: 65 | Thurgood *et al.* 2014: 103

- (6.95) *MəKaam tɔɔ?* *tì la tuh ɔpih ikaan MəHlə?* *təmɪ tənəh*

Kam stay LOC down pour all fish Hlok into basket

‘Kam stayed below and poured all of Hlok’s fish into her basket . . . ’

Phan Rang Cham | AN | Thurgood 2005: 499

A language’s pattern of ordering of elements in a possessive construction can be affected by a history of language contact, as we saw in the last section in relation to relative clause constructions. By contrast with the Chamic languages just discussed, in Hainan Cham, spoken in intensive contact with Sinitic languages, possessor occurs before possessed, and an explicit marker of possession is used (the same as the one used in a relative clause construction) in a pattern influenced by Mandarin:

- (6.96) *ŋa²¹ba²¹* *sa³³ sa:y³³*

father’s.old.brother NOM house

‘father’s eldest brother’s house’

Hainan Cham (colloquial) | AN | Zheng 1997: 71 | Thurgood *et al.*

2014: 101

- (6.97) *taj³³tsiay⁴³* *sa³³ na?*²⁴

captain NOM child

‘the captain’s son’

Hainan Cham (colloquial) | AN | Zheng 1997: 71 | Thurgood *et al.*

2014: 101

³ Some examples in this section denote part-whole relations; for convenience, I use the Pr-Pe terminology in these cases, with Pr referring to whole and Pe to part.

- (6.98) *ʔuy²¹k^hiu²¹ sa³³ k^ha:n²³³*

teacher NOM book

‘the teacher’s book’

Cham (colloquial) | AN | Zheng 1997: 71 | Thurgood *et al.* 2014: 101

However, a different structure, with the opposite ordering of Pe and Pr can be observed when the possessor is first person:

- (6.99) *ŋa:n³³ kaw³³ kiʔ²⁴*

hand 1P painful

‘My hand hurts.’

Cham (colloquial) | AN | Zheng 1997: 97 | Thurgood *et al.* 2014: 103

Similarly, Tai Kadai languages in contact with Sinitic languages show alternative orderings of Pr and Pe in possessive constructions. Mulao is an example. ‘Although post modification is the basic order for nouns being modified by a pronoun or another noun, premodification is not uncommon when the modifier indicates possession or location, in which case the structural particle *kɔ* is normally used’ (Wang and Zheng 1993: 67). This particle is likely borrowed from Sinitic, as is appropriate to its function in the Sinitic modelled Pr Pe construction.

- (6.100) Pe Pr ordering (native Tai ordering), no linker

kuj⁵se⁴ niu²

commune we

‘our commune’

Mulao | TK | Wang and Zheng 1993: 67

- (6.101) Pr Pe ordering (Sinitic ordering), no linker

niu² kuj⁵se⁴

we commune

‘our commune’

Mulao | TK | Wang and Zheng 1993: 67

- (6.102) Pe Pr ordering (native Tai ordering), no linker

le² hɔi²

book I

‘my book’

Mulao | TK | Wang and Zheng 1993: 67

- (6.103) Pr Pe ordering (Sinitic ordering), with explicit linker
 $həi^2 \ kɔ \ lɛ^2$
 I PTC book
 ‘my book’
 Mulao | TK | Wang and Zheng 1993: 67

- (6.104) Pe Pr ordering (native Tai ordering), no linker
 $pu^4 \ lu^6yo^2$
 father Lao Wu
 ‘Lao Wu’s father’
 Mulao | TK | Wang and Zheng 1993: 67

- (6.105) Pr Pe ordering (Sinitic ordering), with explicit linker
 $lu^6yo^2 \ kɔ \ pu^4$
 Lao Wu PTC father
 ‘Lao Wu’s father’
 Mulao | TK | Wang and Zheng 1993: 67

Similarly to the Tai Kadai situation, some Hmong Mien languages show both a native head initial pattern and a Sinitic modelled head final pattern in possessive constructions. In Aizhai Miao, the Sinitic Pr Pe pattern often requires marking with the linking particle *nay*⁴⁴:

- (6.106) $te^{53}ku^{44} \ nay^{44} \ mu^{21} \ ni^{53} \ nəy^{44}$
 YBR PTC firewood.chopper be.at here
 ‘Younger brother’s firewood chopper is here.’
 Aizhai Miao | HM | Yu 2010: 263

- (6.107) $tei^{53}wei^{22} \ (nay^{44}) \ dzo^{35} \ qa^{53} \ zu^{35} \ nəy^{31}$
 Jiwei PTC rice more good eat
 ‘Ji Wei Town’s rice is more delicious.’
 Aizhai Miao | HM | Yu 2010: 263

When both orders are possible for the same possessor possessee pair, the head initial structure does not allow the linker *nay*⁴⁴ to be used, while the head final structure allows optional use of the linker *nay*⁴⁴:

- (6.108) $dzaŋ^{31} \ (*nay^{44}) \ kji^{53}la^{44}$
 market PTC Aizhai
 ‘the market of Aizhai’
 Aizhai Miao | HM | Yu 2010: 264

- (6.109) *kji⁵³ta⁴⁴ (nay⁴⁴) dzay³¹*
 Aizhai PTC market
 ‘the market of Aizhai’
 Aizhai Miao | HM | Yu 2010: 264

In many MSEA languages, when a linker between possessor and possessee is optional, it tends to mark a semantic distinction. As Haiman (1983: 793 5, 1985: 130 6) has argued, the presence of formal marking will tend to correlate with more alienable forms of possession.⁴ This is reflected in the patterns of interpretation of possessive constructions in MSEA languages generally. For example, in Thailand Mien, a possessive particle *ŋei³³* is used when the possessed is a physical object that is alienable, such as a towel:

- (6.110) *nin⁵³ ŋei³³ si²³¹tea:u⁵⁵ dɔt⁵⁵ ŋa²³¹*
 3SG PTC towel drop PTC
 ‘His towel dropped.’
 Thailand Mien | HM | Liu 2012: 251

By contrast, when the possessed noun is human or a dwelling closely related to human life, the particle *ŋei³³* is not used:

- (6.111) *bwa³³ pjau⁵⁵ i⁵³ŋɔi³³ dzeŋ³¹ kɔ³³*
 1PL.INCL house today clean very
 ‘Our house is very clean today.’
 Thailand Mien | HM | Liu 2012: 252

The following example from Hmong shows how the presence of an explicit marker (a classifier appropriate to the possessed noun) can, through its marking of higher alienability, signal a distinction between two meanings of a word. In this case, the noun *tx̥* means ‘father’ when no explicit possessive particle is used, and it means ‘husband’ (more alienable than a father) when an explicit particle is used:

- (6.112) *ku tx̥*
 1SG father/husband
 ‘my father’
 Hmong | HM | Clark and Prasithrathsint (1985: 70)

⁴ Some cases of possession are said to be *alienable*. For example, if I buy a mango from a store I can now say it is ‘my mango’, but this relationship of possession can easily cease to hold. If I give the mango to you, I can no longer say it is mine. In other cases, the nature of the possession is *inalienable*. For example, there is no straightforward sense in which my uncle or my soul can no longer be mine. In many languages, a distinction of roughly this kind is reflected in two formal types of possessive construction.

- (6.113) *kǔ tù txǐ*
 1SG CLF father/husband
 ‘my husband’
 Hmong | HM | Clark and Prasithrathsint (1985: 70)

6.2 Nominal Classification

MSEA is often cited in typologies of nominal classification as an area that has *numeral classifiers* (Grinevald 2000, Aikhenvald 2000). Section 6.2.1 describes features of these systems. As a preliminary point, we note that not all languages of the area have numeral classifiers, at least not in elaborated form. As Jenny *et al.* (2015: 123) write, Austroasiatic languages tend to have partial or limited use classifier systems, if they have such systems at all:

Jones (1970) argues that classifier systems in Austroasiatic languages are not native to the family but borrowed from neighbouring languages such as Thai, Burmese, and Chinese. Adams (1989; 1991) elaborates on Jones’s claims and identifies possible sources of a number of classifier constructions in the eastern groups of Austroasiatic. She concludes that “classifiers have been in Austroasiatic languages long enough for them to possess many features which are not derivative but are unique to the language family” (Adams 1991: 75). Though classifiers may be part of the genetic inheritance of Austroasiatic languages, the individual languages show a variety of structures, some arguably influenced by neighbouring languages, and a number of Austroasiatic languages do not make regular use of classifiers at all.

(Jenny *et al.* 2015: 123)

A second point is that the oft cited *numeral classifier* system is not the only type of nominal classification system found in MSEA languages. The languages also have systems that resemble the *noun class* systems found widely in Africa and the Amazon, and ancillary systems that resemble numeral classifiers but which are involved in the use of simpler grammatical modifiers such as demonstratives and specifiers (Grinevald 2000, Aikhenvald 2000). Lao, for example, has four distinct grammatical systems of nominal classification, of which numeral classifiers are one (Enfield 2007: 119–56; cf. Blench 2015, Fedden and Corbett 2017). This said, numeral classifier systems are especially well elaborated in the MSEA area, and we will give them more attention here accordingly.

6.2.1 Numeral Classifier Constructions

When speakers of many MSEA languages want to refer to things in specified quantities (e.g., ‘three sheep’, ‘many sheep’), they use numeral classifier constructions. A numeral classifier is either a dedicated type of nominal that exclusively serves the numeral classifier function or it is a regular noun that can, perhaps secondarily, be used grammatically as a numeral classifier. Numeral classifiers occur contiguous to numerals and other expressions of quantity (Grinevald 2000: 63–4; Aikhenvald 2000: 98). They often, but may not necessarily, appear together with the nouns that they modify. They tend to be semantically general in meaning and are selected for their appropriateness of fit with the main noun in a construction, that is, the noun which refers to whatever it is that is being enumerated or quantified (e.g., ‘sheep’ in the phrase ‘three sheep’).

Two semantic types of numeral classifier are recognized. Mensural classifiers are much like familiar measure terms from English, such as *spoon*, *bag*, *bowl*, etc. as a way of quantifying an amount of some referent. Mensural classifiers convey how much of a quantifiable referent is being referred to (cf. *two spoons of sugar*, *two cups of sugar*, *two kilos of sugar*). Sortal classifiers do not have a ready equivalent in English. Rather than telling a hearer how much of a noun’s referent is being referred to, a sortal classifier conveys something about the general category of entity that the referent belongs to. In Sgaw Karen, the sortal classifier *thu²* is used when counting or quantifying things that are rolled or curled or furled, such as a cigarette, banana leaf, or cloth; and the sortal numeral classifier *di²*, otherwise meaning ‘body’, is used when counting four footed animals and similar entities, including buffaloes, porcupines, bamboo rats, dogs, elephants, statues, ghosts, and Buddhist monks (for further Sgaw Karen examples, see Ratanakul 1998: 102–3).

Here are some illustrative examples of sortal classifier constructions, ranging across the major MSEA language families (with the classifier element underlined):

- (6.114) *hkwei hnă kauñ*
 dog two animal
 ‘two dogs’
 Burmese | TB | Okell 1969: 209

- (6.115) *lot sòòng khan*
 car two handle
 ‘two cars’
 Thai | TK

- (6.116) *neəknipɔn pi ru:p*
writer two person
‘two writers’
Khmer | AA | Capell 1979: 13
- (6.117) *yim tus menyuam*
eight CLF child
‘eight children’
Hmong | HM | Clark 1989: 183
- (6.118) *lēuhng jek gáu*
two CLF dog
‘two dogs’
Cantonese | SN | Matthews and Yip 1994: 93

Examples (6.114)–(6.118) each show the classifier element occurring directly contiguous to a numeral, where that numeral quantifies the referent of a noun that is also included in the construction. The examples also show that MSEA languages vary in how they order the elements of a numeral classifier construction. Jones (1970: 3) noted that languages of the area fall into two groups with respect to the constituent order of numeral classifier constructions. In his Type I languages, the order of basic elements of the numeral classifier construction is NUMERAL CLASSIFIER NOUN⁵ (as seen in the Hmong and Cantonese examples just above), shown in Table 6.2.

In Jones’s Type II languages, the order of basic elements of the numeral classifier construction is NOUN NUMERAL CLASSIFIER (as found in the Burmese, Thai, and Khmer examples given earlier), shown in Table 6.3. Note that this ordering is different from Type I in having the NOUN at the beginning of the construction, but it is like Type I in having NUMERAL CLASSIFIER in the same relative ordering.

Numerical classifier systems often contain large inventories of nominals that can appear in the numerical classifier slot. For Lao, the set supplied in the dictionary of Lao by Alan Kerr (1972) has more than eighty entries. For Thai, Haas similarly states that there are ‘some eighty or ninety’ classifiers in the language (Haas 1942: 201). A non academic source on Thai lists well over 300

⁵ This representation may be taken to suggest that the structure is flat. But in the Lao case, and surely in many other MSEA cases, NUMERAL CLASSIFIER forms a syntactic unit, which in turn can modify a NOUN. This results in the surface string NOUN NUMERAL CLASSIFIER, but it is rightly analysed as [NOUN [NUMERAL CLASSIFIER]], with Type II having the structure [[NUMERAL CLASSIFIER] NOUN].

Table 6.2 *Type I classifier constructions in MSEA and beyond*

	Num	CLF	Noun	
1. Amoy (SN)	sə	nuî	huê	(two-CLF-flower)
2. Vietnamese (AA)	hai	con	chó	(two-CLF-dog)
3. Yao (HM)	pyei	taub	Juq	(four-CLF-dog)
4. Meo, Blue (HM)	plaub	lub	tsev	(four-CLF-house)
5. Nung, Tai (TK)	slám	án	bòk	(three-CLF-flower)
6. White Tai (TK)	há	tô	pa	(five-CLF-fish)
7. Black Tai (TK)	song	tô	mà	(two-CLF-horse)
8. Brôu (AA)	tapoât	lám	alic	(six-CLF-pig)
9. Katu (AA)	pe	nak	ayi	(three-CLF(persons)-we)
10. Sedang (AA)	péa	ngé	kúan-kojrai	(two-CLF-girl)
11. Cham (AN)	pak	dray	kan	(four-CLF-fish)
12. Malay (AN)	tiga	buah	rumah	(three-CLF-house)
13. Indonesian (AN)	dua	orang	anak	(two-CLF-children)
14. Cebuano (AN)	upát-(ka)	buqúk	bátaq	(four-(linker)- CLF-children)

After Jones 1970: 3.

Table 6.3 *Type II classifier constructions in MSEA and beyond*

	Noun	Number	CLF	
1. Burmese (TB)	khwéi	θôun	gáun	(dog-three-CLF)
2. Lolo (TB)	ts'o	seū	joù	(man-three-CLF)
3. Lahu (TB)	g'a^	suh-	hkeh	(chicken-seven-CLF)
4. Lisu (TB)	a ¹ na ⁵	sa ³ -	ma ³	(dog-three-CLF)
5. Maru (TB)	myaw	pit	tau	(horse-four-CLF)
6. Rāwang Nung (TB)	yit	hti	gung	(wild dog-one-CLF)
7. Paluang (AA)	kā	ü	tō	(fish-one-CLF)
8. Cambodian (AA)	monuh	bøy	něaq	(person-three-CLF)
9. Khmu? (AA)	knē?	mooy	too	(rat-one-CLF)
10. Karen, Sgaw (TB)	pya	lwi	ya	(person-four-CLF)
11. Thai (TK)	mää	s̄ooŋ	tua	(dog-two-CLF)
12. Shan (TK)	mä	hah	tō	(horse-five-CLF)
13. Javanese (AN)	kertas	sepuluh	lembar	(paper-ten-CLF(sheets))

After Jones 1970: 3.

classifiers.⁶ Claims of many dozens if not hundreds of classifiers in these systems might be technically correct, but they can be misleading. This is because only a small number of classifiers are used with any frequency. Such lists therefore show a classic long tail distribution of token frequency among types, whereby

⁶ www.thai-language.com/ref/classifier-list

Table 6.4 *Khmu (AA) classifiers*

CLF	Used with reference to
1. kló:ng	solid, round objects (e.g., fruit, vegetables, sun, moon, house)
2. trl̥m	long, flat objects (e.g., leaves, piece of meat, hammer, comb, bag)
3. tó:	animals
4. p̥v:n	piece of cloth, shirt, trousers
5. sén	flexible, long, narrow objects (e.g., rope)
6. kòn	human beings
7. smtúm	bunch of fruit or ‘other things’
8. plàh	one of a pair of human organs
9. cùm	one of a pair of things such as human organs
10. krl̥h	mouthful of rice
11. bò:n	piece of land
12. kmlò?	piece of meat
13. lém	oblong object (e.g., piece of wood, finger, drinking water container)

Reproduced from Premsrirat 1987: 34.

speakers rely mostly on only a fraction of the available classifiers and a majority of available classifiers are used only infrequently.

Let us consider the classifier system of Khmu. Premsrirat (1987: 34) states that in Khmu, numeral classifiers ‘form a closed set of words and are required when most kinds of items are counted’. Table 6.4 shows a representative list of the more frequent numeral classifiers used in Khmu (Premsrirat 1987: 34, slightly adapted transcription and glossing).⁷

We can note three important characteristics of the Khmu system. The first is that constituents in a numeral classifier construction canonically appear in the order NOUN NUMERAL CLASSIFIER (Jones’s Type II), as in the following examples:

- (6.119) *tlɔ:j mò:j klɔ:y*
 banana one CLF
 ‘a banana’
 Khmu | AA | Premsrirat 1987: 34
- (6.120) *rmɔ:j mò:j plàh*
 ear one CLF
 ‘an ear’
 Khmu | AA | Premsrirat 1987: 34

⁷ Söderberg (2006: 14) argues that they are not a closed class: ‘[T]hey are independent lexemes and can be considered an open class because of the wide use of repeaters [which allow many nouns to classify themselves by appearing in the classifier slot]’.

- (6.121) *scàay mòoy tóo*
 elephant one CLF
 ‘an elephant’
 Khmu | AA | Söderberg 2006: 14

- (6.122) *cm.rɔ? pà:r kòn*
 man two CLF
 ‘two men’
 Khmu | AA | Jenny *et al.* 2015: 125

Second, while there is a fixed, tight relation between the numeral and the classifier in what we can call a quantifier phrase (NUMERAL CLASSIFIER) the syntactic relation between this phrase and the main noun is loose. So, as Jenny *et al.* (2015: 127) note, ‘the quantifier phrase may be separated from the head noun by verbal and adverbial constituents and appear at the end of the clause’. (See Enfield 2007: 120 for the same point for Lao.) Here is a Khmu example, in which the noun classifier phrase *mò:j tó* ‘one CLASSIFIER’ is separated from the main noun *k.né?* ‘rat’ by a prepositional phrase:

- (6.123) *kò pýn k.né? hák:n tá ká:y ʔò? mò:j tó*
 3SG.M shoot rat die LOC house 1SG one CLF
 ‘He shot one rat dead in my house.’
 Khmu | AA | Jenny *et al.* 2015: 127

A third feature of the system is that many nouns can be used as classifiers for themselves, creating a phenomenon known as repeater classifiers (termed ‘self classifiers’ in Premsrirat 1987). Here are some examples:

- (6.124) *màt mò:j màt*
 eye one eye
 ‘an eye’
 Khmu | AA | Premsrirat 1987: 35

- (6.125) *ti:y cét ti:y*
 water.container seven water.container
 ‘seven water containers’
 Khmu | AA | Premsrirat 1987: 35

- (6.126) *t.lɔy nám mò:j t.lɔy*
 log big one log
 ‘one big log’
 Khmu | AA | Jenny *et al.* 2015: 128

- (6.127) *trɔɔŋ kné mòoy trɔɔŋ*
 track rat one track
 ‘a rat track’
 Khmu | AA | Söderberg 2006: 20

In the last two examples, the quantified nominal is a noun phrase consisting of a head (first) and modifier (second). Note that in a repeater construction the classifier slot is not filled by the entire main noun phrase, but only the head noun (*t.lɔŋ* ‘log’ and *trɔɔŋ* ‘track’, respectively).

The Khmu situation is not representative of Austroasiatic languages more generally. As noted, other Austroasiatic languages have simpler systems, and in many Austroasiatic languages, classifier constructions are optional or non-existent. For example, in Semelai, the set of classifiers ‘is small and rigid, and their use is not obligatory’ (Kruspe 2004: 206). Kruspe identifies two subsystems: a numeral classifier system consisting of terms mostly borrowed from Malay and an indigenous system the terms of which can be used either as numeral classifiers or independently of numeral classification (Kruspe 2004: 206). These two subsystems are used for different functions. Malay loan numeral classifiers can also ‘be used to express relative dimension’ (Kruspe 2004: 207). The primary function of indigenous classifiers is akin to a mensural classifier usage, but vaguer: they function ‘to express a subset of referents, rather than a precise quantity’ (Kruspe 2004: 208).

Unlike Khmu, Semelai follows Jones’s Type I pattern, with the constituent order NUMERAL CLASSIFIER NOUN. Here are some examples, illustrating some of the contexts of use of numeral classifiers in Semelai, namely, anaphoric reference, expressing relative dimension, and expressing a subset of the referents, respectively:

- (6.128) *kəh, ki kʰɔm pon ?ikur, dwa? kdor, dwa? rmɔl sdom*
 3 3A get four CLF two be.female two be.male only
 ‘Him, he got four (of them), just two (who) were female, (and) two (who) were male.’
 Semelai | AA | Kruspe 2004: 207

- (6.129) *ny tʰɔy mə bje? dɔl*
 NMZ be.big one CLF house
 ‘(It was) the width (of) a house.’
 Semelai | AA | Kruspe 2004: 207

- (6.130) *mə k<ŋ>ampoy smə?*
 one settlement<UNIT> people
 ‘one settlement (of) people’
 Semelai | AA | Kruspe 2004: 208

Jahai, a neighbouring Aslian language, has a small inventory of classifiers (eleven attested), more than half of which are loans from Malay (Burenhult 2005: 80 2). Only two classifiers are widely used. As in Semelai, the use of classifiers is optional when counting or quantifying. Classifier constructions are ‘pragmatically motivated references to and emphases and specifications of the noun they refer to rather than syntactically obligatory components of quantified NPs’ (Burenhult 2005: 81). Classifiers can be used outside of quantifying contexts (as with Semelai) for anaphoric reference. A classifier ‘may be used to replace a noun which has been introduced earlier in the discourse or an implicit noun not previously introduced overtly’ (Burenhult 2005: 80).

The order of constituents in a Jahai numeral classifier construction is NUMERAL CLASSIFIER NOUN (Jones’s Type I):

- (6.131) *je? bdil spuloh k<n?>mɔ? kasa?*
 1s to.shoot ten CLF<UNIT> sambar.deer
 ‘I shot ten sambar deer.’
 Jahai | AA | Burenhult 2005: 82

As was seen in Khmu, the main noun is not tightly integrated in the construction: ‘a classifier and its modifying numeral seldom form a phrasal unit with the noun they refer to. Instead, the numeral and the classifier typically make up a separate noun phrase which is in some way detached from and syntactically opposed to the noun’ (Burenhult 2005: 80). Here is an example, in which the main noun *?amey* ‘siamang’ (a primate species) is separated from its quantifier phrase *duwa? k<n?>mɔ?* ‘two CLASSIFIER’:

- (6.132) *ja je? bdil ?amey k tɔm mayəh ?on leh, duwa? k<n?>mɔ?*
 RT 1s to.shoot siamang LOC river Mangga there EMP two
 CLF<UNIT>
 ‘Then I shot siamangs by Mangga river! Two of them!’
 Jahai | AA | Burenhult 2005: 81

This separability of head noun and classifier phrase is not exclusive to Jahai or Austroasiatic. Here is an example from Lao, showing a similar phenomenon:

- (6.133) *kuu3 sùù4 paa3 juu1 talaat5 sòòng3 too3*
 1SG.B buy fish be.at market two CLF.ANIM
 ‘I bought fish at the market, two (of them).’ (‘I bought two fish at the market.’)
 Lao | TK | Enfield 2007: 120

Sinitic languages also follow Jones’s Type I order, but differ sharply from the Austroasiatic cases in that their classifiers are obligatory in usage and come from much larger inventories. Cantonese has a set of more than sixty numeral classifiers (Matthews and Yip 1994: 92), with various subtypes. They are obligatory not only when counting or quantifying something, but also for other functions, including when referring to a specific (as opposed to generic) object, when a noun follows the demonstratives *nī* ‘this’ and *gó* ‘that’, and when a noun occurs with *bī* ‘which’ or *bīn* . . . *dōu* ‘any’ (Matthews and Yip 1994: 93–4). Classifiers in these functions are assigned according to broad semantic principles, similarly to the Khmu situation described above. Here are some examples of Cantonese classifiers *in situ*:

- (6.134) *nī júng fā*
 this CLF flower
 ‘this kind of flower’
 Cantonese | SN | Matthews and Yip 1994: 100
- (6.135) *yāt leuih mahntāih*
 one CLF problem
 ‘one genre of problem’
 Cantonese | SN | Matthews and Yip 1994: 100
- (6.136) *yāt faai mihbāau*
 one CLF bread
 ‘a slice of bread’
 Cantonese | SN | Matthews and Yip 1994: 102

The examples just cited illustrate the generic or sortal function of classifiers. The following examples show the same kind of structure, where the classifier either groups sets of nominal referents into collectives (example 6.137), or measures an amount of the nominal referent in terms of a container that holds it (examples 6.138 and 6.139):

- (6.137) *nī bāan hohksāang*
 this CLF student
 ‘this class of students’
 Cantonese | SN | Matthews and Yip 1994: 96

- (6.138) *léuhng gwun bējáu*
 two CLF beer
 ‘two cans of beer’
 Cantonese | SN | Matthews and Yip 1994: 99

- (6.139) *léuhng wún faahn*
 two CLF rice
 ‘two bowls of rice’
 Cantonese | SN | Matthews and Yip 1994: 99

Mandarin has a similar system, with some 200 classifiers, divided into different classes, including those shown in Table 6.5.

Table 6.5 *Mandarin classifiers: Categories and inventory sizes*

Type	Size	Functional definition
Individual classifiers	51	‘modify nouns according to the entity’s shape, or other properties’
Group classifiers	46	‘used for a group or collection of individuals’
Partition classifiers	39	‘represent portions of things’
Container classifiers	36	‘container nouns used as measures’
Temporary classifiers	14	‘use the outside extent of objects to measure quantity’
Standard measures	16	‘are measures proper’

After X. Li 2013: 21, following Chao 1968.

Here are examples of each subclass in Mandarin:

- (6.140) Individual classifier
liang ke shu
 two CLF tree
 ‘two trees’
 Mandarin | SN | X. Li 2013: 17

- (6.141) Group classifier
san qun xuesheng
 three CLF.group student
 ‘three groups of students’
 Mandarin | SN | X. Li 2013: 17

- (6.142) Partition classifier

yi jie shengzi

one CLF.section rope

‘a section of rope’

Mandarin | SN | X. Li 2013: 18

- (6.143) Container classifiers

yi ping jiu

one CLF.bottle wine

‘a bottle of wine’

Mandarin | SN | X. Li 2013: 18

- (6.144) Temporary classifiers

yi bizi hui

one CLF.nose dust

‘a noseful of dust’, an idiom meaning ‘being refused’

Mandarin | SN | X. Li 2013: 19

- (6.145) Standard measures

wu mi bu

five CLF.metre cloth

‘five metres of cloth’

Mandarin | SN | X. Li 2013: 19

In Hmong Mien languages, numeral classifiers form large and elaborate systems comparable to those just described for Sinitic languages. In research on Hmong Mien languages, classifiers have sometimes been termed ‘selectors’ (see Lyman 1979). Lyman (1979: 95–100) gives a list of 167 of these in Mong Njua, grouped into four types: unit selectors, grouping selectors, independent selectors, and ambifunctional selectors (see Lyman 1979: 20–1 for discussion). These appear in a construction of Jones’s Type I, with the constituent order QUANTIFIER SELECTOR NOUN, as shown in the following examples:

- (6.146) *pé lú cě*

three USL house

‘three houses’

Mong Njua | HM | Lyman 1979: 21

- (6.147) *pé thaø qhau*

three GSL shoe

‘three pairs of shoes’

Mong Njua | HM | Lyman 1979: 21

- (6.148) *ḡua t̄e mblē*
 nine AMSL paddy
 ‘nine handfuls of paddy’
 Mong Njua | HM | Lyman 1979: 21

Numeral classifiers are used in Mong Njua not only for counting nouns, but also as quasi definite articles (Lyman 1979: 94).

Other Hmong Mien languages show similar patterns of structure and usage, with large inventories of numeral classifiers, following Jones’s Type I constituent order pattern, and being obligatory in the languages when counting or quantifying things.

Aizhai Miao has over 200 numeral classifiers (Yu 2010: 127). Here are some examples:

- (6.149) *a⁴⁴ tu²² du³⁵*
 one QW tree
 ‘a tree’
 Aizhai Miao | HM | Yu 2010: 129

- (6.150) *a⁴⁴ eɔŋ⁵³/qu⁵³ pei⁵³*
 one QW hair
 ‘a hair’
 Aizhai Miao | HM | Yu 2010: 130

- (6.151) *a^{44/53} le⁵³ nɔŋ²²*
 one QW rain
 ‘a drop of rain’
 Aizhai Miao | HM | Yu 2010: 133

Examples from Thailand Mien show how the choice of numeral classifiers can be a mechanism for expressing politeness. In the following two examples, the same classifier *tau⁵³* is used for people as for animals:

- (6.152) *jet³¹ tau⁵³ sja²⁵⁵ tɔn³³*
 one QW woman
 ‘a woman’
 Thailand Mien | HM | Liu 2012: 99

- (6.153) *jet³¹ tau⁵³ da²³¹ mau⁵³*
 one QW tiger
 ‘a tiger’
 Thailand Mien | HM | Liu 2012: 99

But when referring to a respected person, such as a guest, a special classifier *la:n⁵³* is used:

- (6.154) *jet³¹ la:n⁵³ khe²⁵⁵ mjen⁵³*
 one QW guest person
 ‘a guest’

Thailand Mien | HM | Liu 2012: 98

Illustrating another frequent function of classifiers in MSEA languages, these examples from Thailand Mien show that repetition of a numeral classifier can convey the meaning ‘each and every’:

- (6.155) *pwo:n^{31/53} pwo:n³¹ ja³¹ ma:i⁵³ sjay^{33/53} tshja³³*
 QW QW also have new car
 ‘Each and every family has a new car.’

Thailand Mien | HM | Liu 2012: 110

- (6.156) *le⁵³ le⁵³ sa³⁵ zt²² to³⁵ məŋ³¹*
 QW QW all know get you
 ‘Everyone knows you.’

Aizhai Miao | HM | Yu 2010: 142

These examples from Cantonese show the same ‘each and every’ meaning of classifier reduplication (in combination with *dōu* ‘all’):

- (6.157) *Go go (yāhn) dōu séung máaih láu.*
 CL CL (people) all want buy flat
 ‘Everyone wants to buy a flat.’

Cantonese | SN | Matthews and Yip 1994: 96

- (6.158) *Mhaih jek jek (gípiu) dōu wúih sīng ge.*
 not be CL CL (share) all will rise PRT
 ‘Not all (shares) are going to rise.’

Cantonese | SN | Matthews and Yip 1994: 96

Turning now to Tai languages, in Judu Kelao, spoken in China, the numeral classifier construction has the same Type I structure as just observed in Hmong Mien languages. Here are some examples:

- (6.159) *tsɿ³³ say³³ la³⁵*
 one QW child
 ‘one child’

Judu Gelao | TK | Kang 2009: 120

- (6.160) *tsɿ³³ du³⁵ nzi³¹*
 one QW ox
 ‘a bull’
 Judu Gelao | TK | Kang 2009: 120
- (6.161) *tsɿ³³ phu³¹ tin³¹*
 one QW tree
 ‘a tree’
 Judu Gelao | TK | Kang 2009: 121
- (6.162) *tsɿ³³ əŋ³³ tsa³³tsa³³*
 one QW rubbish
 ‘a pile of rubbish’
 Judu Gelao | TK | Kang 2009: 124

A point of contrast between Judu Gelao and the languages we have examined already is that when a demonstrative is used, it appears *after* the CLASSIFIER NOUN phrase, as shown here:

- (6.163) *kui³¹ vu³³ŋuŋ³³ ŋi³⁵*
 QW horse this
 ‘this horse’
 Judu Gelao | TK | Kang 2009: 127
- (6.164) *səu³¹ kan³¹ qə³¹tshu³⁵ ŋi³⁵*
 two QW person this
 ‘these two people’
 Judu Gelao | TK | Kang 2009: 128

Xia’ao Zhuang, also spoken in China, has the same Type I constituent order in classifier constructions as is found in Sinitic languages. The language illustrates the common phenomenon in which the use of the number ‘one’ triggers a special constituent order. Look first at these examples with ‘two’, which follow Jones’s Type I structure NUMERAL CLASSIFIER NOUN:

- (6.165) *ro:ŋ⁴² tu¹³ jo:ŋ²³¹*
 two QW lamb
 ‘two lambs’
 Xia’ao Zhuang | TK | Wei 2012: 107

- (6.166) *ro:y⁴² ko⁴² fei¹³*
 two QW tree
 ‘two trees’
 Xia’ao Zhuang | TK | Wei 2012: 107

Compare these to the following cases in which the numeral is ‘one’. The numeral is moved to the end of the phrase, with the order CLASSIFIER NOUN ONE (in the same way as we have just seen with the post posing of demonstratives in Judu Gelao):

- (6.167) *tu¹³ lə:k¹³ deu⁴²*
 QW child one
 ‘a child’
 Xia’ao Zhuang | TK | Wei 2012: 107
- (6.168) *liek³³ fei¹³ deu⁴²*
 QW rod one
 ‘a rod’
 Xia’ao Zhuang | TK | Wei 2012: 108

In line with the Judu Gelao case, this same constructional ordering is observed when a demonstrative is used:

- (6.169) *pə:n⁴² ruu⁴² te²⁴*
 QW book that
 ‘that book’
 Xia’ao Zhuang | TK | Wei 2012: 112
- (6.170) *ro:y⁴² pu¹³ wə:n²³¹ ni¹³*
 two QW person this
 ‘these two people’
 Xia’ao Zhuang | TK | Wei 2012: 112

Other Tai Kadai languages in particular, those spoken further south, such as Thai and Lao show the fundamentally distinct ordering of Jones’s Type II pattern in their basic numeral classifier pattern: NOUN NUMERAL CLASSIFIER. Here are two examples from Lao (the second illustrating the use of a noun as a repeater classifier):

- (6.171) *kuu3 siù4 paa3 sòòng3 too3*
 1SG.B buy fish two CLF.ANIM
 ‘I bought two fish.’
 Lao | TK | Enfield 2007: 121

- (6.172) *kuu3 siø hêt1 hang2 sip2 hang2*
 1SG.B IRR make nest ten nest
 ‘I’m going to make ten nests (for the chickens).’
 Lao | TK | Enfield 2007: 122

Lao also illustrates a quirk of ordering in classifier phrases associated with the numeral ‘one’. When *nùng1* ‘one’ is used to quantify something, the usual Jones Type II structure may be used, as follows:

- (6.173) *kuu3 siù4 paa3 nùng1 too3*
 1SG.B buy fish one CLF.ANIM
 ‘I bought one fish.’
 Lao | TK | Enfield 2007: 121

But there is also an alternative pattern, available only for the numeral *nùng1* ‘one’. The ordering of classifier and numeral can be reversed, in which case the classifier is (usually) phonologically reduced, as shown in this re rendering of example (6.173):

- (6.174) *kuu3 siù4 paa3 toø nùng1*
 1SG.B buy fish MC.ANIM one
 ‘I bought a fish.’
 Lao | TK | Enfield 2007: 121

The Jones Type II system is also observed in Tibeto Burman languages (in contrast to their Type I ordered Sinitic cousins). Lahu, for example, has a system of numeral classifiers of similar size and functional diversity as seen for the Sinitic languages mentioned above, though used in structures of the Type II NOUN NUMERAL CLASSIFIER order. The following examples show that, as is often the case in these sorts of systems, there is no one to one assignment of classifiers to nouns. Lahu speakers can often choose from among different classifiers for the same noun, when referring to certain familiar entities, such as pigs or houses:

- (6.175) *và? tê khe*
 pig one CLF
 ‘a pig’
 Lahu | TB | Matisoff 1973a: 91

- (6.176) *và? tê mà*
 pig one CLF
 ‘a pig’
 Lahu | TB | Matisoff 1973a: 91

- (6.177) *yè tê yè*
 house one CLF
 ‘a house’
 Lahu | TB | Matisoff 1973a: 91

- (6.178) *yè tê šī*
 house one CLF (*elegant*)
 ‘a house’
 Lahu | TB | Matisoff 1973a: 91

- (6.179) *yè tê mà*
 house one CLF
 ‘a house’
 Lahu | TB | Matisoff 1973a: 91

Here are some further examples, showing some of the functional categories of classifier identified by Matisoff (1973a):

Measure classifiers

- (6.180) *i kâ? šé? lî?*
 water three CLF
 ‘three litres of water’
 Lahu | TB | Matisoff 1973a: 90

- (6.181) *là tê khê*
 tea one CLF
 ‘a cup of tea’
 Lahu | TB | Matisoff 1973a: 90

Group classifier

- (6.182) *yâ é tê gî*
 kid one CLF
 ‘a bunch of kids’
 Lahu | TB | Matisoff 1973a: 90

Repeater

- (6.183) *qhâ? nî qhâ?*
 village two CLF
 ‘two villages’
 Lahu | TB | Matisoff 1973a: 89

Reduplication of numeral classifiers in Lahu has a number of functions, as these examples illustrate:

Inclusive

- (6.184) *tē qhò? qhò?*
 one CLF.year CLF.year
 ‘every year’
 Lahu | TB | Matisoff 1973a: 93

Indefinite

- (6.185) *chɔ tē ñga ñga*
 person one CLF CLF
 ‘somebody or other’
 Lahu | TB | Matisoff 1973a: 93

Approximate

- (6.186) *mà? cɔ? sī tē chi chi*
 orange one CLF.ten CLF.ten
 ‘about ten oranges’
 Lahu | TB | Matisoff 1973a: 93

Also in the Tibeto Burman family, Burmese has an elaborate and well studied system. Pe (1965) distinguishes three separate systems of ‘numeratives’ – classifiers, quantifiers, and repeaters – defined as follows.⁸

A classifier is a word for an attribute of a specific object, some of which may have more than one; a repeater is the specific object itself or part of it, used as numerative; whilst a quantifier concerns itself with the estimating of things by some sort of measure – size, extension, weight, amount or number *especially* of ten or multiples of ten. All these are numeratives, since the patterns are used in enumerating things.

(Pe 1965: 166)

See Pe (1965) for a detailed analysis. The Austronesian languages of MSEA also have numeral classifier constructions, though it is suggested that these may have been acquired through areal processes. In Hainan Cham, most of the attested mensural classifiers are borrowed from Mandarin, plus one native form, one Malay loan word, and one of unknown

⁸ Becker (1975) presents an analysis of the ‘numerative classifier system’ of Burmese as a paradigmatic system distributed in a metaphoric space.

origin (Thurgood *et al.* 2014: 158). Speakers of Hainan Cham use classifiers of both the mensural and sortal type (Thurgood *et al.* 2014: 157). The constructions are of the Jones Type I structure (NUMBER CLASSIFIER NOUN), like the Sinitic languages that Hainan Cham has been in intensive contact with.

The following example shows a mensural classifier construction in Hainan Cham:

- (6.187) *pia³³ p^hi⁵⁵ ta²¹ maw²¹ ko⁵⁵saf³³*
 plant CMPL one CLF peanuts
 ‘Already planted one mǔ of peanuts’ (1 mǔ 0.067 hectares)
 Hainan Cham | AA | Zheng 1997: 88, cited in Thurgood *et al.* 2014: 158

Of the sortal classifiers in Hainan Cham, two general classifiers are most frequently used. This simple system can be said to sort referents into the categories animate versus inanimate. The classifier *se⁵⁵/se²¹* is used with reference to ‘people, animals, birds’ and more generally for animates (Thurgood *et al.* 2014: 160), while *pho⁵⁵/pho²¹* refers to ‘round objects, pieces’ and more generally for inanimates (Thurgood *et al.* 2014: 160).

6.2.2 Class Terms

Many MSEA languages obligatorily use a class term as the head of the phrase for certain classes of entity, particularly members of natural kind sets such as trees, fish, and beans.⁹ Compare English and Lao. In English, we can say either ‘I planted a eucalyptus’ or ‘I planted a eucalyptus tree’, where the word ‘tree’ is optional. The same is true for many other species, including oak, fir, pine, ash, etc. In Lao, however, the word analogous to *tree* is obligatory. (Also note that in English often the head word cannot be used at all: ‘We had salmon (*fish) for dinner’.)

Many MSEA languages feature class terms. Class term systems have only partial overlap with their inventories of numeral classifiers. They also have quite distinct syntactic organization.

Bisang (1993) describes class terms in White Hmong. (He uses the term ‘class nouns’.) Table 6.6 shows sets of referring expressions consisting of two nouns in combination. In these expressions, the first noun acts as a class term, denoting the broader category to which the referent belongs (e.g., ‘tree’, ‘fruit’, ‘bird’, ‘fish’), and the second noun gives a more specific characterization of the referent (e.g., by specifying a species or variety).

Other languages of the same family show the same kinds of class terms. Here are some examples from Aizhai Miao:

⁹ DeLancey (1986) cites Haas (1964) as the creator of the phrase ‘class term’.

Table 6.6 *Class nouns in White Hmong (Hmong-Mien)*

Class term	Example	Meaning
1. <i>ntoo</i> 'tree'	<i>ntoo ciab</i> <i>ntoo moog/ntoo nceb moog</i> <i>ntoo txiv roj</i> (oil, Heimbach 1979: 279)	'larch' 'species of chestnut tree' 'a tree with fruit producing oil'
2. <i>txiv</i> 'fruit'	<i>txiv duaj</i> <i>txiv txhais</i> <i>txiv tsawb</i> (the banana plant, the banana, Heimbach 1979: 347)	'peach' 'mango' 'banana'
3. <i>noog</i> 'bird'	<i>noog даж</i> (yellow) <i>noog dag laug</i> <i>noog ncaws nplej</i> (bird peck unhulled rice, Heimbach 1979: 146, 165)	'yellow thrush, oriole' 'jay' 'hoopoe'
4. <i>ntsəs</i> 'fish'	<i>ntsəs nab</i> (snake) <i>ntsəs choo</i>	'eel' 'ray'
5. <i>kws</i> 'an expert'	<i>ntsəs liaj</i> (rice field) <i>kws hlau</i> (iron) <i>kws ntoo</i> (tree, wood) <i>kws nyiaj</i> (silver)	'fish living in the ricefield' 'blacksmith' 'carpenter' 'silversmith'
6. <i>tub</i> 'son'	<i>tub qhe</i> <i>tub rog</i> (war, Heimbach 1979: 279) <i>tub luam</i> (business, trade, Heimbach 1979: 121)	'slave, servant' 'soldier' 'merchant, trader'
7. <i>sab</i> 'side, direction'	<i>sab hnub tuaj</i> (side sun rise) <i>sab hnub poob</i> (side sun set) <i>sab hauv qab</i> (side at below)	'East' 'West' 'lower part'
8. <i>yav</i> 'used for periods of time or abstractions related to time'	<i>yav laus</i> (old)	'age'
9. <i>chaw</i> 'place, region'	<i>yav tom ntej</i> (at-in front of) <i>yav tom qab</i> (at-behind) <i>chaw thaiv dej</i> (to dike/dam water) <i>chaw pw</i> (to sleep) <i>chaw chim</i> (to hate/be angry at)	'future' 'past' 'a dam/dike' 'sleeping place' 'cause of hatred'
10. <i>lus</i> 'word, speech'	<i>lus khev</i> (allow/permit) <i>lus qhuab</i> (to discipline/to teach) <i>lus cuav qaub</i> (false no good)	'permission' 'admonition, exhortation' 'lies'

Table 6.6 (*cont.*)

Class term	Example	Meaning
11. <i>kev</i> ‘way’	<i>kev pluag</i> (poor) <i>kev txom nyem</i> (miserable) <i>kev tu npua</i> (to look after pig)	‘poverty’ ‘misery’ ‘pig-breeding’
12. <i>dais</i> ‘bear’ (Heimbach 1979: 481)	<i>dais dev</i> (dog, Heimbach 1979: 34) <i>dais nees</i> (horse, Heimbach 1979: 138)	‘honey bear’ ‘black bear’
13. <i>taum</i> ‘bean’ (Heimbach 1979: 485)	<i>taum lag</i> <i>taum suav</i> <i>taum pauv</i>	‘long beans’ ‘short beans’ ‘soybean’

Data are from Bisang 1993: 42–5, except where specified as from Heimbach 1979. The relevant meanings of component morphemes provided by Bisang (1993) are bracketed, except where otherwise specified.

A. *du*³⁵ ‘tree’

- (6.188) *a*⁴⁴ *tu*²² *du*³⁵
one CLF tree
'a tree'
Aizhai Miao | HM | Yu 2010: 129

- (6.189) *a*⁴⁴ *tu*²² *du*³⁵ *ce*⁵³
one CLF tree fir
'a fir tree'
Aizhai Miao | HM | Yu 2010: 129

- (6.190) *a*⁴⁴ *tu*²² *du*³⁵ *gwa*³¹
one CLF tree peach
'a peach tree'
Aizhai Miao | HM | Yu 2010: 129

B. *zei*⁵³ ‘vegetable’

- (6.191) *a*⁴⁴ *ko*⁴⁴ *zei*⁵³
one CLF vegetable
'a vegetable'
Aizhai Miao | HM | Yu 2010: 129

- (6.192) *a⁴⁴ tu⁴⁴ z̥ei⁵³*
 one CLF vegetable
 ‘a handful of vegetables’
 Aizhai Miao | HM | Yu 2010: 130
- (6.193) *p̥u⁵³ te^{44/21} te⁴⁴ nəŋ³¹ z̥ei⁵³ gəŋ⁵³.*
 1DU CLF~CLF eat vegetable green
 ‘We eat pak choi for every meal.’
 Aizhai Miao | HM | Yu 2010: 145

In Tai languages, class terms are widely attested, for example as described by DeLancey (1986). Here are some of Delancey’s Thai examples (with the class terms underlined):

- (6.194) *ráan ɻaahǎan*
 shop food
 ‘restaurant’
 Thai | TK | DeLancey 1986: 438
- (6.195) *khon sák phâa*
 person wash clothes
 ‘laundress’
 Thai | TK | DeLancey 1986: 438
- (6.196) *lîuk taa sǎam lîuk*
 ball eye three CLASS
 ‘three eyeballs’
 Thai | TK | DeLancey 1986: 442

DeLancey (1986) re analyses Saul’s (1964) description of the classification devices in Nung. Saul’s analysis focused not on the class terms but on the more specific nominals to which the class terms attach. He called the class terms ‘Qualifiers’. ‘Qualifiers, like imposed quantifiers, designate a particular species of the nouns that they precede. The meanings are generally more qualitative than quantitative’ (Saul 1964: 279). Here are two Nung examples, showing that the class terms (underlined) are distinct from the numeral classifiers:

- (6.197) *slam an mak kam*
 three CLF fruit orange
 ‘three oranges’
 Nung | TK | DeLancey 1986: 443

- (6.198) *slam an neo mu'*
 three CLF digit hand
 'three fingers'
 Nung | TK | DeLancey 1986: 443
- (6.199) *slam tew kha lo*
 three CLF road road
 'three roads'
 Nung | TK | DeLancey 1986: 443
- (6.200) *sloŋ kha thu slon*
 two CLF head garlic
 'two bunches of bulb garlic (i.e., a specific type of garlic)'
 Nung | TK | DeLancey 1986: 443

Similar structures are observed in Mulao, as shown in the following examples:

- (6.201) *nɔk⁸ py⁵*
 bird sparrow
 'house sparrow'
 Mulao | TK | Wang and Zheng 1993: 27
- (6.202) *məm⁶ myət⁴*
 fish carp
 'carp'
 Mulao | TK | Wang and Zheng 1993: 27
- (6.203) *kya¹ ming³*
 insect light
 'fluorescent insect'
 Mulao | TK | Wang and Zheng 1993: 27
- (6.204) *mu⁵ la:i⁵*
 pig wild.boar
 'wild boar'
 Mulao | TK | Wang and Zheng 1993: 27
- (6.205) *kwa¹ thaŋ⁵*
 gourd cucumber
 'cucumber'
 Mulao | TK | Wang and Zheng 1993: 28

- (6.206) *mai⁴ kyan'*
 tree tallow
 'Chinese tallow tree'
 Mulao | TK | Wang and Zheng 1993: 28
- (6.207) *ma¹ ka:t⁷*
 vegetable mustard
 'mustard plant'
 Mulao | TK | Wang and Zheng 1993: 28
- (6.208) *hui¹ tɔ:²*
 fruit peach
 'peach'
 Mulao | TK | Wang and Zheng 1993: 28

For details on class terms in another Tai language, see Enfield (2007: 146ff.) on Lao.

Finally, we can note that class terms are also found in Sinitic languages, but with the expected opposite ordering of class term and specifying noun. Following the head final order of Sinitic grammar, languages like Mandarin position the class term after the specifying noun. Here are some examples:

A. *bǐ* 'pen, writing instrument' (Chinese Academy of Social Sciences (CASS) 2012: 68)

- (6.209) *máo bǐ*
 hair CT.pen
 'Chinese writing brush'
 Mandarin | SN | CASS 2012: 875
- (6.210) *qiān bǐ*
 'lead CT.pen'
 'pencil'
 Mandarin | SN | CASS 2012: 1033
- (6.211) *fěn bǐ*
 powder pen
 'chalk'
 Mandarin | SN | CASS 2012: 384

B. *chē* 'car, wheeled land vehicle' (CASS 2012: 155)

- (6.212) *huǒ chē*
 fire CT.car
 ‘train’
 Mandarin | SN | CASS 2012: 589
- (6.213) *qì chē*
 vapour CT.car
 ‘car’
 Mandarin | SN | CASS 2012: 1028
- (6.214) *mǎ chē*
 horse CT.car
 ‘carriage’
 Mandarin | SN | CASS 2012: 862

C. *guǒ* ‘fruit’ (CASS 2012: 498)

- (6.215) *píng guǒ*
 apple CT.fruit
 ‘apple’
 Mandarin | SN | CASS 2012: 1004
- (6.216) *wú huā guǒ*
 NEG flower CT.fruit
 ‘fig’
 Mandarin | SN | CASS 2012: 1374

D. *dāo* ‘knife’ (CASS 2012: 263)

- (6.217) *cài dāo*
 vegetable CT.knife
 ‘kitchen knife’
 Mandarin | SN | CASS 2012: 121
- (6.218) *lián dāo*
 sickle CT.knife
 ‘sickle’
 Mandarin | SN | CASS 2012: 807

E. *yú* ‘fish’ (CASS 2012: 1585)

- (6.219) *lǐ yú*
 carp CT.fish
 ‘carp’
 Mandarin | SN | CASS 2012: 796

Table 6.7 Some variable features of personal pronouns across a sample of MSEA languages

Language	Family	1/2/3p distinguished?	dedicated plural	dedicated dual	incl/excl marked	politeness marked	gender marked	case marked
Lahu	TB	yes	no	no	no	no	no	no
Cantonese	SN	yes	no	no	no	no	no	no
Chadong	TK	yes	yes	no	no	no	no	no
Xuyong Hmong	HM	yes	yes	yes	no	no	no	no
Green Miao	HM	yes	yes	yes	no	no	no	no
Jahai	AA	yes	yes	yes	yes	yes	no	no
Kri	AA	yes	yes	yes	yes	yes	yes	no
Mulao	TK	yes	yes	no	no	no	no	no
Phan Rang Cham	AN	yes	yes	no	yes	yes	no	no
Zhuang	TK	yes	yes	no	yes	yes	no	no
Chrau	AA	yes	no	no	yes	no	yes	no
Mandarin	SN	yes	no	no	yes	yes	no	no
Hakka	SN	yes	no	no	yes	yes	no	no
Burmese	TB	yes	no	no	no	yes	yes	no
Khmer	AA	yes	yes	no	no	yes	yes	no

- (6.220) *dài yú*
 ribbon CT.fish
 ‘ribbonfish’
 Mandarin | SN | CASS 2012: 251

- (6.221) *shā yú*
 shark CT.fish
 ‘shark’
 Mandarin | SN | CASS 2012: 1127

6.3 Personal Pronoun Systems

Pronoun systems of MSEA languages vary greatly in size and complexity. Table 6.7 lays out some of the features of MSEA personal pronoun systems discussed in this section. The second to last column of Table 6.7 shows gender marking, which is only rarely found in MSEA pronoun systems, except where gender is implicated in complex systems that mark politeness. A feature that is absent from all MSEA pronoun systems is case marking (as is found, for example, in many European languages, such as English *I/me* or Dutch *ik/mij*).

Given that MSEA languages have only a minimum of inflectional morphology then they might be expected to omit such marking in their pronominal systems too. Accordingly, Li and Thompson (1981: 134) attribute the simplicity of the Mandarin pronoun system to the fact that Mandarin ‘does not have inflection, conjugation, or case markers’. One area in which MSEA languages elaborate their pronoun systems is in politeness and other expression of social difference, as we shall see later in this chapter. But such elaboration is not observed in all MSEA languages. In commenting on the simplicity of the Lahu pronominal system, Matisoff suggests that the lack of pronouns that make ‘distinctions of politeness, or establish relative rank or social distance’ is a reflection of the ‘egalitarian nature of Lahu society’ (Matisoff 2003c: 214).

Among the simplest pronoun systems in the MSEA area are those of Sinitic languages. The Cantonese pronoun system has three core forms, coding singular number in three persons. Plural forms are made by adding a plural suffix. This is shown in Table 6.8.

Table 6.8 *Personal pronouns in Cantonese (SN)*

	SG	PL
1	<i>ngóh</i>	<i>ngóh-deih</i>
2	<i>léih</i>	<i>léih-deih</i>
3	<i>kéuih</i>	<i>kéuih-deih</i>

After Matthews and Yip 1994: 79.

Mandarin and Hakka have the same simple structure as Cantonese, with the addition of an inclusive first person plural form,¹⁰ along with distinctions marking interpersonal familiarity in the first person. See Tables 6.9 and 6.10.¹¹

Table 6.9 Personal pronouns in Mandarin (SN)

	SG		PL	
	familiar to bare	bare to polite	familiar to bare	bare to polite
1	<i>zán</i>	<i>wǒ</i>		
1 & 2			<i>zánmen</i>	
2	<i>nǐ</i>	<i>nín</i>	<i>nímen</i>	<i>nín</i>
3	<i>tā</i>		<i>tāmen</i>	

Adapted, with reference to Li and Thompson 1981: 134.

Table 6.10 Personal pronouns in Hakka (SN)

	SG		PL	
	familiar to bare	bare to polite	familiar to bare	bare to polite
1	<i>en¹</i>	<i>ngai²</i>		
1 & 2			<i>en¹ deu¹</i>	
2				<i>hen² deu¹</i>
3	<i>hen²</i>	<i>gi²</i>		<i>gi¹ deu¹</i>

After He 1993: 14.

Simple pronoun systems can also be observed among Tibeto Burman languages (though Burmese is a notable exception; see below). An example is Lahu, illustrated in Table 6.11. Matisoff writes, ‘Lahu does not distinguish number with common nouns, but pronouns (and proper names) can take the plural suffix *hi*, or dual suffixes like *hi mà* or *hi nè*. There is also an impersonal third person pronoun which cannot be pluralized: *šu* “remote or contrastive third person”, “they”, “others”.’ (Matisoff 2003c: 214).

¹⁰ In the pronoun tables in this section, a greyed-out cell means logically impossible, while a non-greyed blank means logically possible but not attested. Throughout this chapter, when I refer to the presence of an inclusive pronoun meaning ‘we, including you’ (marked by the ‘1 & 2’ row in the table), I leave open the status of the other first person non-singular form or forms. In some systems, the non-inclusive form may entail an exclusive reading, while in others it may be a general first person non-singular meaning. In Mandarin, *wómen* does not encode an exclusive meaning, but it may pick up that meaning in pragmatic opposition with the option of specifying inclusive (with *zánmen* ‘first person inclusive’).

¹¹ The term ‘bare’ in these tables and elsewhere in this section refers to the property of a pronoun of having no semantic element of politeness or respect, in the context of a system in which other forms do have these sorts of meaning elements. ‘Bare’ forms are not necessarily neutral, as they can convey both prosocial and negative meanings in contrast with other terms in a system.

Table 6.11 *Personal pronouns in Lahu (TB)*

	SG	DU	PL
1	<i>ŋà</i>	<i>ŋà-hí-mà</i>	<i>ŋà-hí</i>
2	<i>nò</i>	<i>nò-hí-mà</i>	<i>nò-hí</i>
3	<i>y᷑</i>	<i>y᷑-hí-mà</i>	<i>y᷑-hí</i> <i>šu</i> ('remote/contrastive')

After Matisoff 2003c: 214.

As in the Sinitic languages just illustrated, the Lahu system features transparent morphological marking for non singular number, done by adding a distinct marker for plural number, and in addition, for dual number.

Dual marking is coded in dedicated pronouns in numerous MSEA languages. For example, in Xuyong Hmong (HM, Ruey and Kuan 1962), in which plural number is marked in the first person only, there is a single, dedicated dual term in the third person (note that it is distinguished from the singular form only in tone). See Table 6.12.

Table 6.12 *Personal pronouns in Xuyong Hmong (HM)*

	SG	DU	PL
1	<i>ko</i> (B1)		<i>pe</i> (A1)
2	<i>ka</i> (A2) [SG/PL] <i>ne/me</i> (B2) [SG/PL]		
3	<i>ne/me</i> (B2) [SG/PL]	<i>me</i> (A1)	

Bracketed are tone categories.

After Ruey and Kuan 1962, cited by Ratliff 1992: 103.

In Green Miao, also in the Hmong Mien family, there are two dual pronouns, in first and second person (where dual and plural in the second person are distinguished only by tone). See Table 6.13.

Table 6.13 *Personal pronouns in Green Miao (HM)*

	SG	DU	PL
1	<i>kǔ</i>	<i>ȝw̥</i>	<i>pé</i>
2	<i>kāo</i>	<i>mé</i>	<i>mē</i>
3	<i>nw̥</i>		

After Lyman 1974: 391.

Other languages mentioned in this section showing coding for dual number include Jahai and Kri (see below).

Tai Kadai languages spoken in Sinitic areas tend to have simple personal pronoun systems. Table 6.14 shows the Mulao system, in which singular and plural are marked in first and second person only, and where a first person inclusive form is found.¹²

Table 6.14 *Personal pronouns in Mulao (TK)*

	SG	PL
1	<i>hai¹</i>	<i>niu²</i>
1 & 2		<i>hya:u¹</i>
2	<i>ja²</i>	<i>sa:u¹</i>
3	<i>mo⁶</i>	

After Wang and Zheng 1993: 51.

The Chadong system is almost the same, but with the addition of a number distinction between singular and plural in the third person. See Table 6.15.

Table 6.15 *Personal pronouns in Chadong (TK)*

	SG	PL
1	<i>je²</i>	<i>lje¹</i>
1 & 2		<i>la:u¹</i>
2	<i>ŋi²</i>	<i>ɛe¹</i>
3	<i>mən²</i>	<i>tje¹</i>

After J. Li 2008: 603.

Gender is seldom marked in these simpler MSEA pronoun systems. An exception is in Chrau, which has a male versus female distinction in the second person singular. See Table 6.16.

MSEA pronoun systems mark distinctions in politeness more often than distinctions in gender. These distinctions are useful for social hierarchical stance taking. In Phan Rang Cham, there is a three way distinction in the first person singular, ranging from

¹² The term *inclusive*, which pinpoints the form *hya:u¹* in the Mulao system shown here, will be used in the following sense. The inclusive form means ‘first and second person together’. So, first person plural inclusive means ‘we, including you’. This may contrast with an *exclusive* form in the system (a possible interpretation of *niu²* in the Mulao system shown here), which would mean ‘we, not including you’.

Table 6.16 Personal pronouns in Chrau (AA)

		SG	PL
1		ănh	<i>khananh, khây ănh</i>
1 & 2			<i>võn</i>
2	M	mai	
	F	ay	
3			<i>něh</i>

After Thomas 1971: 138.

familiar to polite to ‘polite slave’, the latter being a strategy in which the self is referred to using the word for ‘slave’ (cf. Brown and Levinson 1978). See Table 6.17.¹³

Table 6.17 Personal pronouns in Phan Rang Cham (AN)

	SG			PL
	familiar	polite	polite-slave	
1	<i>kaw</i>	<i>tàhla?</i>	<i>halun</i>	<i>kami</i>
1 & 2				<i>ita; tray</i>
2			<i>hĩ</i>	
3		<i>ñu</i>		
Other; group			<i>kàw?</i>	

After Thurgood 2005: 499, Table 17.8.

The ‘slave’ strategy for polite first person reference is also used by speakers of Zhuang. This otherwise relatively simple pronoun system shown in Table 6.18 adds a politeness distinction in the first person.

Table 6.18 Personal pronouns in Zhuang (TK)

	SG		PL
	impolite	polite	
1	<i>ku¹</i>	<i>hot²</i> (lit. ‘slave’)	<i>tu¹, po⁴ tu¹</i>
1 & 2			<i>lau⁴</i>
2	<i>muŋ⁴</i>	kinship term	<i>θu¹, po⁴ θu¹</i>
3		<i>te¹</i>	<i>po⁴ te¹</i>

After Luo 2008: 327.

¹³ Marc Brunelle (pers. comm.) writes that these data reflect ‘a more conservative version of Cham’, based on ‘the language of manuscripts written between the seventeenth and the nineteenth century’.

While pronouns can convey social distinctions, such distinctions can sometimes be pragmatically sensitive, and so are sometimes avoided altogether:

The choice of pronouns in Zhuang is sensitive to social factors such as age, social position and attitude of the speaker towards the addressee. As a general rule, the use of the first and second person singular pronoun is to be avoided, in polite conversation with strangers and acquaintances, between relatives or when speaking to a superior. In such situations, a kinship term is used for the addressee and the term *hoi2*, literally ‘slave’, is used for self reference by the speaker. Parents would use kinship terms to refer to themselves when they speak to their children. The use of first person plural (inclusive) is considered polite.

(Luo 2008: 327)

The more complex Jahai system embodies a combination of the distinctions we have reviewed so far, with dedicated dual pronouns in all persons, dedicated inclusive forms in both first person dual and plural, and a three way social distinction between intimate, familiar, and distant in the second person singular. See Table 6.19.

Table 6.19 Personal pronouns in Jahai (AA)

	SG			DU	PL
	intimate	familiar	distant		
1		<i>je?</i>		<i>jeh</i>	<i>japēh~pēh</i>
1 & 2				<i>hej</i>	<i>he?</i>
2	<i>mi?</i>	<i>məh</i>	<i>paj</i>	<i>jih</i>	<i>gin</i>
3		<i>po?</i>		<i>wih</i>	

After Burenhult 2005: 83.

Similarly, Kri includes the range of distinctions we have seen so far, with the addition of a male versus female distinction in the third person singular polite pronouns. See Table 6.20.

Note that the politeness distinction is further refined in the first person singular, with *teeq* and *pààng* having distinct functions in Kri. It is important to note, as Luo did for Zhuang (noted above), that the distinctions listed in the pronominal paradigm do not tell the whole story. Further nuances are also reflected in how the system is used in social interaction. For example, a Kri speaker may use a dual pronoun to refer to a singular person if a kinship relation warrants it. When a man is referring to himself while talking with one of his parents in law, he should use the first person dual exclusive pronoun *ñaar* ‘we two’ (perhaps indexing his status as member of a two person unit with the addressee’s child, i.e., the speaker’s wife). Table 6.21 shows kin categories in Kri that play a role in selection of (or

Table 6.20 Personal pronouns in Kri (AA)

	SG		DU	PL
	bare	polite		
1	<i>teeq/pààng</i>	<i>koon</i>	<i>ñaar</i>	<i>caa</i>
1 & 2			<i>saa</i>	<i>cawq</i>
2	<i>cak</i>	<i>mii</i>	<i>maar</i>	<i>prii'</i>
3	M	<i>qôâng</i>		
		<i>hang</i>	<i>qaar</i>	<i>paa</i>
F		<i>moog</i>		

After Enfield and Diffloth 2009: 56.

avoidance of) specific pronominals in the system, as well as whether use of the referent's name (for example when referring to the referent in the third person) is allowed.

Table 6.21 Some kin categories that are consequential for selection of person reference form in Kri (AA)

Kin category of addressee	1st person pronoun	2nd person pronoun	can use name?
1. <i>cià-maangq</i> 'kin of same descent group' (marriage not allowed)	<i>teeq</i> 1SG	<i>cak</i> 2SG	yes
2. <i>mree</i> "higher" <i>cià-maangq</i> of your spouse'	<i>ñaar</i> 1DU.EX	kin term	no
3. <i>matààm/qujùù</i> 'so, married to your lower <i>cià-maangq</i> (e.g., younger sib. or child)'	<i>pààng</i> 1SG	<i>maar</i> 2DU	no
4. <i>kmoon</i> "lower" <i>cià-maangq</i> of your spouse'	<i>koon</i> 1SG	<i>mii</i> 2SG	yes
5. <i>tààm/sawq</i> 'someone married to your elder sibling'	<i>teeq</i> 1SG	kin term	no
6. <i>sdoong</i> 'mree of your own child' (i.e., someone whose child you are <i>mree</i> to) [these pronouns are used reciprocally]	<i>khoojq</i> 1SG	<i>nôôq</i> 2SG	no

The distinction between 'higher' and 'lower' is grounded in sibling relations; those related to you via elder siblings (e.g., your father's older brother) are 'higher', and *mutatis mutandis* for younger siblings. Note: *cak* cannot be used as 2SG pronoun for higher (i.e., older) *cià-maangq*.

After Enfield and Diffloth 2009.

In addition to the possibilities outlined in Tables 6.20 and 6.21, Kri speakers can also use kin terms not only for address but also for reference. As shown in Table 6.21, this is the preferred option for certain relations. Table 6.22 lists basic kin terms that can be used in Kri as referring expressions for any of the three persons, as relevant. These function like pronouns in everyday conversation, where a person might say, for example, 'Uncle took some of niece's tobacco', where this might be translatable as 'You took some of my tobacco', 'I took some of your tobacco', 'He took some of her tobacco', etc., depending

on who is addressing whom. These are like pronouns in so far as they are an everyday form of tracking reference in discourse, but they differ from true pronouns in that their reference does not shift depending on who is uttering the word.

Table 6.22 *Kri (AA) basic kin terms*

Kinship relation	Form
F	<i>pòðq</i>
M	<i>mééq</i>
eB	<i>maangq</i>
eZ	<i>cìùq</i>
yG	<i>qeem</i>
FyZ	<i>qoo</i>
FyB/HyB	<i>pòðq</i>
MyB/WyB	<i>kìùq</i>
MyZ	<i>muruq</i>
FeZ/HeZ	<i>jaa</i>
FeB	<i>puu</i>
MeZ/WeZ	<i>naaj</i>
MeB/WeB	<i>taa</i>
PP	<i>vòòk</i>
PPP	<i>kôôq</i>
CC/yGC	<i>ciuq</i>
CCC	<i>cec</i>
CCCC	<i>celavec</i>
CCCCC	<i>calavor</i>

Abbreviations are: F father, M mother, B brother, Z sister, G sibling, H husband, W wife, P parent, C child, y younger, e elder.

After Enfield and Diffloth 2009: 58.

This use of kin terms for person reference with functional equivalence to pronouns points to a means for extending a pronominal system indefinitely. This is the essence of the elaborate systems of person reference observed in some prominent MSEA languages, including Burmese, Khmer, Thai, and Lao. The historical/political status of these languages is unlike that of many of the languages we have reviewed in this section: they are major national languages with historically literate and literary cultures, attached to major religions.¹⁴ An example is Khmer, whose highly elaborated system of person reference is summarized in Table 6.23.

¹⁴ Being a large literary language does not, of course, mean that a large and complex politeness-based pronominal system is found. In Vietnamese, for example, there is no elaborated system of dedicated lexical forms, unlike the languages we are about to discuss (though Vietnamese does rely heavily on the use of kinship terms in the same manner just described for Kri; see Sidnell and Shohet 2013, Sidnell 2020).

Table 6.23 *Khmer (AA) personal reference paradigm*

Person	Form	Comments
1	<i>kñom</i>	General, polite; literal meaning ‘slave’
	<i>qañ</i>	Between intimate friends, superior to inferior, or adult to child; otherwise insulting
	<i>yæŋ</i>	Singular between equals in some dialects; otherwise plural
	<i>kniə</i>	Singular, among intimates
	<i>kñom-præh-baat</i>	Inferior to superior, respectful or extremely formal
	<i>kñom-kñaa</i>	Layman to priest; inferior to superior of exalted rank; commoner to king
	<i>qaatmaa</i>	Priest to layman
	<i>look</i>	Masculine, formal, polite; inferior to superior; between equals of relatively high status
	<i>look-sray</i>	Feminine, formal, polite; inferior to superior; between married women of relatively high status
	<i>něq</i>	Superior to inferior; older to younger; between equals of relatively low status; to one’s mother
2	<i>něq-srey</i>	Feminine, polite; superior to inferior; between equals
	<i>qaen</i>	Reflexive connotation; between intimates; superior to inferior; otherwise insulting; usually paired reciprocally with <i>qañ</i> ‘I’
	<i>qaa-qaey</i>	Derogatory, good-natured insult between friends; adult to child
	<i>præh-dac-præh-kun</i>	Inferior to superior of exalted rank; layman to priest
	<i>koðt</i>	Respectful; younger of older; inferior of superior
3	<i>kee</i>	Informal; other; indefinite (generic)
	<i>wið</i>	Superior of inferior; adult to child; otherwise insulting (‘It’)

Adapted, with reference to Huffman 1970: 356 7.

Burmese similarly has a rich array of forms, with many subtle social distinctions that speakers must learn to navigate. These are shown in Table 6.24. Similar systems to Khmer and Burmese are found in Thai and Lao (see Cooke 1968, Enfield 2015: 133–46). The complexities of their usage go beyond our scope here.

As a final note in this section on personal pronouns, mention can be made of inclusory constructions. These constructions combine a non singular pronoun and a personal name to refer to a group of people who are together with, or associated with, or who otherwise form a unit with, the one person who is explicitly referred to by name. Here are examples from Kri, showing pronouns in inclusory constructions of the type PRONOUN+[TITLE+NAME]:

- (6.222) *paa vòök nun*
 3PL grandpa N
 ‘they of whom Grandpa Nun is one’ (or: ‘Grandpa Nun and the others’)
 Kri | AA

Table 6.24 Burmese (TB) personal reference paradigm

Person	Form	Literal translation	Sex of speaker	Degree of politeness, etc.
1	<i>nga</i>		either	familiar, impolite
	<i>kou</i>	'body, self'	either	fairly polite
	<i>cou'</i>	'worthless slave'	either	fairly polite in Upper Burma, impolite in Lower Burma
	<i>cuñto</i>	'royal slave'	M	polite
	<i>cuñmá</i>	'female slave'	F	polite
	<i>tāpyíto</i>	'honourable disciple'	M	speaking to monks
	<i>tāpyítomá</i>	'female honourable disciple'	F	speaking to monks
	<i>aiñ</i>	'I' (English)	either	university slang, mainly among women
2	<i>niñ</i>		either	familiar, impolite
	<i>mìñ</i>	'king'	either	familiar, impolite
	<i>nyì</i>		F	familiar, impolite
	<i>to</i>		F	familiar, impolite
	<i>hkiñbyà</i>	'lord, master'	M	polite
	<i>hyiñ</i>	'lord, master'	F	polite
	<i>mei'hswei</i>	'friend'	either	polite
	<i>ăhyiñhpăyà</i>	'lord'	either	speaking to monks
	<i>hsáyahkălei</i>	'little teacher'	either	speaking to nuns
	<i>yu</i>	'you' (English)	either	university slang, mainly among women
3	<i>thu</i>	'he, she', sometimes 'it'		
	<i>thumá</i>	'she'		

After Okell 1969: 100 1.

- (6.223) *qaar qaj kham*
 3DU 3.NON.RESP K

'those two of which Kham is one' (or: 'Kham and the other one')
 Kri | AA

- (6.224) *ñaar saaw knin*
 1DU.EXCL aunt K

'we two (exclusive) of which the other is Aunt Knin' (or: 'me and Aunt
 Knin')
 Kri | AA

Inclusory constructions of this kind have also been noted for Lao (Enfield 2007: 82).

6.4 Demonstrative Systems

Every language has demonstrative pronouns. These are nominal forms such as *this* and *that* in English that refer deictically to people and things (Dixon 2010: 223ff.; Levinson *et al.* 2018). They identify things in terms of their accessibility (in an intentionally vague sense: ‘accessibility’ might be construed in spatial, conceptual, or other terms). When we refer to something as ‘this’, we mean that it is more readily accessible than something we refer to with ‘that’. It may be regarded as more accessible, for example, because it is physically closer to the speaker. In most languages of the world, the number of demonstratives is two or three. A two term system will often distinguish between something that is near the speaker versus something that is far from the speaker. An example is English *this* and *that*. Many languages add a third term referring to something that is further away still—the dated English word *yon* has this sense in contrast with *this* and *that*. A good number of languages add further terms: these do not make further distinctions in distance, but tend to invoke other semantic features of accessibility, such as spatial elevation, visibility, and focus of attention.¹⁵

The following sections survey demonstrative systems according to the number of distinctions made.

6.4.1 Two Term Systems

Speakers of Sinitic languages tend to use relatively simple systems, with two terms: proximal and distal. For example, Cantonese in Table 6.25.

Table 6.25 Cantonese (SN)
demonstrative pronouns

<i>nī</i>	‘this’
<i>gó</i>	‘that’

After Matthews and Yip 1994: 89.

In Mulao, a Tai language spoken with a history of intensive contact with Sinitic languages, two demonstratives are described, as shown in Table 6.26.

¹⁵ The statements made in this section about the semantic bases of distinctions in demonstrative systems should be treated with caution. Few descriptions of these systems distinguish conclusively between the typical situations in which these demonstratives may be used versus the concepts coded in the semantics of these terms. The fact, for example, that one demonstrative tends to be used for referring to objects that are further away does not mean that the coded meaning of the demonstrative includes the concept ‘far away’. In most cases, it is clear how many distinctions an inventory makes, but because most descriptions of demonstrative systems are brief and cursory, consisting often of no more than proposed labels for the observed distinctions, it is difficult to evaluate the precise semantic content of the distinctions, or the contextual conditions in which they are used. Exceptions include Lao (Enfield 2003) and Jahai (Burenhult 2003, 2008).

Table 6.26 *Mulao (TK)*
demonstrative pronouns

<i>na:i⁶</i>	‘this’
<i>ka⁶</i>	‘that’

After Wang and Zheng 1993: 52.

MSEA demonstratives always appear as modifiers of head nominals in noun phrases, but not all languages allow them to occur as independent nominals on their own. Wang and Zheng (1980: 52) note this constraint for Mulao: ‘The two demonstrative pronouns, *na:i⁶* and *ka⁶*, are generally used as nominal modifiers only. They cannot stand alone as subject or object of a sentence.’ Similarly, Zhuang has two generic demonstrative adjectives which ‘are seldom used alone to function as demonstrative pronouns’ (Luo 2008: 327). The two Zhuang forms are given in Table 6.27.

Table 6.27 *Zhuang (TK)*
demonstrative pronouns

<i>ni⁶</i>	‘this’
<i>te¹</i>	‘that’

After Luo 2008: 327.

Among Austronesian languages, Hainan Cham and Phan Rang Cham have two term systems, shown in tables 6.28 and 6.29, respectively.

Table 6.28 *Hainan Cham (AN)*
demonstrative pronouns

<i>ni³³</i>	‘this’
<i>nan³³</i>	‘that’

After Thurgood *et al.* 2014: 200.

Table 6.29 *Phan Rang Cham (AN)* demonstrative pronouns

<i>ni</i>	‘this’
<i>nan</i>	‘that’

After Thurgood 2005: 499.

While the contrast in these two term systems is usually described in exophoric/spatial terms, as a contrast between proximal and distal for referring to objects that are near or far, one or another (or both) of the terms may also be used endophorically, for cohesion and coherence in discourse (Halliday and Hasan 1976). Languages differ as to which term is used in anaphoric or other discourse functions. Thurgood (2005: 499) notes for Phan Rang Cham that ‘the distal demonstrative *nan* is the one that has taken up the duties involved in marking NPs as anaphoric; the proximal *ni* is only used when a proximal meaning is emphasized’.

White Hmong has a simple two way set of forms (by contrast with other Hmong Mien languages that have three term systems – see below), shown in Table 6.30.¹⁶

Table 6.30 *White Hmong (HM) demonstrative pronouns*

<i>no</i>	‘this’
<i>ntawd</i>	‘that’

After Mottin 1978: 47.

In Lao, a basic two term system for demonstrative adjectives (which can directly modify nominal referents) is enriched with a three term set of adverbial forms. These are shown in tables 6.31 and 6.32.

Table 6.31 *Lao (TK) demonstrative determiners*

<i>nii4</i>	general ('this')
<i>nan4</i>	nonproximal ('that')

After Enfield 2007: 100.

Table 6.32 *Lao (TK) demonstrative adverbs*

<i>phii4</i>	proximal ('here')
<i>han5</i>	distal ('there')
<i>phun4</i>	far distal ('yonder')

After Enfield 2007: 100.

¹⁶ There is a derivational relationship between the demonstrative noun *ntawd* (with the *-d* tone) and a preposition *ntawm* (with the *-m* tone). (See Ratliff 1992 : 104 22.)

6.4.2 Three Term Systems

Semelai has two basic demonstratives, which are ‘based on simple distance orientation from the speaker, or the speaker’s adopted point of reference’ (Kruspe 2004: 192). In addition, there is a third term, ‘a proximate deictic locative preposition which expresses a non specific proximate location’ (Kruspe 2004: 192), see Table 6.33.

Table 6.33 *Semelai (AA) basic demonstratives*

<i>ŋnɔŋ~nɔŋ</i>	‘this’
<i>ke</i>	‘that’
<i>ha?</i>	‘in the immediate vicinity, but not at any particular point within it’

After Kruspe 2004: 192.

In Sedang, the three forms are described by Smith (1979: 85), as shown in Table 6.34.

Table 6.34 *Sedang (AA) demonstratives*

<i>kó</i>	close	‘this, here’
<i>me</i>	intermediate	‘that’
<i>tá</i>	distant	‘that’

After Smith 1979: 85.

In Hakka there are three distinctions in spatial demonstratives, shown in Table 6.35.

Table 6.35 *Hakka (SN) demonstratives*

<i>l̄i³</i>	proximal	‘this’
<i>gai²</i>	quasi-distal	‘that’
<i>gun²</i>	distal	‘that’

After He 1993: 16.

Tables 6.36 and 6.37 show two examples of Hmong Mien languages with three term systems.

Table 6.36 *Aizhai Miao (HM) demonstratives*

<i>nəŋ⁴⁴</i>	‘this (close to speaker)’
<i>ei⁵³</i>	‘that (far from speaker and addressee)’
<i>a⁴⁴</i>	‘that (far from speaker, close to addressee)’

After Yu 2010: 53.

Table 6.37 *Green Miao (HM) demonstratives*

<i>núa</i>	‘this’
<i>hǎo, kǎo</i>	‘that (far)’
<i>ndǎw</i>	‘that’

After Kunyot 1984: 74.

6.4.3 Four Term Systems

Four term systems are less frequently found. In the Chrau system, the spatial meaning of ‘far’ is divided into two, as shown in Table 6.38.

Table 6.38 *Chrau (AA) demonstratives*

<i>heq</i>	‘this’
<i>nōq</i>	‘that’
<i>toq</i>	‘that (farther)’
<i>tiq</i>	‘that (far)’

After Thomas 1971: 139.

Two further examples of four term systems are from the Tibeto Burman language family. In Akha, demonstratives have a three way distinction for distance, where the differences are described as person anchored: (1) close to me, (2) close to you, (3) away from both of us. See Table 6.39. As Hansson (2003: 244) describes it, ‘there is a fourth demonstrative pronoun, only occurring with a suffix for space and usually pronounced with an exaggerated high, prolonged vowel’.

Table 6.39 *Akha (TB) demonstratives*

<i>hə</i>	‘here (close to me)’
<i>thə</i>	‘there (close to you)’
<i>xhó</i>	‘away from both of us’
<i>xhü ba ~ hü ba</i>	‘far over there’

After Hansson 2003: 244.

In Lai, shown in Table 6.40, a fourth term is used when the referent is not visible.

Table 6.40 *Lai (TB) demonstratives*

<i>hi:</i>	'this, near me'
<i>kha:</i>	'that, near you'
<i>khi:</i>	'that, over there (visible)'
<i>cu:</i>	'that, over there (not visible)'

After Bedell 2001: 157.

6.4.4 Five Term Systems

An example of a demonstrative system with five spatial distinctions is Lahu. In this system, shown in Table 6.41, the three term distinction between proximal, distal, and far is further augmented with terms that refer to elevation, one for reference to things 'up there', another for things 'down there' (see also Matisoff 1973a: 51).

Table 6.41 *Lahu (TB) demonstratives*

<i>chò</i>	'here'
<i>ô</i>	'there'
<i>cô</i>	'way over there; yonder'
<i>nô</i>	'up there'
<i>mô</i>	'down there'

After Matisoff 2003c: 216.

The same system is found in Kri. The 'up there' versus 'down there' distinctions are clearly associated with the upstream/downstream orientation of life for these upland riverine people. See Table 6.42.

Table 6.42 *Kri (AA) demonstratives*

<i>niì</i>	general ('this')
<i>naaq</i>	external ('that', away, far)
<i>seeh</i>	distal ('yon', across, far)
<i>lééh</i>	external up, upstream, above
<i>côôh</i>	external down, downstream, below

After Enfield and Diffloth 2009: 59.

6.4.5 Larger Systems

We finish this section with the Jahai demonstrative system, which has eight distinctions, encoding ‘accessibility, exteriority and elevation of locations in relation to the speaker and the addressee’ (Burenhult 2005: 84). Table 6.43 lists the Jahai forms and their glosses. Burenhult’s treatment of Jahai provides unusually detailed insight into the semantic and pragmatic distinctions between demonstratives within a system. Table 6.44 outlines the distinctions in more detail.¹⁷

Table 6.43 *Jahai (AA) demonstratives*

/?əh/	speaker-anchored accessible	‘here’
/?on/	addressee-anchored accessible	‘there (you know)’
/?əni?/	speaker-anchored inaccessible	‘there (away)’
/?ün/	addressee-anchored inaccessible	‘there (you don’t know)’
/?adeh/	speaker-anchored exterior	‘there (beyond me)’
/?ŋi?/	addressee-anchored exterior	‘there (beyond you)’
/?itih ~ ?otih/	superjacent	‘there (up)’
/?ujih/	subjacent	‘there (down)’

After Burenhult 2005: 84 6.

Table 6.44 *The semantics and pragmatics of Jahai (AA) demonstrative categories*

Category	Meaning and function
Speaker-anchored accessible	‘This distinction is associated with referents conceived of as in some way accessible to the speaker, e.g., with regard to their proximity, perceptibility, reachability/approachability, possession and topicality in discourse. Speaker-proximal location is typical.’ (Burenhult 2005: 85)
Addressee-anchored accessible	‘This distinction is associated with referents considered by the speaker to be “cognitively accessible” to the addressee, i.e., referents which have the addressee’s current or previous attention/knowledge. Proximity to addressee is common.’ (Burenhult 2005: 85)
Speaker-anchored inaccessible	‘This distinction is associated with referents conceived of as inaccessible to the speaker, e.g., with regard to distance, imperceptibility, unreachability etc. Speaker-distal location is typical’ (Burenhult 2005: 85)

¹⁷ ‘The allomorphs /?itih ~ ?otih/ are in free variation and idiolectally determined’ (Burenhult 2005: 86).

Table 6.44 (*cont.*)

Category	Meaning and function
Addressee-anchored inaccessible	'This distinction is associated with the introduction of new referents, i.e., referents which do not have the addressee's current or previous attention/knowledge and therefore are "cognitively inaccessible" to the addressee. There is no typical spatial patterning of referents; location is flexible' (Burenhult 2005: 85)
Speaker-anchored exteriority	'This distinction is associated with referents located on the other side of the speaker from the addressee's position; distance is irrelevant' (Burenhult 2005: 85)
Addressee-anchored exteriority	'This distinction is associated with referents located on the other side of the addressee from the speaker's position; distance is irrelevant' (Burenhult 2005: 86)
Superjacent	'This distinction is associated with referents located above the speech situation, either in the immediate area of the speech situation (including vertically above) or with reference to landscape contour (uphill) or river profile (upstream).' (Burenhult 2005: 86)
Subjacent	'This distinction is associated with referents located below the speech situation, either in the immediate area of the speech situation (including vertically below) or with reference to landscape contour (downhill) or river profile (downstream).' (Burenhult 2005: 86)

After Burenhult 2005: 85–6.

6.4.6 Comment on Demonstrative Systems

The family affiliation of a language does not appear to be a good predictor of the size and semantic makeup of the language's inventory of demonstratives in MSEA. Possible determining factors are community size and degree of proximity to natural environments in day to day life (Denny 1978, Perkins 1992). If a community is smaller, and lives closer to nature, a larger demonstrative inventory appears to be more likely. These two conditions hold in MSEA in the case of minority languages of upland areas, and these languages feature the larger, more detailed systems: for instance, Kri, Lahu, and Jahai, noted above. These three languages are not (closely) historically related, and they are spoken far from each other, but they are each spoken by small communities who live and work mostly outdoors in hilly environments.

Predication and Clausal Syntax

7.1 Marking Syntactic Functions

Recall at the beginning of Chapter 3, I introduced three ways in which languages can grammatically mark whether a participant in an event or situation is in the A, S, or O syntactic function. These were (1) constituent order (e.g., in English, A/S arguments precede the verb, while O arguments follow it), (2) case marking (e.g., in English, the third person plural pronoun is *they* for A/S arguments but *them* for O arguments), and (3) agreement (e.g., in English, verbs in the present simple tense are marked with *s* if the A/S argument is singular, while number of the O argument is not marked).

Let us start with verb agreement. There is virtually nothing in the way of verb agreement in the MSEA area. Kruspe does not use the term ‘agreement’ but uses the related term ‘cross referencing’ in her description of transitive clauses in Semelai. In Semelai, A arguments are obligatorily marked with a proclitic *la* , signalling their grammatical role as A (Kruspe 2004: 259). (O arguments are optionally marked with a proclitic *hn* .) On the verb, a proclitic marks the A argument, distinguishing in terms of person, number (with a minimal/augmented distinction in first and second person and a singular/plural distinction in third person), and familiarity (in first and second person) (Kruspe 2004: 171). Here are two examples, with the proclitic verbal marker underlined:

- (7.1) *ki grn la lslps*
 3SG A bite A flying.ant
 ‘The flying ant bit (him).’
 Semelai | AA | Kruspe 2004: 259
- (7.2) *dom, de ca la deh*
 AFF 3PL.A eat A 3PL
 ‘Indeed, they ate (it).’
 Semelai | AA | Kruspe 2004: 259

This type of system is unusual in MSEA languages. Very few languages of the area show any form of cross referencing or case marking. Exceptions can be found among Tibeto Burman languages of western Myanmar and northeast India. Jinghpaw is one of a group of these languages that feature a clause final verbal auxiliary complex, in which agreement in person and number is marked in verbal prefixes. Here is an example, with the first person singular agreement marker *ijy* prefixed to the auxiliary declarative marker in final position (see Kurabe 2017: 1003ff. for details and discussion):

- (7.3) *yay ei phé? gùmrój yà ijy ay*
 1SG 3SG ACC boast CONT 1SG DECL
 ‘I had boasted to him (about you).’
 Jingphaw | TB | Kurabe 2017: 1005

We also noted at the end of Chapter 3 that many languages of the world use voice systems to manipulate A, S, and O (as, for example, in the case of a passive derivation, which transforms the O of a transitive clause into an S and demotes the erstwhile A by expressing it in an adjunct or removing it altogether). Mechanisms and constructions under the rubric of grammatical voice are widespread in the world’s languages, and their vast complexities are the topic of decades of typological work and countless research publications. Yet the hundreds of languages of MSEA have little if anything that can be described as grammatical voice.

Thompson (1987: 217) writes the following on Vietnamese: ‘An important way in which Vietnamese verbs differ from English verbs is that they do not in themselves imply a clear notion of “voice” in the grammatical sense.’ He gives this example of the verb *làm* ‘do, make’ with two arguments (the first of which is elided in the example):

- (7.4) *Làm việc này rồi.*
 do work this already
 ‘We’ve already done this work.’
 Vietnamese | AA | Thompson (1987: 218)

Then he notes that the same verb can be used with a single argument. What was previously the (postverbal) O in the two argument construction is now the (preverbal) S in a one argument construction:

- (7.5) *Việc này làm rồi.*
 work this do already
 ‘This work has already been done.’
 Vietnamese | AA | Thompson (1987: 218)

The English translation given by Thompson is in the passive voice. Here are more examples, this time with the verb *cát* ‘build’. First, with two arguments:

- (7.6) *Ông ấy cát nhà ở gần trường học.*
 sir that build house be.at near school
 ‘He’s building a house near the school.’
 Vietnamese | AA | Thompson (1987: 218)

Second, with one argument, again with the erstwhile O argument now in S position, and with no explicit marking of any derivation:

- (7.7) *Hai cái nhà kia cát gần nhau cùng trên một miếng đất.*
 two CLF house that build near RCP together on one CLF lot
 ‘Those two houses are built close together on the same lot.’
 Vietnamese | Thompson (1987: 218)

This pattern of valency alternation is achieved simply by the use of verbs as ambitransitive (S O, in this case), and it means that there is little need for voice alternations such as a passive construction would provide. That said, there is a kind of explicit construction that has been identified with passive, known as the adversative passive (Clark 1989).

In Lao, this so called passive is marked by the verb *thùùk5* ‘strike, come into contact with’. Here is the verb used in a simple transitive clause:

- (7.8) *khòòj5 thùùk5 toq2*
 1SG.P strike table
 ‘I bumped into the table.’
 Lao | TK | Enfield 2007: 439

The following example shows its ‘adversative passive’ function, as a sort of complement taking predicate where the complement is a clause that describes what happens to the main subject:

- (7.9) *thùùk5 khaw3 ñing2 taaj3*
 strike 3PL.B shoot die
 ‘(You would) get shot by them.’
 Lao | TK | Enfield 2007: 439

This example only incidentally serves the passive like voice function of rearranging the argument structure of the clause, where the subject of the main clause is the undergoer of the marked/subordinate verb.

In other examples, the derived construction does not rearrange the assignment of syntactic functions. In the following example, the subject of the main clause has the role of actor, not undergoer, in the marked predicate:

- (7.10) *haw2 tòòng4 thùùk5 son2*
 1.FA OBLIG strike fight

‘I would be made to fight.’; ‘It would fall to me to fight.’

Lao | TK | Enfield 2007: 440

The same kind of pseudo passive constructional alternation is found in Western (Cambodian) Cham. First, here is a simple transitive clause using the verb *khan* ‘tell’:

- (7.11) *Sohput khan hlün*
 friend tell 1SG.LORESP
 ‘A friend told me.’

Western (Cambodian) Cham | AN | Baumgartner 1998: 5

Next, here is a derived passive like structure in which the undergoer of the verb meaning ‘tell’ is now the main subject:

- (7.12) *Hlün djauk sohput khan*
 1SG.LORESP ‘passive’ friend tell
 ‘I was told by a friend.’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 5

The so called passive marker *djauk* can be used before or after the sentential subject of the subordinate clause. It also means ‘must’ when the main subject remains in the role of actor in relation to the marked verb (as shown for Lao in example 7.10). This is illustrated in the following example:

- (7.13) *Yah rók nao, bóh dráp kau,*
 if dig away see thing 1.HIRESP
hur djauk ta bray ni ka kau wok
 2.LORESP must only to.give this to 1.HIRESP back
 ‘If you go dig it up and find my things, you have to give it back to me.’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 13

In Khmer, the so called passive marker is also derived from a verb to ‘strike’, which in turn can mean ‘correct, on target’ (‘acquire’ verbs are also used in some cases; see Enfield 2003). The next two examples show the Khmer verb *trəw* used as a main verb:

- (7.14) *camlaøy look min trəw tee*
 answer 2HON NEG correct NEG
 ‘Your answer isn’t correct.’
 Khmer | AA | Huffman 1970: 301

- (7.15) *wiə̯ crəluə̯h mɔ̯ɔ̯k trə̯w cəŋkuɔ̯ŋ wi̯ñ*
 3 slip come contact knee return
 'It slipped and (came and) hit my knee.'

Khmer | AA | Huffman 1970: 302

The following example shows it being used as a so called passive marker, where the main subject is an undergoer of the *trə̯w* marked predicate:

- (7.16) *pi̯i msə̯l mə̯ñ kñom trə̯w ckae kham*
 yesterday 1 contact/undergo/suffer dog bit
 'Yesterday I was bitten by a dog.' (I was subjected to a dog biting.)

Khmer | AA | Huffman 1970: 302

But as the next examples show, this reversal of roles between main subject and subordinate subject is not a necessary function of *trə̯w*, which here has a modal meaning 'must':

- (7.17) *kñom trə̯w ti̯w psaa tñay nih.*
 1 must go market day this
 'I have to go to the market today.'

Khmer | AA | Huffman 1970: 302

- (7.18) *koun trə̯w sli̯q pe̯ɔ̯q qaoy sqaat baat*
 child must dress give/cause clean/tidy
 'You must dress neatly, child.'

Khmer | AA | Huffman 1970: 302

These examples show that what is sometimes called passive in MSEA languages does not have the dedicated argument structure function that a passive alternation has in a language such as English – that is, to promote an O argument to (intransitive) subject position of a main clause. Such manipulation of syntactic prominence is achieved in MSEA languages often by other means, including ellipsis of arguments, freedom of pragmatically determined argument movement, and versatility in the valency of verbs. See Chapter 3, and the rest of this chapter, for more on this.

The second general problem relating to verbs that the grammar of any language will help to solve can be termed the grounding problem. All languages will provide forms of grounding that characterize the relationship between (1) the target proposition, that is, the event, state, or property that the verb describes, and (2) the context of speaking, that is, the time and place of speaking, and the identity and knowledge/belief state of the speaker and hearer. (In Jakobson's 1971 terms, these are the 'narrated event', or En, and the 'speech event', or Es.) The fundamental categories of meaning that are dealt with in the world's languages in relating the target proposition to the context of speaking include those listed in Table 7.1.

Table 7.1 Some idealized grammatical categories of verbal marking that ground a proposition by linking (1) the ‘narrated event’* to (2) the coordinates of the speech event

(1) Property of target event	(2) Property of speech event	Verbal marking relating (1) to (2)
(1) when target event occurs	(2) when speaking event occurs	tense
(1) whether target event is true	(2) how certain speaker is of (1)	epistemic
(1) whether target event is allowed	(2) whether speaker grants that (1)	modality
(1) whether target event is true	(2) what evidence speaker has for (1)	deontic modality
		evidentiality

* i.e., the target event or situation being spoken about.

Meanings that fall into the categories shown in Table 7.1 and that express combinations of meanings in these categories are often marked directly on verbs in the world’s languages. (This does not mean that all languages mark all of these meanings, nor that they necessarily do so with direct morphological marking on verbs.) Also marked directly on verbs are certain other types of information. An important one of these is status, including negation and forms of contrast or emphasis. Another is aspect, which is often conveyed by means of markers that also include tense or tense like information. We will see some of these below.

Before we look at some of the kinds of verbal marking found in MSEA languages, we can first note what is *not* found in the area. MSEA languages do not have dedicated systems of evidential marking (although some forms of aspectual marking can convey or acquire evidential meanings by inference, and as we noted in Chapter 2, some sentence final particles have elements of evidential meaning).¹ Nor does any MSEA language show tense marking in the usual sense of that term, that is, where an obligatory and explicit form of morphological marking specifies whether the event or situation being referred to occurs before, at the same time as, or after the event of speaking. (As we shall see below, there is widespread marking for aspectual distinctions, which often imply temporal relations to the speech event.)

As a rule, in MSEA languages, if a speaker wants to describe an event that takes place in the past, this does not need to be marked at all. Depending on the context, the following example from Kri has multiple readings:

¹ Evidential systems are found on the fringes of our area and beyond, notably in some Tibeto-Burman subgroups to the north and west of MSEA, including Qiangic and Tibetan.

- (7.19) *teeq qeec baang mooq*
 1SG go village 3SG.F.POL
 i. ‘I went to her village.’
 ii. ‘I am going to her village.’
 iii. ‘I will go to her village.’
 Kri | AA

The following examples show that explicit marking of the time of the event, by words like ‘now’, ‘tomorrow’, and ‘yesterday’, is compatible with the same non tense marked verb phrase:

- (7.20) *cicòòr teeq qeec baang mooq*
 now 1SG go village 3SG.F.POL
 ‘Now I am going to her village.’
 Kri | AA

- (7.21) *kucook teeq qeec baang mooq*
 tomorrow 1SG go village 3SG.F.POL
 ‘Tomorrow I will go to her village.’
 Kri | AA

- (7.22) *kudéê teeq qeec baang mooq*
 yesterday 1SG go village 3SG.F.POL
 ‘Yesterday I went to her village.’
 Kri | AA

As these examples show, no explicit marking of tense on the verb is necessary for a listener to be able to understand correctly the temporal relationship between the target event and the event of speaking.

7.2 Verbal Marking

In the following sections, we review a range of types of grammatical meaning that tend to be marked on verbs in MSEA languages. These are grouped under the headings of negation, aspect, and modality.

7.2.1 Negation

Negation of a verb is marked in most MSEA languages by a morpheme that directly precedes the verb. In a number of MSEA languages, this preverbal negator is analysed as an affix, as in the following examples from Austroasiatic languages:

- (7.23) *kə wih*
 NEG to.go.home
 ‘not go home’
 Bahnar | AA | Alves 2015: 549 | Banker 1964: 116

- (7.24) *k miən*
 NEG to.have
 ‘not have’
 Khmer | AA | Alves 2015: 549 | Huffman 1970: 412

- (7.25) *?a ?ijn*
 NEG want
 ‘hate’
 Pacoh | AA | Alves 2015: 549 | R. Watson 2011: 229

In many other MSEA languages, negation is marked by a free morpheme that occurs immediately before the verb:

- (7.26) *ji^[13] və? tɔŋɛɪŋ^[1 5 1]*
 he not careful
 ‘He is not careful.’
 Shanghainese | SN | Zee and Xu 2003: 143

- (7.27) *mi³³ pu³³ na:w³³ ɿa³³*
 we NEG go DECL
 ‘We did not go.’
 Hainan Cham | AN | Zheng 1997: 92, cited by Thurgood *et al.* 2014: 208

- (7.28) *ɳa¹¹ ɳay¹¹ kwe¹¹ hay¹³ pai⁵⁵ hi¹¹*
 you really NEG willing go FW
 ‘You really don’t want to go (I’m surprised).’
 Kam | TK | Yang and Edmonson 2008: 534

- (7.29) *koj tsis muaj [chaw rau peb nyob] ...*
 2SG NEG have place for 1PL dwell
 ‘... you do not have a place for us to stay ...’
 White Hmong | HM | Jarkey 2015: 58

When this preverbal negator becomes phonologically dependent on the verb that it marks, it may be analysed as a clitic, as in this Lao case:

- (7.30) *kuu3 bo^o sìù4*
 1SG.B NEG buy
 'I do/will not buy (it).'
 Lao | TK

Other languages depart from this pattern, showing final negation. In Pwo Karen, a main clause can be negated by a clause final particle *?e*:

- (7.31) *?əwe. khlain chəkhlain xe xe ?e:*
 3SG speak language slowly NEG
 'He does not speak slowly.'
 Pwo Karen | TB | Kato 2003: 640

Pwo Karen also shows the preverbal particle pattern of negation, with use of the preverbal particle *lə* in subordinate clauses, though this is combined with the particle *ba*: after the verb or in the clause final position. Here is an example (with the markers of negation underlined):

- (7.32) *?əwe. ?e lə ?an: mi ba: nɔ:, jə mə ?an:*
 3SG if NEG eat rice NEG that 1SG IRR eat
 'If he doesn't eat the rice, I will eat it.'
 Pwo Karen | TB | Kato 2003: 640

Numerous other MSEA languages feature double marked negation, with a marker both before the verb and at the end of the clause (Vossen 2011). Here are some examples (with the markers of negation underlined):

- (7.33) *iim, ay tha khīay iim*
 no, I NEG1 accustomed NEG2
 'No, I'm not in the habit of going.'
 Brao | AA | Keller 1976: 69, cited by Vossen 2011: 8

- (7.34) *amā buh nāo paq apu oh*
 father NEG1 go to rice.field NEG2
 'Father didn't go to the rice field.'
 Roglai | AN | Lee 1996: 293, cited by Vossen 2011: 8

- (7.35) *kāo bu homāo prāk ôh*
 I NEG1 have money NEG2
 'I don't have any money.'
 Jarai | AN | Lee 1996: 302, cited by Vossen 2011: 8

- (7.36) *aw big loq oh*
 1 SG NEG1 know NEG2

'I don't know.'

Rengao | AA | Gregerson 1979: 54, cited by Vossen 2011: 8

7.2.2 Aspect and Related Marking

Many grammatical markers in MSEA languages convey meanings that frame events or situations in terms of their inherent temporal properties, often as these relate to the speech situation. For example, in Semelai, the verb *cəŋ* 'pass through' can be used in combination with a main verb to convey the idea that an event happened without delay. Here are two examples:

- (7.37) *cəŋ jtek ?a kəh*
 pass.through sleep DET 3
 'As for him, he immediately slept.'

Semelai | AA | Kruspe 2004: 365

- (7.38) *ki cəŋ jəl la cɔ dey ke*
 3A pass.through bark.at A dog like that
 'The dog immediately barked at (it) like that.'

Semelai | AA | Kruspe 2004: 365

In Lao, the particle phrase *haa3 kòò1* is used preverbally to convey the idea that an event just happened, in the very recent past:

- (7.39) *khaw3 haa3 kòò1 huu4*
 3PL.B PAST.RCNT know
 'They just found out.'

Lao | TK

The marker in this last example emphasizes the short amount of time that has elapsed between the narrated event and the speech event (thus clearly having an element of tense in their meaning; Enfield 2007: 204 5). Many other verbal markers will emphasize the completion of the event with no indication of length of time elapsed, nor necessarily of their relation in time to the speech event. These are concerned with a more purely aspectual distinction. (On aspect as a grammatical category, see Comrie 1985, Dixon 2010.)

In Semelai, the marker *kmɔŋ*, from a verb meaning 'to finish', can be placed before the verb to indicate completion of the event or situation predicated by the main verb, with a clear implication of current relevance of the resulting state:

- (7.40) *ʔmaðbapa? kmɔŋ kʰbəs*
 parents finish be.dead
 '(Her) parents were (both) dead.'

Semelai | AA | Kruspe 2004: 366

- (7.41) *rɔti? da? da? wɔ?. kmɔŋ br ca*
 bread NEG EXIST longer finish MID eat
 'There's no more bread. (It's) all been eaten.'

Semelai | AA | Kruspe 2004: 366

As the following example shows, Semelai speakers may also place the marker after the verb, in which case it conveys the idea of 'totally' or 'completely':

- (7.42) *ki ca kmɔŋ la pɔðŋŋ*
 3A eat totally A tiger
 'The tiger ate (it) all up.'

Semelai | AA | Kruspe 2004: 367

MSEA languages often have multiple ways of marking aspectual meanings. They will often have more than one perfect type marker, an irrealis marker, and sometimes a progressive marker or rough equivalent. The markers of these things in MSEA languages are often derived from verbs that appear in multi verb constructions (see below, and Enfield 2003), or less often from nominal sources.

Here are some further examples of the many MSEA languages in which a postverbal marker can convey that an action or situation is complete, or has occurred at least once:

- (7.43) *dɔ? tsɿ¹ pɔŋ³ sɿ⁵l*
 read ASP copy book
 'have read a book'

Shanghainese | SN | Zee and Xu 2003: 139

- (7.44) *kaw³³ na:w³³ la:w³³ pak²⁴ kin³³*
 1P go EXP Beijing
 'I have been to Beijing.'

Hainan Cham | AN | Zheng 1997: 72, cited by Thurgood *et al.* 2014: 181

- (7.45) *Kéuihdeih ló jó gó bāt chín heui máaih làuh.*
 they take PFV that CL money go buy house
 'They used the money to buy a house.'

Cantonese | SN | Matthews and Yip 1994: 145

We can note two things about the Cantonese example (7.45). First, the aspectual marker *jó* is analysed by Matthews and Yip (1994) as a suffix rather than a stand alone word. In

fact, many of the markers discussed in this section could be analysed as affixes in so far as they are often bound up in a single phonological word, together with the verb that they mark. At the same time, such affixes are grammatical words in themselves, not just bound markers but also able to occur as stand alone elements.²

The second thing to note about the Cantonese example (7.45) is that the aspectual marking occurs on only one of the multiple verbs that occur together in the example. In various kinds of Cantonese multi verb construction (see section 7.2) aspect markers and verbal particles can only mark the first verb in the series (Matthews and Yip 1994: 146 7). This is also shown in the following directional serial verb constructions, with perfect marking only on the first of a series of verbs:

- (7.46) *Dī sailouhjái jáu jó yahp heui.*
 CL children run PFV in go
 ‘The children came running in.’
 Cantonese | SN | Matthews and Yip 1994: 147

- (7.47) *A Māh fāan jó làih.*
 Grandma return PFV come
 ‘Grandma has come back.’
 Cantonese | SN | Matthews and Yip 1994: 146

As the next example shows, the same marking on a non first verb is not possible:

- (7.48) **A Māh fāan làih jó.*
 Grandma return come PFV
 ‘Grandma has come back.’
 Cantonese | SN | Matthews and Yip 1994: 146

Marking of perfect or completive aspect may be preverbal in some languages, as shown in this example from Kam:

- (7.49) *mau³³ tɔu³³ pjən⁵⁵ eok¹¹*
 he FW rain whip
 ‘He got drenched.’
 Kam | TK | Yang and Edmonson 2008: 534

Where both preverbal and postverbal forms of marking an aspectual distinction of completion or achievement are found in a language, there will typically be a meaning difference between the two. The contrast can be illustrated with examples from Lao.

² See above comments in section 4.8 regarding the relation between phonological and grammatical words.

A postverbal marker *lèèw4*, otherwise a verb meaning ‘finish, be completed’, conveys that an event is already completed and that this state of completion is currently relevant:

- (7.50) *khòòj5 kin3 mii1 lèèw4*
 1SG.P eat noodle PRF
 ‘I ate noodles already.’ (implying, e.g., that I am full now)
 Lao | TK |

By contrast, a preverbal marker, from a verb meaning ‘acquire, come to have’, conveys that the event happened as a result of some prior event or opportunity (see Enfield 2003):

- (7.51) *khòòj5 daj4 kin3 mii1*
 1SG.P ACHV eat noodle
 ‘I did eat noodles.’ ‘I got to eat noodles.’ ‘I managed to eat noodles.’
 Lao | TK |

Many MSEA languages have a form of temporal/aspectual/modal marking that can be labelled unrealis. This is sometimes also described as (or at least translated by means of) future tense. In Thai, the preverbal particle *ca* can appear to mark future, but it is not limited to signalling a time after the time of speaking. When it occurs out of context, the default reading of *ca* is future tense, that is, that the event will happen after the time of speaking:

- (7.52) *kuu ca sùù rot maj*
 1SG.B IRR buy vehicle new
 ‘I am going to buy a new car.’
 Thai | TK

However, as the following example shows, it can just as well refer to an event that was planned or expected to occur after a point of reference which is in the past:

- (7.53) *kuu ca sùù rot maj tee man pheeñ kæñ.paj*
 1SG.B IRR buy vehicle new but 3SG dear too.much
 ‘I was going to buy a new car, but it was too expensive.’
 Thai | TK

Pwo Karen has a similar preverbal particle, as shown in the following example:

- (7.54) *jø mə lɪ*
 1SG IRR go
 ‘I will go.’
 Pwo Karen | TB | Kato 2003: 636

Danau is described as having future tense, although it is unclear whether the meaning of tense (signalling that the narrated event occurs at a time after the speech event) is entailed by the marker, or whether it is simply an appropriate translational equivalent, as given in this example:

- (7.55) *ō ãe kî tɔ sh̩v̩i*
 1SG FUT go ALL shop
 ‘I will go to the shop.’
 Danau | AA | Jenny *et al.* 2015: 95

Progressive aspect marks that an event is ongoing (though not necessarily at the time of speaking). Examples from Danau and Thai show it being marked preverbally:

- (7.56) *en t̩em swe bū*
 3SG PROG eat rice
 ‘He is still eating.’
 Danau | AA | Jenny *et al.* 2015: 97

- (7.57) *phom kamlang kin khaaw*
 1.POL PROG eat rice
 ‘I am/was eating rice.’
 Thai | TK

Progressive aspect may also be marked postverbally in many MSEA languages (or after the verb phrase, as in the Thai example 7.60), as shown in the following examples:

- (7.58) *ŋa¹¹ sui⁵³ to³²³*
 you sit FW
 ‘you keep sitting/staying there’
 Kam | TK | Yang and Edmonson 2008: 534

- (7.59) *Kéuihdeih yìhgā hāahng gán lohk làih.*
 they now walk PROG down come
 ‘They’re walking down now.’
 Cantonese | SN | Matthews and Yip 1994: 147

- (7.60) *phom kin khaaw juu*
 1.POL eat rice be.at
 ‘I am/was eating rice.’
 Thai | TK

In example (7.60), the Thai postverbal progressive marker *juu* is a verb meaning ‘to be located (somewhere), to remain’. Compare the following example from White Hmong,

in which a preverbal adverbial marker *tseem* conveys the idea that something is ‘still’ happening, that is, that the event has not yet stopped:

- (7.61) *Ntxawm tseem quaj~quaj*³³
 PN still RDUP~CRY
 ‘... Nzue was still crying and crying ...’
 White Hmong | HM | Jarkey 2015: 68

Finally, we can note that postverbal and preverbal marking of aspect can often be combined in those languages that use both. Example (7.62) shows combined preverbal and postverbal marking of progressive aspect in Thai (with the two markers underlined):

- (7.62) *phom kamlang kin khaaw juu*
 1.POL PROG eat rice be.at
 ‘I am/was eating rice.’
 Thai | TK

Similarly, example (7.63) shows preverbal and postverbal marking of completive/perfect marking together in Lao:

- (7.63) *khòòj5 daj4 kin3 mii1 lèèw4*
 1SG.P ACHV eat noodle PRF
 ‘I did eat noodles already.’ ‘I got to eat noodles already.’
 Lao | TK |

7.2.3 Modality and Related Marking

Grammatical meanings in the domain of modality, such as necessity, obligation, and possibility, are marked by similar sorts of particles and secondary verbs as observed for aspectual markings in the last section. In the following examples, modal meanings relating to possibility and necessity are marked in preverbal position:

- (7.64) *jø ba: l̥ ua*
 1SG must go QUE
 ‘Do I have to go?’
 Pwo Karen | TB | Kato 2003: 636

- (7.65) ... *kaw³³ pu³³ kian³³ sua³³*
 ... 1P NEG able write
 ‘I can’t write ...’
 Hainan Cham | AN | Zheng 1997: 93, cited in Thurgood *et al.* 2014: 185

- (7.66) *ŋ.y.kar ɻa.pən ho:j ɻi.ʃip*
 bearskin able to.wear
 ‘Bearskin is wearable.’
 Pacoh | AA | Alves 2006: 111

Ability or permission are also often marked in postverbal position, where a marker occurs either immediately after the verb or after a verb phrase. Again, this marking is often transparently related to serial verb constructions that describe relations of result between conjoined verbs (Enfield 2003). For example, when a verb meaning ‘acquire’ or ‘get’ or similar is placed after a main verb, it conveys the meaning of a modal of ability or possibility. Here are some examples:

- (7.67) *m᷑ swe n᷑ɔ b᷑ m᷑nɔ?*
 2SG eat that get STAT
 ‘You can eat that.’
 Danau | AA | Jenny *et al.* 2015: 100
- (7.68) *ma⁵⁵ nai³³ tcan⁵⁵ li³³*
 dish this eat get/fw
 ‘The dish (of food) is edible’ or ‘You are allowed to eat this dish.’
 Kam | TK | Yang and Edmonson 2008: 533
- (7.69) *... nce tsis tau*
 ascend NEG get/can
 ‘... You’re not allowed to go up.’
 White Hmong | HM | Jarkey 2015: 64

Modality marking in MSEA languages often includes a distinction between a general marker of possibility, meaning ‘can’, with the expected range of readings (epistemic, deontic; see the above few examples) and a marker that specifically means ‘know how to’, not exclusively in a mental sense, but rather more broadly in knowing what to do. Sometimes the ‘know how to’ marker occurs in preverbal position, as in the following Austroasiatic examples:

- (7.70) *teeq deeh ruuq mee*
 1SG not know do
 ‘I don’t know how to do it.’
 Kri | AA
- (7.71) *kñom qot ceh twəə*
 1SG not know.how do
 ‘I don’t know how to do it.’
 Khmer | AA

- (7.72) *tu.miəŋ ki: ləj? cɔ:m ta?*
 crossbow 1s no know make
 ‘As for crossbows, I don’t know how to make them.’
 Pacoh | AA | Alves 2006: 109

In other cases the marking is postverbal:

- (7.73) *aan may pen*
 read not know.how/be.able
 ‘I’m unable to read (it).’ (not because I’m not allowed, but because e.g.,
 I can’t read)
 Thai | TK

In this section, I have tried to illustrate the main sorts of tense aspect modality meanings that tend to be marked in MSEA languages and to give a sense of the usual ways in which they are marked. Aspectual and modal meanings are marked liberally in these languages and in a range of ways, while true marking of either tense or evidentiality are hardly, if ever, found. Morphosyntactically, these forms of marking tend to involve independent grammatical words that are transparently related to existing nouns and verbs in the languages, showing that multi verb constructions, especially, are the means, or the source, of such marking. The semantics and pragmatics of aspectual modal marking in MSEA languages are, of course, replete with subtleties. Those subtleties are beyond our scope here, and in any case they are only partially captured in most available grammatical descriptions which, by necessity, can give limited attention to the finer details of grammatical semantics.

7.3 Multi-verb Constructions

MSEA is a global hotspot for serial verb constructions, which can be broadly defined as follows:

A serial verb construction (SVC) is a sequence of verbs which act together as a single predicate, without any overt marker of coordination, subordination, or syntactic dependency of any other sort. Serial verb constructions describe what is conceptualized as a single event. They are monoclausal; their intonational properties are the same as those of a monoverbal clause, and they have just one tense, aspect, and polarity value. SVCs may also share core and other arguments. Each component of an SVC must be able to occur on its own.

(Aikhenvald 2006: 1)

In relation to MSEA languages, the notion of serial verb construction has been defined along related lines (Matisoff 1991a: 403, Bisang 1991), sometimes as broadly as this (e.g., Li and Thompson 1981: 594, Vittrant 2007: 5), sometimes in a more restricted sense (e.g., Lutz Hughes 2016). As Aikhenvald (2006: 2) states, serial verb constructions ‘are a grammatical technique covering a wide variety of meanings and functions. They do not constitute a single grammatical category.’ That said, in much of the literature, the term serial verb construction has been reserved for certain quite narrowly defined structures. In this chapter I use the term *multi verb construction* (or MVC) in order to avoid making claims as to whether the constructions being described fit a specific technical description of ‘serial verb construction’.

There is a widespread flexibility of the surface ordering of verbs and nouns in MSEA languages, particularly where multi verb constructions are concerned, often because of the widespread nominal ellipsis that these languages allow. However, a few languages show tighter constraints on what is possible. For example, in Semelai, nominal arguments must not intervene between a sequence of verbs: ‘all NP arguments must follow the sequence of verbs’ (Kruspe 2004: 362). In the following example, while the undergoer ‘offspring’ is the direct object of ‘take’, it does not appear immediately after that verb:

- (7.74) *ki goy paloh la bapa? hn knɔŋ hn*
 3A take flee A father 3POSS offspring 3POSS
 ‘The father fled taking his child.’
 Semelai | AA | Kruspe 2004: 362

In Thai, by contrast, in a clause describing the same scene, the O argument of the first verb can occur as the immediate complement of ‘take’, thus intervening between the two verbs:

- (7.75) *phɔɔ paw luuk nii*
 father take child flee
 ‘The father fled taking his child.’ (‘The father took the child and fled.’)
 Thai | TK |

A similar structural constraint is seen in White Hmong. In a type of multi verb construction in which the O argument of the first verb is the same as the O argument of the second verb, the verbs must be consecutive, such that the O argument does not intervene between the verbs, as if the verbs (underlined here) formed a compound:³

³ Jarkey (2015: 173) notes another type in White Hmong, called Disposal SVCs, in which both A and O arguments are shared by the two verbs and in which the O argument does intervene between the verbs. See examples (7.236) (7.238).

- (7.76) *nws_{A/A}* *tua* *raug* *liab_{O/O}*
 3SG shoot.at hit.the.mark monkey
 ‘He shot some monkeys.’

White Hmong | HM | Jarkey 2015: 147

- (7.77) *kuv_{A/A}* *raws* *cuag* *lawv_{O/O}* *lawm*
 1SG pursue reach 3PL PRF
 ‘I have caught up with them.’

White Hmong | HM | Jarkey 2015: 148

In Lao, by contrast, in clauses describing the same scenes as these White Hmong examples, the O argument can (and preferably does) occur between the two verbs:

- (7.78) *man2* *ñing2* *liing2* *thiùuk5*
 3SG shoot monkey hit.the.mark
 ‘He shot (and hit) some monkeys.’

Lao | TK |

- (7.79) *kuu3* *nam2* *khaw3* *than2*
 1SG pursue 3PL be.on.time
 ‘I have caught up with them.’

Lao | TK |

A distinction between symmetrical and asymmetrical serial verb constructions can be descriptively useful (Aikhenvald 2006; see Foley and Olson 1985). In a symmetrical construction, the verbs involved have equal weight in some sense, for example, where the sequence resembles a compound or denotes a chain of events. In an asymmetrical construction, one of the verbs (either the first or the second, depending on the language and the construction) comes from a closed set. The symmetrical/asymmetrical distinction is a useful heuristic, but must be interpreted with caution, as the bases for determining symmetry are not always applied consistently or reliably across studies (see Enfield 2009: 449).

Another important element of the definition of serial verb constructions given above concerns their intonational properties. On Burmese, Vittrant states that a serial verb construction has the ‘intonation of a single phrase’ (Vittrant 2007: 5). Similarly, in describing Semelai, Kruspe (2004) notes that a serial verb construction ‘has the intonation contour of a single clause construction. There is no intonational indication of a clause boundary separating the two predicates. The introduction of such would change the whole nature of the clause’ (Kruspe 2004: 361 2). We noted in Chapter 4 that intonation is a domain of grammar that is not yet well represented in descriptive and typological work. A future task for MSEA language grammarians is to provide adequate

evidence to support statements about the properties of the ‘single clause intonation contour’ in the area’s languages.

In the rest of this section we survey a number of the ways in which verbs combine in MSEA languages, ranging from canonical serial verb constructions to other multi verb structures such as strategies for complementation and depictive secondary predication.

7.3.1 Possible Number of Verbs in Series

Most multi verb constructions feature two verbs, or less often three. While there is perhaps no intrinsic limit on the number of verbs in a multi verb construction, more than about five or six would seem to be vanishingly rare (and are not flat structures but analysable as nested two or three part constructions; Enfield 2007: 481–4). In my own work on Lao, I have occasionally found examples with six or more verbs in series, as illustrated here (with verbs underlined):

- (7.80) *caw⁴* *lòòng²* *gaw³* *paj³* *hêt¹* *kin³* *beng¹* *mèè⁴*
 2SG.POL try take go make eat look IMP.UNIMP
 ‘You go ahead and take (them) and try cooking (them) to eat!’

Lao | TK | Enfield 2017: 193

- (7.81) *tok²* *long²* *tèèk⁵* *pakòòp⁵* *kan³* *ñang²* *thoo²* *daj⁴* *leej²*
 fall descend break assemble COLL remain call acquire/can pass
 ‘(When/if) it falls down and breaks apart (one) can still put it (back)
 together and make a call without ado.’ (Describing a hardy brand of
 mobile phone.)

Lao | TK | Enfield field notes 2019

I argue that these examples combine multiple constructions, each of which has just two, or sometimes three, elements (Enfield 2007: 481–4). Thus, in example (7.80), in one construction *lòòng²* ‘try’ takes a clause as its complement; in another, *hêt¹* ‘make/do’ and *kin³* ‘eat’ combine in a purposive construction; in another, *beng¹* ‘look’ occurs after a predicate to indicate that the action is done without knowing in advance how it would turn out; in another, *gaw³* ‘take’ hosts the (here, unexpressed) theme argument in a three place predication; and finally, *paj³* ‘go’ acts as a directional marker on the verb *hêt¹* ‘make/do’ as primary constituent of a clause.

While the Lao case is closer to the norm in MSEA, some languages have constraints on the number of verbs allowed in a multi verb construction. In Semelai, ‘the maximal number of verbs attested in a serial complex is two’ (Kruspe 2004: 362). While the following example features three verbs, ‘the third verb belongs to the following

concatenated clause, reflected by an intonation break and indicated by the presence of the comma':

- (7.82) *ki goŋ lumpot, paloh ?a kəh nɔ? neŋ*
 3_A bring leap flee DET 3 this before
 'He took (it) leaping, and fled (or: in order to flee), the
 aforementioned him.'

Semelai | AA | Kruspe 2004: 362

We now turn to some of the forms and functions of multi verb constructions in MSEA languages, proceeding roughly on the basis of distinctions in the kinds of events or situations that different multi verb constructions can convey.

7.3.2 Events Occurring in Series

A frequent function of a multi verb construction is to depict a series of events by simply chaining verbs or verb phrases in the order of the events that occur in the scene being described. In the following examples from across MSEA, note that the subject of each verb stays the same through the series, tracking a single actor through a scene:

- (7.83) *ger bùh oak cha ka*
 he roast clean eat fish
 'He roasted, cleaned, (and) ate the fish.'
 Haläng | AA | Cooper 1966: 33
- (7.84) *pataw ruup bùh ?maang pahaang giaw?* àih
 rich man capture pound beat roast cook on a stick you
 'The rich man will capture, pound, beat, roast, (and) cook you.'
 Haläng | AA | Cooper 1966: 33
- (7.85) *kataam yok cha ger*
 crab get eat her
 'The crab caught (and) ate her.'
 Haläng | AA | Cooper 1966: 33
- (7.86) *pøtaaw tày, dii? yun, naaw thaay*
 king stand climb.into hammock go home
 'The king stood up, climbed into his palanquin, (and) went home.'
 Phan Rang Cham | AN | Blood 1977: 61, cited by Thurgood 2005: 510

- (7.87) *əmee?* *MəKaam chò? pà təkò càn kɔ? ɳa? baŋ*
 Mom Kam scoop take up chop head make eat
 ‘Kam’s mother scooped it up, cut off its head, (and) made it into food.’
 Phan Rang Cham | AN | Thurgood 2005: 510
- (7.88) *khràmq qooç quu miì*
 pack up go LOC cassava
 ‘pack up (and) go to the cassava (fields)’
 Kri | AA | Lutz Hughes 2016: 25
- (7.89) *m̥i³⁵ a³³dzl³³ pi³³hu³¹ vu³¹ha³¹ pin³¹ vu³³ laŋ³³ a³⁵*
 3SG push door open walk go in PTC
 ‘He pushed the door open (and) walked in.’
 Judu Gelao | TK | Kang 2009: 199
- (7.90) *nin⁵³ soy⁵⁵ key⁵³ jaŋ²⁴ pje²³¹ miŋ⁵³*
 3SG push door walk in go
 ‘He pushed the door open (and) walked in.’
 Thailand Mien | HM | Liu 2012: 318
- (7.91) *je³³bwa³³ ɳan³¹ liu²⁴ ɳay²⁴ miŋ⁵³ tsware²⁴ bjau⁵³ ɳa²³¹*
 1PL.EXCL eat ASP food go sow paddy ASP
 ‘Upon finishing eating, we went to sow the paddy.’
 Thailand Mien | HM | Liu 2012: 318
- (7.92) *Bātyùh ngóhdeih heui tái hei.*
 rather we go see film
 ‘Let’s go (and) see a film.’
 Cantonese | SN | Matthews and Yip 1994: 143
- (7.93) *‘sha ne’ to’ ‘sa pa Ø*
 salt with touch eat PV:POL (PVF:IMP)
 ‘Soak (it) in salt (and) eat (it).’
 Burmese | TB | Vittrant 2007: 5

As some of the above examples show, many languages allow full verb phrases (i.e., V NP phrases) to be chained (allowing an O argument of a V1 to intervene between V1 and V2). Some languages instead may require, in some situations at least, that the verbs be adjacent (recall examples 7.74, 7.76, 7.77 above). A case is Pwo Karen, as illustrated in the following two examples:

- (7.94) *jə xwe ?an: ko:*
 1SG buy eat confectionery
 ‘I bought (and) ate cake.’
 Pwo Karen | TB | Kato 2003: 643

- (7.95) *jə l̥ xwe ja: l̥ phja.*
 1SG go buy fish at market
 ‘I went (and) bought a fish at the market.’
 Pwo Karen | TB | Kato 2003: 642

The next example shows that Pwo Karen speakers do not allow an O argument to immediately follow V1, thus intervening between the two verbs:

- (7.96) **jə l̥ phja. xwe ja:*
 1SG go market buy fish
 ‘I went to the market and bought a fish.’
 Pwo Karen | TB | Kato 2003: 641

By contrast, in Thai, for example, the (7.96) structure, disallowed in Pwo Karen, with an object intervening between verbs, would be fine, as in these renderings of examples 7.94 and 7.95:

- (7.97) *phom suuu khanom kin*
 1SG.M buy confectionery eat
 ‘I bought confectionery and ate it.’
 Thai | TK |

- (7.98) *phom paj talaat suuu plaa*
 1SG.M go market buy fish
 ‘I went to the market and bought fish.’
 Thai | TK |

Recall from section 3.5 that Karen is unusual for a Tibeto Burman language in having AVO constituent order in transitive clauses. Normally a Tibeto Burman language will be verb final, meaning that multiple verbs can usually combine directly at the end of a clause. Kato (2003: 642) writes that in Pwo Karen, ‘the only time a noun phrase can separate the two verbs is when the second verb denotes the result of the first verb or an objective description about the situation denoted by the first verb’. In the following examples, and where relevant in further examples below, verbs are underlined, and I specify the pattern of argument sharing across the verbs (where ‘A S’ means that the A argument of the first verb is the S argument of the second verb, and *mutatis mutandis* for other combinations of A, S, and O):

A S

- (7.99) *jə ŋan: mi ble jao*
 1SG eat rice full PERF
 ‘I’ve eaten rice and got full.’
 Pwo Karen | TB | Kato 2003: 645

O S

- (7.100) *jə do: thwi: θi. poon*
 1SG strike dog die SFP
 ‘When I struck the dog, it happened to die.’
 Pwo Karen | TB | Kato 2003: 645

There is a semantic difference between the verbs together and verbs apart constructions in Pwo Karen. The concatenation type (verbs together) implies that the event denoted by V2 is purposive, but not always realized. The separated type (with a noun phrase intervening between the verbs) implies that the event denoted by V2 is unexpected or accidental, but realized (Kato 2003: 645). Compare the following examples:

Concatenation type

O S

- (7.101) *jə do: θi. thwi:, (la nan. θi: θi. ɻe:)*
 1SG strike die dog but also die NEG
 ‘I struck the dog to kill it (but it did not die).’
 Pwo Karen | TB | Kato 2003: 645

Separated type

O S

- (7.102) *jə do: thwi: θi. poon*
 1SG strike dog die SFP
 ‘When I struck the dog, it happened to die.’
 Pwo Karen | TB | Kato 2003: 645

O S

- (7.103) **jə do: thwi: θi., la nan. θi: θi. ɻe:*
 1SG strike dog die but also die NEG
 (This would mean: ‘I struck the dog dead but it didn’t die.’)
 Pwo Karen | TB | Kato 2003: 645

Many verb (phrase) series can be interpreted as indicating a purposive relation between the predicates in series; that is, the subject carries out the action of V1 (often a predicate of motion) in order to do V2. In the following examples from across MSEA, we see purposive/irrealis interpretations of V2, and as far as can be discerned, we note that realis/completive interpretations are possible too. These are laid out in terms of the logical possibilities of argument sharing between the two verbs.

S S

- (7.104) *dehn sar hūm*
 3PLS descend bathe
 ‘They went down to bathe.’
 Semelai | AA | Kruspe 2004: 371

- (7.105) *ni chiw dak cha*
 you go ascend eat
 ‘You go up to eat.’
 Halāng | AA | Cooper 1966: 33

S A

- (7.106) *dɔ: la.luh₂ təm.muh he:*
 3SG run meet 1PL
 ‘He ran to meet us.’
 Pacoh | AA | Alves 2006: 109

- (7.107) *je³³ ta:i⁵³ xo²³¹ tean^{24/53} khe²⁵⁵ wa³¹*
 1SG come learn Chinese language
 ‘I came to learn Chinese.’
 Thailand Mien | HM | Liu 2012: 318

- (7.108) *ba²²taŋ⁴⁴ ləŋ⁴⁴ nəŋ³¹ le³⁵*
 PN come eat rice
 ‘Bazhang came to eat rice.’
 Aizhai Miao | HM | Yu 2010: 290

- (7.109) *kəhn dɬəh mn (t)ugpl*
 3s go.across IMPERF dibble.rice
 ‘He went across to sow rice.’
 Semelai | AA | Kruspe 2004: 371

- (7.110) *he: fjør hɔ:m da:?* *?a.ba:f bʊəj?*
 3P descend bathe water fish.for fish
 ‘We went down to bathe and fish.’
 Pacoh | Alves 2006: 109

- (7.111) *lau³²³ jan²¹² pai⁵⁵ səm³³ ɳa²¹²*
 enter house go look.for you
 ‘went into the house to look for you’
 Dong | TK | Long and Zheng 1998: 161

O S

- (7.112) *ta.lah ?n.truəj ?a.li:k ?i.ca: ?i.ca:*
 let.out domestic.animals eat
 ‘I let the domestic animals out to eat.’
 Pacoh | AA | Alves 2006: 112

IO A

- (7.113) *ki: ɿɔ:n do:j ?a.dɔ: ca:*
 1s give rice to 3SG eat
 ‘I gave rice to him to eat.’
 Pacoh | AA | Alves 2006: 108

A A

- (7.114) *we⁴⁴ tsɛ⁵³ kjɔŋ²¹ tu³¹ dɔ⁴⁴ ɿu³⁵ ga³⁵ ma³¹ ɿu³⁵ te³¹*
 1SG work.hard read book good arrive PTC good place
 ‘I study hard to move to good places.’
 Aizhai Miao | HM | Yu 2010: 292

- (7.115) *pa:i' fat³ hyat³ ma'*
 go market buy vegetable
 ‘to go to market to buy vegetables in the market’
 Mulao | TK | Wang and Zheng 1993: 78

A A & O O

- (7.116) *wǒmen zhòng nèi zhǒng cài chī*
 we raise that kind vegetable eat
 ‘We raise that kind of vegetable to eat.’
 Mandarin | SN | Li and Thompson 1981: 618

The next case, from Burmese, shows an ambiguity between A A & O O and A S:

- (7.117) *ce?* *mwe* *'pyi* *sa* *Tε*
 chicken give.birth, rise.up SUB eat PVF.R
 'He raises some chickens and eats (them).'
 'He raises chickens to eat [to earn his living].'
 Burmese | Vittrant 2007: 7

Often, the relation between verbs in a multi verb construction can be interpreted as one of cause and effect, where the second verb is interpreted as the caused result of the first verb. Here are some examples:

A S

- (7.118) *a³³kɔ²⁴ xop⁵⁵ tju⁵⁵* *xop⁵⁵* *gwin³³* *ŋa²³¹*
 EBR drink alcohol drink drunk PTC
 'The elder brother got drunk from drinking.'

Thailand Mien | HM | Liu 2012: 319

- (7.119) *i⁵⁵ q_A³¹ləuu³⁵ su³¹* *q_A³¹ləuu³⁵* *liau³⁵* *ᴀ³⁵*
 1SG see book see tired PTC
 'I read books until I got tired.'

Judu Gelao | TK | Kang 2009: 199 200

O S

- (7.120) *nws ntaus tus dev kхиav kiag*
 3SG hit CLF dog flee completely
 'He hit the dog (so that it) ran away.'
 White Hmong | HM | Jarkey 2015: 125

- (7.121) *jø do: θi. thwi:*
 1SG strike die dog
 'I struck the dog to kill it.'

Pwo Karen | TB | Kato 2003: 643

- (7.122) *ki sayor cin*
 3A stew.with.vegetable be.cooked
 'She stewed the vegetable (so it) was cooked.'

Semelai | AA | Kruspe 2004: 372

S S

- (7.123) *bɔy ma ra?wen ɻlap*
 NEG:IMP IRR abandon be.far
 ‘Don’t wander (so as to) be far (from me)!’
 Semelai | AA | Kruspe 2004: 372

- (7.124) *ɻɔwe. lan thi.pha θi.*
 3SG tumble die
 ‘He tumbled and died.’
 Pwo Karen | TB | Kato 2003: 642

S A

- (7.125) *pa⁵³ze⁴⁴ te⁴⁴qe⁴⁴ tcu⁵³ phu²² tu³⁵ za⁴⁴*
 PN angry NEG say word PTC
 ‘Bayan got angry and stopped saying anything.’
 Aizhai Miao | HM | Yu 2010: 292

As should be clear from many of the examples already shown in this section, a notable property of event chain type multi verb constructions is that they are frequently ambiguous. Li and Thompson make this point for Mandarin Chinese:

In this type of serial verb construction, the two separate events may be understood to be related in one or more of the following four ways:

- (i) Consecutive: One event occurs after the other.
- (ii) Purpose: The first event is done for the purpose of achieving the second.
- (iii) Alternating: The subject alternates between two actions.
- (iv) Circumstance: The first verb phrase describes the circumstances under which the event in the second verb phrase or clause occurs.

(Li and Thompson 1981: 595)

Here are some of their illustrative examples:

- (7.126) *wǒ mǎi piào jìn qu*
 I buy ticket enter go
 ‘I bought a ticket and went in.’ (i)
 ‘I bought a ticket to go in.’ (ii)
 Mandarin | SN | Li and Thompson 1981: 595

- (7.127) *tā tiān tian chàng gē xiě xìn*
 3SG day day sing song write letter
 ‘Everyday s/he sings songs and writes letters.’ (i, iii)
 Mandarin | SN | Li and Thompson 1981: 595

- (7.128) *hē diǎn jiǔ zhuàng zhuang dǎnzi*
 drink a:little wine strengthen strengthen gall:bladder
 ‘Drink a little wine, and it will give you courage.’ (i)
 ‘Drink a little wine to give yourself courage.’ (ii)
 ‘Get some courage by drinking a little wine.’ (iv)
 Mandarin | SN | Li and Thompson 1981: 596

Making the same point for Zhuang, Luo (2008) gives the following examples:

- (7.129) *?dɔt² ?ɔm⁵ lau³ ?deu¹ lau³ ?da:y¹ kon²*
 drink mouthful wine one warm body first
 ‘Have a little wine first and it will warm you up.’ (cause effect)
 ‘Have a bit of wine to warm yourself up first.’ (purpose)
 ‘Warm yourself up first by drinking a little wine.’ (manner)
 Zhuang | TK | Luo 2008: 371
- (7.130) *tsun²te¹ kun¹ lau³ kuak⁵ vuuan¹*
 3PL eat wine make folk.song
 ‘They drank and sang.’ (alternative)
 ‘They drank in order to sing.’ (purpose)
 ‘They sang while drinking.’ (comitative)
 Zhuang | TK | Luo 2008: 371

7.3.3 Events or Event Facets Occurring as Elements of a Single Event

In a second major category of multi verb construction, multiple verbs denote distinct facets or aspects of a single event rather than distinct events or sub events that occur in sequence. For example, one verb may denote an agent’s motion while another verb denotes an action or state that the agent is simultaneously engaged in, or can also be described as doing, as they move. In this example from Kri, the two verbs come together to denote an event in which the agent’s action can be described as both walking and coming:

- (7.131) *qung trjaangq laajq*
 3SG.P.M walk come
 ‘He came walking.’ (‘He walked here.’)
 Kri | AA

Cantonese has a system of seven general path of motion verbs that feature in this type of construction (Matthews and Yip 1994: 145). These are shown in Table 7.2.

Table 7.2 Cantonese (SN) path-of-motion verbs

Path verb	Meaning
<i>yahp</i>	in/enter
<i>chēut</i>	out/exit
<i>séuhng</i>	up/ascent
<i>lohk</i>	down/descend
<i>gwo</i>	over/cross
<i>fāan</i>	back/return
<i>màaih</i>	close

After Matthews and Yip 1994: 145.

These path of motion verbs combine with the general directional verbs of motion, *heui* ‘go’ and *lāih* ‘come’, to yield the set of possibilities laid out in Table 7.3.

Table 7.3 Cantonese (SN) directional SVCs

Path verb	SVC	
	+ <i>heui</i> ‘go’	+ <i>lāih</i> ‘come’
<i>yahp</i>	<i>yahp heui</i> ‘go in’	<i>yahp lāih</i> ‘come in’
<i>chēut</i>	<i>chēut heui</i> ‘go out’	<i>chēut lāih</i> ‘come out’
<i>séuhng</i>	<i>séuhng heui</i> ‘go up’	<i>séuhng lāih</i> ‘come up’
<i>lohk</i>	<i>lohk heui</i> ‘go down’	<i>lohk lāih</i> ‘come down’
<i>gwo</i>	<i>gwo heui</i> ‘go over’	<i>gwo lāih</i> ‘come over’
<i>fāan</i>	<i>fāan heui</i> ‘go back’	<i>fāan lāih</i> ‘come back’
<i>màaih</i>	<i>màaih heui</i> ‘go closer’	<i>màaih lāih</i> ‘come closer’

After Matthews and Yip 1994: 146.

Building on these two verb motion sequence descriptions, manner of movement can be expressed in a third verb, added before the path of motion verb. Here are some examples:

- (7.132) *jáu yahp heui*
 run in go
 ‘rush in (there)’
 Cantonese | SN | Matthews and Yip 1994: 147

- (7.133) *pàh chēut lāih*
 climb out come
 ‘climb out (here)’
 Cantonese | SN | Matthews and Yip 1994: 147

- (7.134) *tiu séuhng heui*
 jump up go
 ‘leap up (there)’
 Cantonese | SN | Matthews and Yip 1994: 147
- (7.135) *dit lohk làih*
 fall down come
 ‘fall down (here)’
 Cantonese | SN | Matthews and Yip 1994: 147
- (7.136) *būn gwo heui*
 move cross go
 ‘move over (there)’
 Cantonese | SN | Matthews and Yip 1994: 147
- (7.137) *hàahng màaih làih*
 walk close come
 ‘walk closer (to here)’
 Cantonese | SN | Matthews and Yip 1994: 147

Lao has a virtually identical system to the Cantonese one. The sets of verbs that may appear in the three slots are supplied in Table 7.4.

Table 7.4 *Elements of the Lao (TK) Manner Path Direction construction*

1	2	3			
verb of manner	verb of path	verb of direction			
<i>lèèn1</i>	‘run’	<i>khùn5</i>	‘ascend’	<i>paj3</i>	‘go’
<i>ñaang1</i>	‘walk’	<i>long2</i>	‘descend’	<i>mùa2</i>	‘return’
<i>king4</i>	‘roll’	<i>khaw5</i>	‘enter’	<i>maa2</i>	‘come’
<i>luan1</i>	‘slide’	<i>qòòk5</i>	‘exit’		
<i>tén4</i>	‘jump’	<i>khaam5</i>	‘cross over’		
<i>lòòj2</i>	‘float’	<i>lòòt4</i>	‘cross under’		
<i>khii1</i>	‘ride’	<i>taam3</i>	‘follow’		
<i>bin3</i>	‘fly’	<i>phaan1</i>	‘pass’		
<i>khaan2</i>	‘crawl’	<i>liap4</i>	‘go along edge’		
<i>taj1</i>	‘creep’	<i>qòòm4</i>	‘go around’		
<i>com1</i>	‘sink’				
<i>doot5</i>	‘leap’				
<i>etc....</i>					

Note that each verb can take its own direct nominal complement.
 After Enfield 2007: 388.

Here are two examples of the kinds of multi verb constructions that are generated by this schema:

- (7.138) *king4 khùn5 paj3*
 roll ascend go
 ‘... roll up going ...’
 Lao | TK
- (7.139) *king4 long2 maa2*
 roll descend come
 ‘... roll down coming ...’
 Lao | TK

The principles underlying the Cantonese and Lao systems emerge naturally from the structure of multi verb constructions combined with the existence of motion related vocabulary of distinct levels of semantic specificity. Accordingly, we see similar systems across MSEA languages. Here are two examples from White Hmong:

- (7.140) *cov tub.rog khiḁv tawm ntawm lub kwj.ha rov los tsev*
 CLF.PL soldier run leave nearby CLF gully valley return
 come home
 ‘The soldiers fled from the ravine and came back home.’
 White Hmong | HM | Jarkey 2015: 107
- (7.141) *nqa ib lub pob.zeb nce mus lawm saum lub roob siab~siab saud*
 carry one CLF stone ascend go depart top CLF
 mountain RDUP~be top
 ‘... take a stone up to the top of that tall mountain up there.’
 White Hmong | HM | Jarkey 2015: 107

Closely related to these manner path direction constructions are constructions in which a first verb describes the position of the person in motion. In the following two examples, the first verb phrase describes a posture of sitting (on a bus), and the second verb describes the motion event of going. This leads to a kind of instrumental reading, as suggested by the English glosses ‘take a bus’ or ‘go by bus’:

- (7.142) *te⁵³ eo³¹ səŋ⁵³ tɕəŋ³⁵ the³⁵ qd³⁵ kei³³ tʂ²² məŋ⁴⁴ za⁴⁴*
 student sit car go Aizhai go PTC
 ‘The student took a bus to Aizhai.’
 Aizhai Miao | HM | Yu 2010: 290

- (7.143) *?a:u^l nay⁵ ee^l pai^l huu^l*
uncle sit bus go market
‘Uncle took a bus to the market.’
Zhuang | TK | Luo 2008: 372

In Semelai, Kruspe describes a template for a similar kind of motion multi verb construction which involves a verb ‘take, carry’ as its initial/main element. See Table 7.5.

Table 7.5 *Semelai (AA) associated motion multi-verb constructions*

Main verb	Serial verb + motion verb	+ process verb
<i>goy</i> ‘to take/bring, carry’	a) ‘to take someone/something V-ing’ b) ‘to take oneself V-ing’	‘to engage oneself in V-ing’

Reproduced from Kruspe 2004: 368, Table 11.2.

As Kruspe’s examples show, the construction can be used for adding a theme argument to an intransitive motion predicate. The following example illustrates the elements of this construction: there is a simple intransitive predication of an event in which a man returns. This is complemented by an event of him taking some meat. This, in turn, implies that he had the meat with him while in motion. The construction is a way to add an argument (here, ‘meat’) to the clause:

- (7.144) *?pspl ga gon ?yot lawok kmpən*
if IMM bring return meat wife
‘whenever (he) was going to bring back meat (for) (his) wife’
Semelai | AA | Kruspe 2004: 368

The following two examples illustrate the same principle:

- (7.145) *ki gon ?yot mə g<n>indøy kayen ke*
3A bring return one length<NMZ> cloth that
‘He brought back a length of cloth.’
Semelai | AA | Kruspe 2004: 368

- (7.146) *bɔy ma gon ?endɔl ye ɻen*
NEG:IMP IRR bring enter.house 1 AUG
‘Don’t bring (him) inside our (house)!’
Semelai | AA | Kruspe 2004: 368

Often in multi verb constructions of this kind, one of the verbs has an adverbial function, in the sense that it modifies another verb. These modifying verbs include such meanings as ‘again’, ‘frequently’, ‘surreptitiously’, ‘improperly’, ‘fast’, ‘easy’. Here are some examples (with the adverb like verb underlined):

- (7.147) Chú ý *nghe tôi giảng lại.*
 pay.attention listen 1SG explain again
 ‘Listen attentively to my explanation again.’
 Vietnamese | AA | Srichampa 1997: 143
- (7.148) do: nɔ:y *po:r jəw ba:j*
 3SG frequent help friend
 ‘He frequently helps friends.’
 Pacoh | AA | ND&P 1986 | Alves 2006: 108
- (7.149) dehn sec *palo:h*
 3s steal flee
 ‘They surreptitiously fled.’
 Semelai | AA | Kruspe 2004: 360
- (7.150) k<r>dor ?yot *dm dəm*
 be.female<NMZ> revert IMPERF lie.down
 ‘The woman lay down again.’
 Semelai | AA | Kruspe 2004: 364
- (7.151) qooh càaw manàaq
 do.not speak disobey
 ‘Do not speak improperly.’
 Kri | AA | Lutz Hughes 2016: 25
- (7.152) ʔəwe. kli: phle:
 3SG run fast
 ‘He runs fast.’
 Pwo Karen | TB | Kato 2003: 645

In the examples just given, different orders of main verb and modifying verb can be observed. The following examples from Vietnamese show that different orderings of modifying verb and modified verb are possible within a single language, with different orderings determined by the modifying verb used. The examples show that *nhanh* ‘quick’ goes in postverbal position (to mean ‘quickly’), while *chăm chí* ‘diligent’ goes in preverbal position (to mean ‘diligently’):

- (7.153) *Anh* *áy* *hiểu* *nhanh* *lắm*.
 EBR that understand quick very
 ‘He understands very quickly.’
 Vietnamese | AA | Srichampa 1997: 137
- (7.154) *Lan* *chăm.chi* *viết* *thư* *về* *nhà*.
 Lan diligent write letter return home
 ‘Lan is diligent in writing letters back home.’
 Vietnamese | AA | Srichampa 1997: 139

Similarly, in Pacoh, two orderings of modifying verb and modified verb are possible:

- (7.155) *do:* *la.luh₁* *?a.ja:2*
 3SG run quick
 ‘He ran quickly.’
 Pacoh | AA | Alves 2006: 109
- (7.156) *?a.lo:y* *?n.neh* *?iən* *?i.kiə*
 wood this easy saw
 ‘This wood is easy to saw.’
 Pacoh | AA | Alves 2006: 111

And in Lao, again two orderings are possible, depending on which verb functions as modifier. Here are two ways of describing the fact that a roast potato takes a long time to cool down, either by using *naan2* ‘to take a long time to happen’ as a preverbal modifier, or *don3* ‘to continue or be the case for a long time’ as a postverbal modifier:

- (7.157) *man2* *naan2* *jēn3*
 3SG take.a.long.time cool
 ‘It takes a long time for (the roast potato) to cool.’
 Lao | TK
- (7.158) *man2* *hòòn4* *don3*
 3SG hot continue.for.a.long.time
 ‘They (the roast potatoes) stay hot for a long time.’
 Lao | TK

Sometimes the co occurring facets of an event being described in a multi verb construction are not related in terms of cause, effect, or purpose. They function like the modifying adjectives used in depictive constructions in English such as *They arrived drunk*, *She ate the fish raw*, and *He gave the lecture nude* in which the secondary element of the clause makes a distinct predication about a participant, independently

from what is predicated by the main verb (Himmelmann and Schultze Berndt 2005). The next example shows an associated motion construction in Hmong, which can be interpreted as depictive:

- (7.159) *nqa ib lub yeeb.thooj haus puj.pauv tuaj*
 carry one CLF pipe inhale puff.puff come
 ‘... bringing a pipe along and puffing on it’
 White Hmong | HM | example from Johnson 1985: 124, cited by Jarkey 2015: 115

Some cases of associated posture (already encountered in motion description constructions earlier in this section) can be interpreted in this way. In the following examples, a verb of motion combines with a verb that predicates an activity which occurs quite independently of the motion event:

- (7.160) *jə chi.nan ko ea*
 1SG sit cry
 ‘I sat and cried.’ (or: ‘I cried sitting down’, ‘I sat crying.’)
 Pwo Karen | TB | Kato 2003: 643

- (7.161) *jə chi.nan ɻax:kho ɻəwe*
 1SG sit wait 3SG
 ‘I waited for him while seated.’ (or: ‘I sat waiting’, ‘I waited sitting.’)
 Pwo Karen | TB | Kato 2003: 643

- (7.162) *te^{53/21}ku⁴⁴ tcəŋ³⁵ ɳi⁵³ nəŋ³¹ daŋ³¹*
 YBR sit at eat candy
 ‘The younger brother is sitting there, eating candies.’
 Aizhai Miao | HM | Yu 2010: 292

Relatedly, in some multi verb constructions each verb makes reference to a complementary facet of some broader, distributed types of activity. The verbs or verb phrases together index the larger set of actions. Here is an example from Dong, referring to indicative types of action that may comprise a villager’s livelihood:

- (7.163) *ɻat³²³ ɳay¹³ say³¹ pa⁵⁵*
 cut grass raise fish
 ‘cut grass, raise fish’
 Dong | TK | Long and Zheng 1998: 161

Often, these constructions show forms of parallelism, making them resemble the four syllable expressions that occur in these languages (see section 3.8.2). Here is an example from White Hmong:

- (7.164) cov poj niam npaj zaub npaj mov
 CLF.PL woman prepare vegetables prepare rice
 'The women prepared the food.'
 White Hmong | HM | Jarkey 2015: 117

Similar structures are used for describing more specific situations in which disparate actions happen to both occur in the event being described. Here are examples from Semelai:

- (7.165) ki r<h>pɔh tampɔŋ rɔm midur
 3A beat.with.twig<IMPEF> beat.with.big.stick with plant SP
 'He beat (her) with small twigs (and) he beat (her) with big sticks of *Goniothalamus macrophylla*.'
 Semelai | AA | Kruspe 2004: 373
- (7.166) ki sep ca wec
 3A suck eat guts
 'She sucked (and so) ate the guts (of the fish).'
 Semelai | Kruspe 2004: 372

In a similar example from Pwo Karen, the two verbs have same argument structure (fully shared arguments and roles) and interconnected meanings:

- (7.167) jø θi:ja. na:θi: ɻɔyain ɻɔcon
 1SG know understand reason
 'I know and understand the reasons.'
- Pwo Karen | Kato 2003: 643

And in an example from Burmese, co occurring event facets are closely related in causal terms (also possibly construable as events in series):

- (7.168) chiNθe' ha 'θoKa'le Ko kai? sa lai? Te
 lion TOP lamb OBJ bite eat AUX:term. PVF:R
 'The lion devoured the lamb.'
- Burmese | Vittrant 2007: 6

7.3.4 Complementation Strategies

A widespread function of multi verb constructions is to provide a complementation strategy (Dixon 2008), in which one verb takes another as its complement. Some cases involve a verb meaning 'want' or similar, as in the following examples:

- (7.169) *?a.ca;j ?ijn plɔj duŋ ?ən*
 SPMY want buy home (yes no)
 ‘Do you want to buy a home?’
 Pacoh | AA | Alves 2006: 107
- (7.170) *qunêeq mon tìrr*
 Vietnamese want get
 ‘The Vietnamese want to get (it).’
 Kri | AA | Lutz Hughes 2016: 25
- (7.171) *mei⁵³ ɔi²⁴ miŋ⁵³ xai²⁴dau³³ puŋ²⁴ yoŋ⁵³ le⁵³*
 2SG want go where pasture cattle PTC
 ‘Where do you want to go to pasture the cattle?’
 Thailand Mien | Liu 2012: 318
- (7.172) *plòh ŋu oh khin pà təkhɔ? tuy tra o*
 after she NEG dare take shoe follow again NEG
 ‘After that she didn’t dare take the other shoe with her anymore.’
 Phan Rang Cham | AN | Thurgood 2005: 511
- (7.173) *wǒ yào shàng jiē*
 I want ascend street
 ‘I want to go out.’
 Mandarin | SN | Li and Thompson 1981: 598

In some languages, such as Mandarin, a subordinate subject in a complement construction need not be coreferential with a main subject. In (7.173), the main subject is also the subject of the verb in the lower clause, and is, accordingly, elided. If the lower subject is not coreferential with the main subject, then it is simply inserted before that lower verb, and no further marking is required. This is shown in the next example:

- (7.174) *wǒ yào tā guò lai*
 I want 3SG cross come
 ‘I want him/her to come over here.’
 Mandarin | SN | Li and Thompson 1981: 598

In other languages, such as Thai, there is a constraint against this type of complement construction: the two verbs’ subjects must be co referential. The following examples from Thai show, firstly, a simple same subject ‘want’ complement construction in which the

subject of the two verbs is the same, and secondly, a case in which the lower subject is different from the main subject. In the second case, the verb ‘give’ (used in a pseudo causative sense) must be used in the lower clause as a strategy for switching subjects. The effect is that in (7.176) the same subject structural constraint is still respected, as the main subject is also the subject of the first occurring verb (*hay* ‘give’) in the lower clause.

- (7.175) *phom yaak khaam pay*
 I want cross go
 ‘I want to cross over there.’
 Thai | TK

- (7.176) *phom yaak *(hay) khaw khaam pay*
 I want give 3SG cross go
 ‘I want him/her to cross over there.’
 Thai | TK

Similarly, in a Hmong complement construction with ‘want’ as the main verb, the verb that is directly subordinate to ‘want’ must have the same subject as ‘want’. If the subject of the subordinate clause is different from that of the main verb, then the verb *kom* ‘order’ is used as a kind of switch reference strategy (though without entailing that any act of ordering occurs), as just seen for Thai:

- (7.177) *kuv xav mus*
 I want go
 ‘I want to go.’
 Hmong | HM

- (7.178) **kuv xav koj mus*
 I want you go
 ‘I want you to go.’
 Hmong | HM

- (7.179) *kuv xav kom koj mus*
 I want order you go
 ‘I want you to go.’
 Hmong | HM

While verbs such as ‘want’ project complement structures that are tightly controlled, other kinds of complement taking predicate show looser relationships between main verbs and their complement clauses. The following examples feature

verbs of perception or emotion as main verbs, taking complete clauses as their complements:

- (7.180) *ki: rɔw li: ?a.ca;j cq: to? mi:?*
 1s sad very SPMY return to America
 'I'm sad that you're going back to America.'
 Pacoh | AA | Alves 2006: 111
- (7.181) *ki: jnɔj ?a.?e:m ?ɔ: li:*
 1s see SPNY nice very
 'I see that you're very nice.'
 Pacoh | AA | Alves 2006: 112
- (7.182) *vē sī?ichē kē cuu to*
 1s fear AMB rain NEG
 'I'm not afraid that it will rain.' (matrix clause negated)
 or 'I'm afraid that it won't rain.' (embedded clause negated)
 Eastern Kayah Li | TB | Solnit 1997: 260
- (7.183) *Káu the pūi in ūi i thiaⁿ tiòh lāng lōng mñg.*
 dog PROG bark because 3SG hear ACHV person knock door
 'The dog is barking because he heard someone knocking the door.'
 Taiwanese | SN | Lin 2015: 239
- (7.184) *i muaqh buoth nin mbu mborqh fu juiv*
 i³³ mua?³ puɔ?³ nin³³ bu³³ bɔ?³ fu³³ cui⁵³
 1SG look see 3PL beat child
 'I saw them beating a child.'
 Ruyuan Mien | HM | Liu 2016: 104
- (7.185) *Bôh Malayu mai gaik*
 see Malay come again
 '(The Khmers) see the Malays coming again.'
 Western (Cambodian) Cham | AN | Baumgartner 1998: 7

When a main verb is a verb of speech or cognition, similarly the complement clause can stand apart as a representation of what was said or thought by the main subject. Note that there is seldom any formal distinction in MSEA languages between direct and indirect quotation of the discourse or thought being referred to (though the use of elements like interjections, prosody, and shifted pronouns in the quoted clause can make it clear that

the speech or thought quoted is direct rather than indirect). The following are some examples:

- (7.186) *tā tīyì wōmen dōu qù chī jiǎozǐ*
 3SG suggest we all go eat dumplings
 ‘S/he suggested that we all go eat dumplings.’
 Mandarin | SN | Li and Thompson 1981: 601
- (7.187) *te¹ cə:y⁶ci⁴ eiŋ³ vai⁶tsi² kuŋ¹ lau³*
 3SG often invite friends eat wine
 ‘He/she often invites friends for a drink.’
 Zhuang | Luo 2008: 371
- (7.189) *i cai naiv dauh dungx maih i baeqv juang*
 i³³ ts^hai³³ nai⁵³ tau³¹ tuŋ³⁵ mai³¹ i³³ pæʔ⁵ cuŋ³³
 1SG guess this CLF pig have two hundred jin
 ‘I guess that this pig weighs two hundred jin [100 kg].’
 Ruyuan Mien | HM | Liu 2016: 104
- (7.190) *A pa kóng gún ē tàng chhī káu á.*
 father say 1PL can feed dog
 ‘Dad said that we are allowed to have a dog.’
 Taiwanese | SN | Lin 2015: 288
- (7.191) *chan goh ngu na khī mit hah*
 can³ goʔ³ njuu¹ naa³ khii³ mit³ haʔ¹
 clean FUT.IMP say SEQ 3SG think DECL
 ‘He thought to clean the fish.’
 Turung | TK | Morey 2010: 597
- (7.192) *Ông Chăm non sua laik: “Djauk.gék phông hu ðuaik?”*
 title.RESP Cham that ask that why group 2.LORESP to.run
 ‘An old Cham man asks them: “Why are all of you running?”’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 8
- (7.193) *he:y bɔ:k wâ: ca tʰɔ: ka:n mōy.kʰōn ɻah.lɔ:*
 1PL tell SUB FUT do work auspicious DISC
 ‘We tell him we will have an auspicious ceremony.’
 Chong | AA | Premsrirat and Rojanakul 2015: 619

Many of the examples of looser complementation given in this section have not featured an explicit marker of complementation (equivalent of ‘that’ in English). In many MSEA languages an overt complementizer is optional. Like many languages around the world,

when verbs of perception and cognition are involved, the complement marker is often a grammaticalized version of a verb of saying or speaking. The following example from Mandarin shows a main verb *gàosu* ‘tell’ optionally taking the general verb *shuō* ‘say’ before the complement clause that describes what was said by the main subject:

- (7.194) *tā* *gàosu* *wǒ* (*shuō*) *nǐ* *tóu* *téng*
 3SG tell I say you head ache
 ‘S/he told me (that) you had a headache.’
 Mandarin | SN | Li and Thompson 1981: 602

While *shuō* ‘say’ can also be used as a main verb, it cannot co occur twice (as both main verb and complementizer) in the same construction:

- (7.195) *tā* *shuō* (**shuō*) *tā* *méi* *zuò* *gōngkè*
 3SG say say 3SG not do homework
 ‘S/he said s/he hadn’t done his/her homework.’
 Mandarin | SN | Li and Thompson 1981: 602

In White Hmong the complement markers *hais* and *tias* are derived from verbs meaning ‘say, speak’ (Jaisser 1984: 39, Jarkey 2006: 120):

- (7.196) *kuv* *to.taub* (*hais.*)*tias* *lawv* *yog* *Hmoob*
 1SG understand THAT(‘say’) 3PL COP Hmong
 ‘I understand that they are Hmong.’
 White Hmong | HM | Jarkey 2006: 120
- (7.197) *tus* *nus* *txawm* *hais* *rau* *tus* *muam* *hais.tias*,
 CL brother then say to CL sister THAT(‘say’)
 ‘*koj* *yuav* *kuv*’
 2SG arry 1SG
 ‘Then the brother said to the sister, “Marry me.”’
 White Hmong | HM | Jaisser 1984: 31, cited by Jarkey 2006: 122

In Western Cham, the complementizer is *laik* ‘say’ (see Baumgartner 1998: 4, Ex. 10):

- (7.198) *nhu* *thau* *laik* *patao* *Chăm* *non* *tăk* *phǔn* *krék*
 3 know COMP(‘say’) king C. that cut tree K.
 non *pajaloh*
 that destroy
 ‘They knew that the Cham king had cut the krek tree down.’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 8

- (7.199) *dray yǒl laik ngăk yau non jiang ô*
 1PL understand COMP(‘say’) do like that can NEG
 ‘We understand that we cannot do like that.’
 Western (Cambodian) Cham | AN | Baumgartner 1998: 8

And in Turung, ‘the phrase *ngu na*, literally “say SEQ” might be analyzed as a complementizer’ (Morey 2010: 592):

- (7.200) *ngai jolpan* sa goh ngu na mit*
 ɻai³ jolpan saa² go²³ ɻuu¹ naa³ mit³
 [1SG jolpan eat FUT.IMM] say SEQ think
 ‘I hope to eat Jolpan (“snacks”).’
 Turung | TK | Morey 2010: 591

7.4 Valency-Changing Strategies

7.4.1 Participant Adding: Syntactic Causatives and Related Constructions

A frequently used means of adding an argument to a clause is to use a syntactic causative construction. An example in English is the distinction between the simple transitive clause *He ate it* (two arguments) versus the syntactic causative *They made him eat it* (three arguments), in which ‘make’ takes a causee as its direct complement. Or between the simple intransitive clause *It fell* (one argument) and *She let it fall* (two arguments). As in the English causative construction, MSEA languages often use a multi verb construction whose main constituent is a verb meaning ‘make’ or ‘give’, which then takes a complement clause. In English, the lower verb in a syntactic causative construction is in non finite form. As we have seen, in MSEA languages there is no marking of non finiteness or similar grammatical status. This means that syntactic causatives are multi verb constructions, as defined above.

In Mandarin, several possible complement taking verbs including *shǐ*, *ràng*, and *jiào* can be used in syntactic causative constructions:

- (7.201) *zhèi jian shiqing shǐ/ràng/jiào wǒ hěn nánguò*
 this CL matter CAUSE I very sad
 ‘This matter makes me very sad.’
 Mandarin | SN | Li and Thompson 1981: 602

In Lao, similarly, different verbs are possible with a causative function. The verb *hēi* ‘make/do’ conveys a sense of direct causation, appropriate with reference to situations in which someone’s action is the cause of a purely physical event such as something breaking. First, consider a simple description of such an event:

- (7.202) *còòk5 tèèk2*
 cup break
 ‘The cup broke.’
 Lao | TK

And here is a syntactic causative construction in which the verb *hêt1* ‘make’ takes the whole clause shown in (7.202) as its complement:

- (7.203) *kuu3 hêt1 còòk5 tèèk5*
 1SG.B do/make cup break
 ‘I broke the cup.’
 Lao | TK

By contrast, the verb *haj5* ‘give’ conveys a sense of indirect causation, appropriate with reference to social situations in which someone’s action is the cause of another person’s action. The verb’s meaning is general enough to cover situations that are translatable into English as *make, let, have, and get*. Here is a simple transitive description of a person’s action:

- (7.204) *man2 kuat5 huàn2*
 3SG.B sweep house
 ‘He swept the house.’
 Lao | TK

And here is a syntactic causative construction in which the verb *haj5* ‘give’ takes the whole clause shown in (7.204) as its complement:

- (7.205) *kuu3 haj5 man2 kuat5 huàn2*
 1SG.B give 3SG.B sweep house
 ‘I made/let/had/got him (to) sweep the house.’
 Lao | TK

Here are further examples of MSEA languages in which a general syntactic causative verb (underlined in each example) is a verb that also means ‘give’:

- (7.206) *la:u⁶θai¹ hau³ lu:k⁴ha:k⁵ ɿju⁵ la:n⁴ to:k⁵ θu:¹*
 teacher give student stay home read book
 ‘The teacher allowed the students to stay at home to read.’ or ‘The teacher made the student work at home.’
 Zhuang | TK | Luo 2008: 372

- (7.207) *dééh coon paroooh*
 not give tell
 ‘(They) didn’t let (him) tell (them).’
 Kri | AA | Lutz Hughes 2016: 5

- (7.208) *man² ?na:k⁷ he² pa:i¹ ja:n¹ man²*
 3SG give 1SG go house 3SG
 ‘He let me go to his house.’
 Maonao | TK | Lu 2008: 254
- (7.209) *Nws txiv tsis pub/muab nws mus*
 3P male not give 3P go
 ‘Her father won’t let her go.’
 Hmong | HM | Clark 1989: 203
- (7.210) *vē dā cwá ne to*
 1s give go 2s NEG
 ‘I won’t let you go.’
 Eastern Kayah Li | TB | Solnit 1997: 65 (Jenny 2015: 169 notes that *dā* means ‘give’)
- (7.211) *khu:n phi: le:h de:h a:n ho:m ple: bri: fi:n*
 master ghost descend give 3SG eat fruit forest ripe
 ‘The great spirit came down and gave her ripe jungle fruit to eat.’ or
 ‘The great spirit came down and got her to eat ripe jungle fruit.’
 Palaung | AA | Milne 1921: 146, cited by Jenny 2015: 176
- (7.212) *min əmee? MiKaam oh pray naaw*
 but Mom Kam NEG give go
 ‘but Kam’s mother would not let (her) go.’
 Phan Rang Cham | AN | Thurgood 2005: 508
- (7.213) *i hō' góa tsin siong sim*
 3SG give 1SG really hurt heart
 ‘He makes me very sad.’
 Taiwanese | SN | Cheng 1974: 281

Pacoh presents an unexpected case of combining a syntactic causative with a morphological causative. Pacoh has a morphological causative derivation (like many other Austroasiatic languages; see section 5.5.4 above). It can use these morphologically causative verbs as main complement taking verbs in multi verb syntactic causative constructions. In these cases, the source verb occurs twice, first as the main verb (with the causative morphology applied) and second as the subordinate verb (in its bare form). Here are two examples (with the verb in question underlined):

- (7.214) *ki: pa.ηɔ:j? ɻa.dɔ:₂ jɔw ηɔ:j?*
 1s make:drink to 3SG friend drink
 ‘I make my friend drink.’
 Pacoh | AA | Alves 2006: 108

- (7.215) *ki: pa.ca: ?a.dɔ:₁ ca: dɔ:j*
 1s feed to 3SG eat rice
 ‘I made him eat the rice.’
 Pacoh | AA | Alves 2006: 108

Something similar happens in Mandarin, in which a lexically causative verb (here *wei* ‘feed’; i.e., ‘have/let/make someone eat’) can take *chi* ‘eat’ as a caused verb in a lexical causative construction. Here, it is not the case that the same verb appears twice, but the core semantic notion ‘eat’ occurs twice in the construction:

- (7.216) *wǒ wèi nǐ chī*
 1s feed 2s eat
 ‘I feed you.’ (‘I feed you to eat.’)
 Mandarin | SN

Negative imperative constructions can sometimes be interpreted as similar in spirit to syntactic causative constructions, as they often feature a main verb that means ‘prevent’ or ‘desist’. Here are examples from Pacoh and Lao:

- (7.217) *ɟɔ:₂ ?a.kaj to:ŋ ki:*
 prevent kid speak 1s
 ‘Keep that kid from talking.’
 Pacoh | AA | Alves 2006: 112

- (7.218) *qoo⁴ caw⁴ jaa¹ vaw⁴ cang¹ san⁴*
 INTJ 2SG.P desist speak like thus
 ‘Oh, don’t you speak like that?’
 Lao | TK | Enfield 2007: 217

Going beyond core notions of causation, multi verb constructions can use other kinds of verb to introduce new arguments into clauses. For example, a verb meaning ‘help’ can be used when describing a complex event of doing something on someone’s behalf:

- (7.219) *Ngóh bōng léih dá dihnwá.*
 I help you call phone
 ‘I’ll phone for you.’
 Cantonese | SN | Matthews and Yip 1994: 143

This construction is ambiguous in Mandarin:

- (7.220) *wo bang ni xie zuoye*
 1SG help 2SG write homework
 i. ‘I’ll do your homework for you.’
 ii. ‘I’ll help you with doing your homework.’
 Mandarin | SN

Note how the verb meaning ‘help’ corresponds to the preposition ‘for’ in the English translation. The function is similar: both elements host a participant in an event that is not selected by the basic action denoting verb of the clause. A range of other verbs can function similarly in multi verb constructions as translational equivalents of prepositions here including ‘accompany’ (to mean ‘from/with’), ‘be’ (to mean ‘for’), ‘use’ (to mean ‘with’; see Matthews and Yip (1994: 143), who use the term *coverb* in their description of Cantonese and say that it plays the role of a preposition in English). In the following Cantonese examples, the coverb is underlined:

- (7.221) *Jeui hóu tühng ngàhnhòhng je chin.*

most good accompany bank borrow money

‘It’s best to borrow money from the bank.’

Cantonese | SN | Matthews and Yip 1994: 143

- (7.222) *Kéuihdeih syún jó go Hēunggóng yàhn jouh jyújikh.*

they elect PFV CL Hongkong person be chairman

‘They elected a Hong Kong person as chairman.’

Cantonese | SN | Matthews and Yip 1994: 145

- (7.223) *Kéuihdeih yuhng dī chin heui máaih láu.*

they use CL money go buy flat

‘They’re buying a flat with the money.’

Cantonese | SN | Matthews and Yip 1994: 144

When incorporating instrument arguments into clauses, many MSEA languages use similar multi verb constructions to these, with main verbs such as ‘use’ and ‘take’ taking the added arguments as complements. Here are some examples (with ‘coversbs’ underlined):

- (7.224) *je³³ noŋ³¹ pat⁵⁵ fje⁵⁵ dza:ŋ³¹*

1SG use pen write word

‘I wrote with a pen.’

Thailand Mien | HM | Liu 2012: 319

- (7.225) *mi³⁵ lat³³ qə³¹tsə³⁵ qə³³te³³ qə³¹me³⁴ lo³¹zo³¹ su³¹*

3SG use pen hair write word, character

‘He uses brush pen to write.’

Judu Gelao | TK | Kang 2009: 200

- (7.226) *dɔ: jiəl tər.naf taf*

3SG take a hammer to hammer

‘He took a hammer to hammer.’

Pacoh | AA | Alves 2006: 112

As the following two examples show, there is a fine line between instrumental and consecutive event readings of a multi verb construction, given that both are possible construals of an instrumental situation:

- (7.227) *la¹ tshja¹ pa:i¹ jyn⁴ moi²*
 drag cart go carry coal
 ‘carry some coal with a cart’ (instrumental)
 ‘get a cart and carry some coal’ (consecutive)
 Mulao | TK | Wang and Zheng 1993: 78

- (7.228) *luŋ⁴ ?au¹ ea:k⁵ la:m⁵ va:i⁴*
 uncle take rope tie.up buffalo
 ‘Uncle tied the buffalo with a rope.’ (instrumental)
 ‘Uncle took a rope and tied the buffalo.’ (consecutive)
 Zhuang | TK | Luo 2008: 372

Relatedly, ‘take’ verbs can also introduce so called theme arguments, such as a physical object that one brings or takes somewhere:

- (7.229) *MəHla?* *pà cərɔ?* *naaw thaay*
 Hlok take fish.sp go house
 ‘Hlok took the fish home.’
 Phan Rang Cham | AN | Thurgood 2005: 510

- (7.230) *khoN yu 'θwa Pa Ø*
 stool take go PV:POL (PVF:IMP)
 ‘Take the stool away.’
 Burmese | TB | Vitrant 2007: 5

When ‘take’ or ‘bring’ is used for introducing a human participant into an event, this leads to a comitative, or even causative, reading, as the following examples show:

- (7.231) *Chúng.ta đưa con đi chơi công.viên.*
 1PL.INCL bring child go play park
 ‘We take a child and go for a walk in the park.’
 (‘We go with a child for a walk in the park’, ‘We take a child for a walk in the park.’)
 Vietnamese | AA | Srichampa 1997: 141

- (7.232) *ki goy ?yot he? dɔl*
 3A bring return AT:above house
 ‘She took (him) back up to the house.’
 Semelai | AA | Kruspe 2004: 362

- (7.233) *daai dī haakyāhn yahp heui*
 bring some guests in go
 ‘bring some guests in’
 Cantonese | SN | Matthews and Yip 1994: 147

Some constructions found in MSEA languages involve multi verb constructions that resemble the many valency increasing examples we have surveyed in this section, except that they do not add arguments at all. An example is the so called handling construction or disposal construction (an example being the *ba* construction in Chinese; Li and Thompson 1981). It is comparable to the English construction involving the verb *take* in *She took a twig and snapped it* (as derived from *She snapped a twig*). This complex ‘take’ construction conveys some sort of enhanced control over the undergoer. Its use is also no doubt motivated by information structure concerns: the ‘take’ verb allows the argument to be first introduced into the discourse as a theme, before then being expressed as an undergoer.

An example from Cantonese is a construction involving the verb *jēung* ‘take/hold/put’. Indicating placement, this verb is ‘the nearest equivalent to the *bā* object structure in Mandarin’, except that it is restricted to verbs of physical or metaphorical movement (Matthews and Yip 1994: 144). Here are two examples:

- (7.234) *Kéuih jēung dī wūjōu sāam jāuwàih pehk.*
 s/he take CL dirty clothes around throw
 ‘He throws his dirty clothes all over the place.’
 (‘He takes his dirty clothes (and) throws them around.’)
 Cantonese | SN | Matthews and Yip 1994: 144
- (7.235) *Go sīnsāang jēung ngóhdeih lēuhng go fān hōi jó.*
 CL teacher take us two CL separate PFV
 ‘The teacher separated the two of us.’
 (‘The teacher took the two of us and separated us.’)
 Cantonese | SN | Matthews and Yip 1994: 145

We finish with three further examples of the handling construction, in which the first verb in a multi verb construction does not in fact change the basic argument structure of the event being described in the subsequent main lexical verb (here, ‘dispose of’, ‘hide’, and ‘close’, respectively):

- (7.236) *Kuv yuav muab qho no pov tseg*
 I IRR take thing this throw abandon
 ‘I will dispose of this thing.’
 Hmong | HM

- (7.237) *muŋ² ɻau¹ va:n¹ ɻjau⁴ ɻdi¹*
 you take axe hide good
 ‘You hide the axe well.’
 Wangmo Bouyei | TK | Zhou 2000: 444
- (7.238) *ku¹ pa⁴ pa³tu³ yup⁷ ɻdi¹ leu⁴*
 I take door close well PRF
 ‘I have closed the door already.’
 Zhenfeng Bouyei | TK | Zhou 2000: 444

7.4.2 Single Participant in a Transitive Event: Reflexives and Related Constructions

In addition to valency increasing strategies, there are ways in which non canonical argument alignment can be marked in the morphosyntactic resources of the world’s languages. When a single referent serves both A and O function in a transitive clause, reflexive constructions are used, as in ‘She_A saw herself_O’ (e.g., in a mirror). In MSEA languages, this is often handled by a stand alone pro form, typically meaning ‘body’ or ‘self’, occurring in O function. For example, in Khmer, several related forms may serve this function, including *aeng* ‘self’, *kluan aeng* ‘body self’, *atma: kluan aeng* ‘self or body (monk) self’ (Haiman 2011: 197):

- (7.239) *cru:k kheu:nj kluan aeng knong kanjoak*
 pig see self in mirror
 ‘The pig sees itself in the mirror.’
 Khmer | AA | Haiman 2011: 197
- (7.240) *steu: jo:ng atma: kluan aeng pum ruac*
 almost pull body self self not manage
 ‘almost but not quite succeed in pulling himself along’
 Khmer | AA | Haiman 2011: 197

In Kuy, a similar ‘analytical reflexive’ *cak ɻe:ŋ* is used, literally meaning ‘body self’ (Bos 2009: 71):

- (7.241) *m3dɛ:ŋ ɻeŋ t3m cak ɻe:ŋ*
 woman that hit.whip body self
 ‘That woman is beating herself.’
 Kuy | AA | Bos 2009: 71

- (7.242) *hay səsə:r cak ðe:y*
 1 praise body self
 ‘I praise myself.’
 Kuy | AA | Bos 2009: 71

In Vietnamese, the marker *mình* serves this function. The following example shows *mình* in O function when coreferential with A:

- (7.243) *Anh đánh nó, nó phải giữ mình.*
 E.BRO hit 3SG 3SG must protect oneself
 ‘You hit him; he has to defend himself.’
 Vietnamese | AA | Thompson 1987: 251

In Thai, the reflexive is expressed with a structure similar to that seen in Khmer and Kuy, above. The reflexive expression in Thai is *tua eey*, where *tua* means ‘body’ and *eey* has the reflexive meaning or the ‘emphatic pronoun’ meaning:

- (7.244) *kháw mɔɔŋ duu tua eey nay kracòk*
 3 stare look REF in mirror
 ‘She looked at herself in the mirror.’
 Thai | TK | Iwasaki and Ingkaphirom 2005: 55

In Lao, the equivalent (and cognate) expression *to^o qêêng3* means ‘oneself’. When expressing reflexive events, *to^o qêêng3* fills a non subject role:

- (7.245) *phen1 hén3 to^o qêêng3 naj2 vèèn2*
 3.POL see REFL in mirror
 ‘She saw herself in the mirror.’
 Lao | TK | Enfield 2007: 334

- (7.246) *phen1 paat5 mùù2 to^o qêêng3*
 3.POL slice hand REFL
 ‘She cut her (own) hand.’
 Lao | TK | Enfield 2007: 334

- (7.247) *phen1 khaaj3 hùan2 to^o qêêng3*
 3.POL sell house REFL
 ‘She sold her own house.’
 Lao | TK | Enfield 2007: 334

Another way to express the reflexive relation in Lao is to use the noun phrase *caw4 khòòng3* ‘owner’ as a reflexive pronoun:

- (7.248) *phen1 paat5 müù2 caw4 khòòng3*
 3.POL slice hand owner
 ‘She cut her (own) hand.’ (lit. ‘She cut the owner’s hand.’)
 Lao | TK | Enfield 2007: 334

Certain middle type verbs do not require explicit marking for reflexivity, as in this case with *qaap5* ‘to bathe’, which takes *nam4* ‘water’ as its complement:

- (7.249) *phen1 qaap5 nam4*
 3.POL bathe water
 ‘He bathed (himself).’
 Lao | TK | Enfield 2007: 334

Also in the Tai Kadai family, Xia’ao Zhuang has two reflexive/emphatic forms: a native form *ka:k^{l3}* and a Chinese loan *sŋ^{l3}tei⁴²*. These forms can combine with personal pronouns and are able to fill the subject and the possessor roles, but rarely the object role (Wei 2012: 87–9):

- (7.250) *te⁴² ka:k^{l3} kwon³³ te⁴²*
 3SG REFL manage 3SG
 ‘He takes care of himself.’
 Xia’ao Zhuang | TK | Wei 2012: 88
- (7.251) *mə:ŋ²³¹ ka:k^{l3} kwai^{l3} mə:ŋ²³¹*
 2SG REFL blame 2SG
 ‘You blame yourself.’
 Xia’ao Zhuang | TK | Wei 2012: 88

- (7.252) *wə:n²³¹ la:i⁴² mi²³¹ tse³¹jin^{l3} kwon⁴² di⁴² sŋ^{l3}tei⁴²*
 human many have responsibility manage good REFL
ti³³ lə:k^{l3}
 PTC child
 ‘All people have the responsibility to look after their own kids.’
 Xia’ao Zhuang | TK | Wei 2012: 13

- (7.253) *ni:ŋ³¹ko³¹ me³¹ mə:ŋ²³¹ sŋ^{l3}tei⁴² je⁴² pu^{l3} nə:n³¹ kou^{l3}*
 would.rather tire 2SG REFL also NEG can enough
me³¹ pu^{l3}ən³³
 tire others
 ‘You’d rather bother yourself than others.’
 OR ‘It’s better to bother yourself than others.’
 Xia’ao Zhuang | TK | Wei 2012: 89

In Phan Rang Cham, the reflexive form is *trày/trèy* ‘self’, derived from the lexical noun meaning ‘body’ (< PC *drəy ‘body’; Thurgood 2005: 504):

- (7.254) *saniŋ yaaw nan tapay klaw tha trày*
 think like DIST rabbit laugh one self
 ‘Thinking like that, the rabbit laughed to himself.’
 Phan Rang Cham | AN | Blood 1977: 54, 62 (cited by Thurgood 2005: 504)

In Hainan Cham, the reflexive marker is *se²¹* (also from PC *drəy ‘body’; Thurgood *et al.* 2014: 147).

In another Austronesian language, Bih, the reflexive construction is built from a combination of *ačō* ‘self’ and an appropriate pronoun. Here are some examples:

- (7.255) *Ñ;u khăt ñu ačō*
 3 cut 3 REFL
 ‘She cuts herself.’
 Bih | AN | Nguyen 2013: 123
- (7.256) *Tam jhuñ kĩn ñu ačō*
 PN hate DAT 3 REFL
 ‘Tam hates herself.’
 Bih | AN | Nguyen 2013: 124
- (7.257) *Di ih prni kĩn di ih ačō*
 PL 2 praise DAT PL 2 REFL
 ‘You admire yourselves.’
 Bih | AN | Nguyen 2013: 124

As in examples from other languages shown so far in this section, the Bih reflexive marker is also used in disambiguating reference in possessive constructions. In the following, the addition of the reflexive marker in the possessive phrase *arđeh ñu* ‘her motorbike’ indicates that the possessor of the motorbike is coreferential with the main subject:

- (7.258) *Mř Mlih pamřt arđeh ñu ačō lăm sang*
 mother PN CAUS push motorbike 3 REFL LOC house
 ‘Mlih’s mother pushed her (own) motorbike to the house.’
 Bih | AN | Nguyen 2013: 126

Similar disambiguating functions are seen in embedded clauses, in which a main subject and an embedded subject compete for co reference with an embedded O. In (7.259), without the reflexive marker, the third person pronoun *ñu* in O function is coreferential

with the main clause subject. If the reflexive marker *ačō* is added after *ñu*, as in (7.260), then it is coreferential with the embedded clause subject.

- (7.259) *Mī Rai pamīn mī Nghiēp paj huñ kīn ñu*
 mother PN think mother Nghiēp CAUS hate DAT 3
 ‘Rai’s mother_i thinks that Nghiēp’s mother_j hates her_i.’
 Bih | AN | Nguyen 2013: 127

- (7.260) *Mī Rai pamīn mī Nghiēp paj huñ kīn ñu ačō*
 mother PN think mother Nghiēp CAUS hate DAT 3 REFL
 ‘Rai’s mother_i thinks that Nghiēp’s mother_j hates herself_j.’
 Bih | AN | Nguyen 2013: 127

Turning to Sinitic, the Mandarin reflexive pronoun *ziji* ‘occurs in the verb phrase in any position where a noun phrase could occur and may optionally be preceded by a pronoun that is coreferential with the subject of the sentence’ (Li and Thompson 1981: 137). The form signals ‘that its referent is coreferential with that of the subject of the clause’ (Li and Thompson 1981: 137). The reflexive marker occurs in the following examples as direct object, indirect object, coverb object, and possessor in a possessive phrase, respectively:

- (7.261) *Lǐsì zài zébèi (tā) ziji*
 Lisi DUR blame (3SG) self
 ‘Lisi is blaming himself.’
 Mandarin | SN | Li and Thompson 1981: 137

- (7.262) *tā gěi (tā) ziji xiě le yi fēng xìn*
 3SG to (3SG) self write PFV one CL letter
 ‘S/He wrote himself/herself a letter.’
 Mandarin | SN | Li and Thompson 1981: 137

- (7.263) *wǒ gēn (wǒ) ziji shēngqì*
 1SG with 1SG self angry
 ‘I’m angry with myself.’
 Mandarin | SN | Li and Thompson 1981: 137

- (7.264) *wǒ chuān (wǒ) ziji de yīfu*
 1SG wear (1SG) self GEN clothing
 ‘I wear my own clothing.’
 Mandarin | SN | Li and Thompson 1981: 137

This situation in Cantonese is similar. The Cantonese reflexive marker *jihgéi* occurs in O function in the following two examples:

- (7.265) *Ngóh hōichí líuhgáai jihgéi.*
 1SG begin understand REFL
 ‘I’m beginning to understand myself.’
 Cantonese | SN | Matthews and Yip 1994: 84
- (7.266) *Sailouhjái m̄h sīk bóuwuh jihgéi.*
 children NEG know protect REFL
 ‘Children don’t know how to protect themselves.’
 Cantonese | SN | Matthews and Yip 1994: 84

The reflexive/emphatic form *jihgéi* may also be combined with pronouns, as shown here:

- (7.267) *Ngóh yuhng ngóh jihgéi (lāih) jouh sahtyihm.*
 1SG use 1SG REFL (come) do experiment
 ‘I do experiments with myself (as a guinea pig).’
 Cantonese | SN | Matthews and Yip 1994: 86
- (7.268) *N̄ dī haih léih jihgéi ge sih.*
 this CL is 2SG REFL LP matter
 ‘This is your own business.’
 Cantonese | SN | Matthews and Yip 1994: 86

The Hui’an variety of Southern Min has two reflexive pronouns, *kai5 4 ki5* ‘self’ and *kai5* ‘self’ (Chen 2011: 92). Both are possible when preceded by a coverb, as in this example:

- (7.269) *ua3 ka5 4 kai5 4 ki5/kai5 tsu3*
 1SG for self cook
 ‘I cook for myself.’
 Southern Min | SN | Chen 2011: 92

As direct complement of a main verb, *kai5 4 ki5* is preferred:

- (7.270) *i1 kai5 4 ki5 pha?7 8 kai5 4 ki5*
 3SG self hit self
 ‘He hits himself.’
 Southern Min | SN | Chen 2011: 92 3

In Burmese, the reflexive markers *ko* ‘body, self’ or *kó go* ‘one’s self, one’s own body’ may optionally occur in place of pronominals (Jenny and Hnin Tun 2016: 300). The following examples show that a reflexive event, in which the same entity is in both A and O function, may be expressed either with the form in O function being a pronoun (as in example 7.271, in which *ha* ‘thing’ serves as a ‘non referential dummy noun’; Jenny and Hnin Tun 2016: 300) or the reflexive structure (as in example 7.272):

- (7.271) *θú ha θu θa? θe de.*
 3.DEP thing 3 kill die NFUT
 ‘She killed herself.’
 Burmese | TB | Jenny and Hnin Tun 2016: 300
- (7.272) *θú kó ko go θa? θe ၕwà de.*
 3.DEP body.DEP body OBJ kill die go NFUT
 ‘She killed herself.’
 Burmese | TB | Jenny and Hnin Tun 2016: 300

In Literary Burmese, the same construction features the alternative forms *ko.dain* or *mímí* (*ko*) for ‘self’ or ‘own’:

- (7.273) *sʰa lùn ၕj mwe gá mímí ko go mí.mí pyan*
 hungry exceed REL snake SUB self body OBJ self return
myo sà lai? θi.
 swallow eat follow NFUT
 ‘A snake that was too hungry swallowed and ate itself up.’
 Burmese | TB | Jenny and Hnin Tun 2016: 301
- (7.274) *mímí ko go hnéin teʰá ywé mə pyɔ̄j ðín ba.*
 self body OBJ lower put.down SEQ NEG speak proper IMPORT
 ‘You should not put yourself down.’
 Burmese | TB | Jenny and Hnin Tun 2016: 301

In Kucong (a Yi/Loloish language spoken in the mountains of Yunnan and neighbouring areas of Laos and Vietnam), two forms can be used to convey a reflexive reading (Chang 2009: 29–30). An emphatic pronominal form *tie³¹tie⁵⁵* marks a nominal in agent function:

- (7.275) *yɔ̄³¹ sv³³ lɔ̄³³ le³¹ ta³⁵fay³³ yɔ̄³¹ tie³¹tie³³ lɔ̄³³ na⁵⁵tehe³¹*
 3SG others PATIENT PTC generous 3SG self PATIENT mean
 ‘He is generous to others but mean to himself.’
 Kucong | TB | Chang 2009: 31

The indefinite/generic pronominal form *a³³mi³¹* ‘self’ has a reflexive possessive meaning in the following example:

- (7.276) *ta⁵⁵te³³ tɔ̄³¹ tei³¹ a³³mi³¹ yu³³ tsh³¹ kuaj³¹ na³³ puu³³*
 all.people NEG hurry self POSS goat manage good ASP/MOD
 ‘Don’t hurry, everyone, just take care of your own goats!’
 Kucong | TB | Chang 2009: 31

Kucong verbs may also be marked for ‘reflexive voice’ with an auxiliary form *kho³¹* (related to the verb *kho³¹* ‘return’; Chang 2009: 60). In these two examples, the relevant subject pronouns are also marked with the emphatic/reflexive *tie³¹tie⁵⁵*:

- (7.277) *ŋa³¹ yu³⁵tee³³ lɔ³³ mə³¹ dɔ³³ ŋa³¹ tie³¹tie⁵⁵ dɔ³³ kho³¹*

1SG others PATIENT NEG beat 1SG self beat REFL

‘I don’t beat others; I beat myself.’

Kucong | TB | Chang 2009: 60

- (7.278) *ŋa³¹ tie³¹tie⁵⁵ sv³³ ti³¹ pʂŋ³¹ vʂ³¹ kho³¹*

1SG self book one CLF buy REFL

‘I bought a book for myself.’

Kucong | TB | Chang 2009: 61

Eastern Kayah Li expresses reflexivity using a post verbal particle *duu* ‘own.accord’ together with a noun phrase *X nè* ‘X’s body’ in direct object function to mark coreference with a subject (Solnit 1997: 9).

- (7.279) *v̥e metha v̥e nè*

1SG see 1SG body

‘I see myself.’

Eastern Kayah Li | TB | Solnit 1997: 185

- (7.280) *v̥e təri v̥e nè*

1SG dress 1SG body

‘I get dressed (dress myself).’

Eastern Kayah Li | TB | Solnit 1997: 185

- (7.281) *v̥e təri v̥e phú nè*

1SG dress 1SG child body

‘I dress my (own) child.’

Eastern Kayah Li | TB | Solnit 1997: 185

- (7.282) *súplə cɔ̄ duu*

rope tie own.accord

‘a self tying [magical] rope’

Eastern Kayah Li | TB | Solnit 1997: 109

Turning to Hmong Mien languages, Xong uses a combination of a reflexive pronoun and a classifier: *duat leb* [REFL CLF] (with alternative forms *baod duak* [BUG REFL] and *baod duak leb* [BUG REFL CLF] also found)⁴. Reflexive pronouns are obligatory

⁴ Sposato (2015) identifies the morpheme *baod* in these forms with the homophonous ‘bug’ class term prefix.

for expressing reflexivity; an added personal pronoun is optional before the reflexive pronoun (Sposato 2015: 305–6):

- (7.283) *Monx ghad.maons beux (monx) daut leb lah!*
 2SG NEG.IMP hit 2SG REFL CLF PRF
 ‘Don’t hit yourself!’

Xong | HM | Sposato 2015: 306

- (7.284) *Wel kif (wel) daut leb naond guaot.*
 1SG angry (1SG) REFL CLF ASSOC pass
 ‘I’m really angry at myself.’

Xong | HM | Sposato 2015: 306

- (7.285) *Beul ghab (beul) daut leb.*
 3 bite (3) REFL CLF
 ‘It (i.e., the dog) bit itself.’

Xong | HM | Sposato 2015: 307

As in many other examples shown in this section, Xong reflexive marking may link a possessor with a subject:

- (7.286) *Monx lieif (monx) daut leb naond deb geud at jix ndiot!*
 2SG even (2SG) REFL CLF ASSOC DIM Y.SIB SAT NEG₁
 recognize
 ‘You don’t even recognize your own little sister!’

Xong | HM | Sposato 2015: 307

In Iu Mien, the reflexive marker is *ganh* ‘self’, as illustrated in the following examples, first showing A–O coreference (where the marker occurs twice), and second showing possessor–subject coreference:

- (7.287) *Ninh ganh daix ganh*
 3SG self kill self
 ‘He killed himself.’

Iu Mien | HM | Arisawa, 2016: 463

- (7.288) *Mbenc ning ganh nyei kuv lai*
 prepare 3SG self POSS be.delicious vegetable
 ‘(The king) prepared his own delicious dishes’

Iu Mien | HM | Arisawa, 2016: 487

Mong Njua does not have a dedicated reflexive marker. In Mong Njua, a repeated pronoun in undergoer or other non subject function may signal reflexivity, otherwise context will disambiguate as needed (Kunyot 1984: 27):

- (7.289) *kǔ yùa kù nyēng ndăw*
 I order I read book
 ‘I order myself to read.’
 Hmong Njua | HM | Kunyot 1984: 27
- (7.290) *kǔ yǔa qhau crua kǔ*
 I buy shoe for I
 ‘I bought shoes for myself.’
 Hmong Njua | HM | Kunyot 1984: 27
- (7.291) *kǔ nā chī crua nw̥*
 I mother angry for she
 ‘My mother is angry with herself.’⁵
 Hmong Njua | HM | Kunyot 1984: 27

7.4.3 Two Way or Distributed Action: Reciprocals and Related Constructions

When two entities serve both A and O functions in a single transitive event as in ‘They saw each other’ (meaning ‘She saw him and he saw her’) then a dedicated reciprocal construction may be used to express this. In MSEA languages, reciprocal situations tend to be described with an adverbial particle, often nominal in source. The syntactic position of the particle varies among languages, with two main types of pattern. One pattern involves a marker that occurs in the syntactic position where an O argument would normally occur (in most MSEA languages this is a postverbal slot). The other is a preverbal adverb like marker. In a number of cases, both of these alternatives are found in a single language, sometimes combinable in a single construction. The relevant markers are underlined in the examples given throughout this section.

In Lao, reciprocal events are usually expressed by constructions involving the morpheme *kan3*, glossed as COLL ‘collaborative’. The marker occurs in the usual O position, as shown here, first in a simple transitive clause and then in two multi verb constructions:

- (7.292) *khaw3 hēn3 kan3*
 3PL.B see COLL
 ‘They saw each other.’
 Lao | TK
- (7.293) *qaw3 huā3 laan4 son2 kan3*
 take head bald butt COLL
 ‘(They’d) butt each other with (their) bald heads.’
 Lao | TK | Enfield 2007: 321

⁵ Though it is not explicitly stated, Kunyot’s (1984: 27) account suggests that this example could also have the non-reflexive simple transitive meaning ‘My mother_i is angry with her_j.’

- (7.294) *khacaw4 haa3 haw3 haj5 kan3*
 3PL.P seek louse give COLL
 ‘They’re seeking lice for each other.’
 Lao | TK | Enfield 2007: 332

The marker is not dedicated to the reciprocal function. It may also be used to denote that an action is done ‘together’ rather than ‘to each other’:

- (7.295) *khaw3 salòòng3 kan3*
 3PL.B celebrated COLL
 ‘They celebrated together.’
 Lao | TK

A reciprocal event can also be described in Lao using a more elaborate and iconic structure, spelling out the two way action sequence (note that the first and second pronouns here do not entail first and second person referents; this example could mean ‘They seek lice for each other’):

- (7.296) *khòòj5 haa3 haw3 haj5 caw4, caw4 haa3 haw3 haj5 khòòj5*
 1SG.P seek louse give 2SG.P 2SG.P seek louse give 1SG.P
 ‘You seek lice for me, I seek lice for you.’
 Lao | TK | Enfield 2007: 332

Closely related Thai has a cognate marker *kan*:

- (7.297) *law kâ maa nây khuy kan nà*
 1 LP come sit chat REC PP
 ‘We sat and talked to each other, you know.’
 Thai | TK | Iwasaki and Ingkaphirom 2005: 305
- (7.298) *kháw thák thaay kan pen phíthii*
 3 greet REC COP ceremony
 ‘They greeted each other as a formality.’
 Thai | TK | Iwasaki and Ingkaphirom 2005: 305

In a more formal register, *kan* is replaced by the more elaborate expression (*sîñ*) *kan lé kan*:

- (7.299) *manút tàaŋ cam pen tâŋ phûŋ sîñ kan lé kan*
 human different necessary must depend REC
 ‘Human beings have to depend on each other.’
 Thai | TK | Iwasaki and Ingkaphirom 2005: 306

In Kuy, the reciprocal marker is *k̚nia* ‘together’ (with similar properties to the closely related Khmer form *knia*; Haiman 2011: 198 9):

- (7.300) *kuɔy na:w pɔ:?* *k^bnia*
 person 3 see RECP
 ‘Those people see each other.’
 Kuy | AA | Bos 2009: 70
- (7.301) *hay pɔ:?* *cp: p^bo:k.ma:k hay rɔ: tɔlɔy to?* *kap k^bnia*
 1 see dog friend 1 and monkey that bite RECP
 ‘I saw my friend’s dog and that monkey biting each other.’
 Kuy | AA | Bos 2009: 42
- (7.302) *tɔh.təbəy hay pɔ:?* *li:k cp: yiaw na:w pp:?*
 yesterday 1 see pig dog cat 3 walk
pɔntɔ:r *k^bnia*
 one.behind.other RECP
 ‘Yesterday I saw a pig, a dog and a cat walking in a row.’
 Kuy | AA | Bos 2009: 41

Note that example (7.302) shows a frequently encountered latitude in interpretation of ‘reciprocal’ markers across languages. In (7.302), if the three animals are in a line, then of course not all of the animals can literally be walking behind every other one, so the event is not strictly reciprocal. This is a widely observed extension of reciprocal meaning (cf. English *they chased each other down the street*).

Vietnamese shows a similar grammatical pattern to those illustrated above, with a plural subject, and the marker *nhau* ‘reciprocally, one another’ occurring in the O argument position:

- (7.303) *Chúng ta không thường gặp nhau.*
 1.INCL NEG often meet RECP
 ‘We [inclusive] don’t meet one another often.’
 Vietnamese | AA | Thompson 1987: 248

- (7.304) *Hai đĩa con đánh nhau.*
 two CLF child hit RECP
 ‘The two children are fighting [hit one another].’
 Vietnamese | AA | Thompson 1987: 271

In a similar pattern in Phan Rang Cham, *kàw?* ‘other’ marks reciprocal events (Thurgood 2005: 504):

- (7.305) *plɔh twà hətyu? pəthaay kol kàw? hya cɔ?*
 finish two wife husband hug other cry cry
 ‘Then the husband and wife hugged each other and cried.’
 Phan Rang Cham | AN | Thurgood 2005: 505

- (7.306) twà cə? təcə rɔɔŋ pà kàw?
- two grandmother grandchild keep take other
 ‘the two of them taking care of each other’
 Phan Rang Cham | AN | Thurgood 2005: 505

In Hainan Cham, the marker is *k^haw*⁴³:

- (7.307) miaw³³ tan²³ ɻa²¹saw³³ po⁵⁵ k^haw⁴³
- cat and dog fight RECP
 ‘The cat and the dog were fighting.’
- Hainan Cham | AN | Ouyang and Zheng 1983: 37, cited by Thurgood *et al.* 2014: 148
- (7.308) na²⁴ sa:n²³ suŋ²¹vo²¹ ka:j²¹ k^haw⁴³, po⁵⁵ k^haw⁴³
- child PROH curse RECP fight RECP
 ‘Children, don’t quarrel and hit each other.’
- Hainan Cham | AN | Zheng 1997: 88, cited by Thurgood *et al.* 2014: 148

Bih, another Chamic language (spoken in Vietnam), shows a different syntactic placement of the reciprocal marker. It occurs in preverbal position:

- (7.309) di ñu tām tháo
- PL 3 REC know
 ‘They know each other.’
- Bih | AN | Nguyen 2013: 128
- (7.310) nū tām čhök asei
- 3 REC scoop cooked.rice
 ‘They scoop cooked rice for each other.’
- Bih | AN | Nguyen 2013: 130
- (7.311) armōng leh anān mabui tām pakađah
- tiger and wild.buffalo REC CAUS fight
 ‘Tiger and wild buffalo fight each other.’
- Bih | AN | Nguyen 2013: 128

As with many other MSEA languages, the ‘reciprocal’ construction may also express collaborative events ‘in which multiple participants have the same role’ (Nguyen 2013: 131). In this next example, the marker *tām* means ‘together’ rather than ‘to each other’ (noting that the verbs here are intransitive):

- (7.312) alek *yōh* *di* *gōr* *tām* *pīt* *tām* *dīh* *nān*
quiet PTCL PL 3 REC sleep REC lie then
‘Deeply, they lie together and sleep together.’
Bih | AN | Nguyen 2013: 131

The preverbal placement of the reciprocal marker in Bih resembles the pattern seen in Sinitic. In Mandarin, *hùxiāng* ‘mutual(ly)’ occurs in preverbal position in reciprocal expressions (with the option of also adding *duifāng* ‘the other party’ after the verb):

- (7.313) *zhāngsān* *hé* *lǐsì* *hùxiāng* *ānwèi* (*duifāng*)
Zhangsan and Lisi mutually comfort the.other.party
‘Zhangsan and Lisi comfort each other.’
Mandarin | SN | Jiang 2000: 45

Cantonese has a similar pattern, using pre verbal adverbials *wuhsēung* or *béichí* (combined with *sēung*):

- (7.314) *Léuhng gūng pó* *yīnggōi* *wuhsēung* *chīnjauh*.
two husband wife should mutually accommodate
‘Couples should accommodate each other.’
Cantonese | SN | Matthews and Yip 1994: 87
- (7.315) *Sing Gīng wah yiu* *béichí* *sēung* *ngoi*.
Bible say need mutually each.other love
‘The Bible says we should love one another.’
Cantonese | SN | Matthews and Yip 1994: 87

This Mandarin style adverbial construction is associated with formal contexts, whereas in colloquial contexts a structure that more iconically spells out the two way directedness of action is used (Matthews and Yip 1994: 87; similar to the Lao example 7.296 above):

- (7.316) *ngóh bēi míń* *kéuih héuih bēi míń* *ngóh*.
1SG give face 3SG 3SG give face 1SG
‘He and I respect each other.’ (Literally: ‘I respect him, he respects me’.)
Cantonese | SN | Matthews and Yip 1994: 87

In Hui'an Southern Min, the Mandarin style preverbal use of an adverb meaning ‘mutual(ly)’ is also associated with a more formal register:

- (7.317) *hɔ5 4 sioŋ1* *bo2 4* *li3 2 kai3*
mutual not.have understand
‘(They) don’t understand each other.’
Southern Min | SN | Chen 2011: 93

The preferred option in Hui'an Southern Min involves a simpler preverbal marker. A prefix *sa1* is used:

- (7.318) *sa1 pha?7*
RECP hit
'fight with each other'
Southern Min | SN | Chen 2011: 93

- (7.319) *ua3 ka?7 8 i1 sa1 say2*
1SG and 3SG RECP same
'I am the same as him.'
Southern Min | SN | Chen 2011: 136

The Sinitic style preverbal pattern of reciprocal marking is also observed in languages of families that are in contact with Sinitic. In Xia'ao Zhuang, a reciprocal prefix *to³¹* can be used:

- (7.320) *həm²¹ni¹³ juu³³ tei³¹rəu²³¹ to³¹jin⁴²*
tonight be.at where RECP see
'Where are we meeting each other tonight?'
Xia'ao Zhuang | TK | Wei 2012: 47

In Sanjiang Kam, the preverbal marker is *toy⁵²*:

- (7.321) *ja.²³ pu.²¹² nəi²¹² tca:³³ toy⁵² tən⁵²*
two father mother that RECP quarrel
'Those parents quarrel with each other.'
Sanjiang Kam | TK | Wu 2015: 267

The distinction between placing a reciprocal marker in pre verbal versus post object position is neutralized superficially, at least in languages with verb final word order. This includes most Tibeto Burman languages. In Burmese, the adverbial *?ətc'hin.dzin* 'one another, each other' is used 'in the place of an object' (Jenny and Hnin Tun 2016: 301). Additionally, the verb phrase 'may optionally include the plural marker *tcá/dzá* in either case' (Jenny and Hnin Tun 2016: 301):

- (7.322) *θu dó lin məyà ?ətc'hin.dzin ?əyàn te^{hi?} tcá de*
3 ASS.PL husband wife each.other very love PL NFUT
'The couple love each other very much.'
Burmese | TB | Jenny and Hnin Tun 2016: 301

Another pattern in Burmese involves a combination of numerals and classifiers *tə* CLF *né tə* CLF 'one CLF with one CLF' placed in the same (post object/pre verbal) position.

This pattern ‘puts more emphasis on the individual subjects’ (Jenny and Hnin Tun 2016: 302):

- (7.323) *θu dó lin mɔyà tɔ yau? né tɔ yau? ðɔyàn tɛ̄bi? teá de*
 3 ASS.PL husband wife one CLF with one CLF very love PL NFUT
 ‘The couple loves each other very much.’

Burmese | TB | Jenny and Hnin Tun 2016: 302

- (7.324) *hni? θi? kù hma tɔ yau? né tɔ yau? le? s̄aun pè dzá de*
 year new cross at one CLF with one CLF present give PL NFUT
 ‘On New Year’s Day, we give presents to each other.’

Burmese | TB | Jenny and Hnin Tun 2016: 302

Some MSEA languages express reciprocal event semantics by means of a direct marker on the main verb, sometimes likened to a voice inflection. In Kucong, the marker is *ta³¹* or *tshu³¹*:

- (7.325) *phu³¹ n̄i³¹ khu³³ tch³¹ ta³¹ tsA³¹*
 dog two CLF bite RECP eat
 ‘Two dogs bite each other.’

Kucong | TB | Chang 2009: 60

- (7.326) *yui³⁵ tce³³ tu³¹ ma³¹ s̄j³⁵ tshu³¹*
 1PL word NEG say RECP
 ‘We don’t talk to each other.’

Kucong | TB | Chang 2009: 60

A direct verbal marker of reciprocal events is also found among Hmong Mien languages. In Mong Njua, reciprocal is marked in this way, with the reciprocal marker *shi* occurring before the main or first verb:

- (7.327) *áo tû ndāw shi nyá*
 two CLF that RECP love
 ‘The couple love each other.’ (i.e., ‘Those two love each other.’)

Hmong Njua | HM | Kunyot 1984: 27

- (7.328) *áo tû kw:tî chi⁶ yùa mōng shái nāng*
 two CLF brothers RECP invite go look movie
 ‘Two brothers invite each other to go to see the movie.’

Hmong Njua | HM | Kunyot 1984: 28

⁶ The source gives *chi* instead of *shi* in this example.

In Xong, reciprocal relations are mainly marked by a verbal prefix (*d*)*id* (which may also mark purposive and durative meanings; its phonological variants include *did* , *id* , and less often *jeud*):

- (7.329) *manx oub leb jont bioud, ghad.maons id beux id ndaot*
 2PL two CLF live home NEG.IMP DID hit DID curse
 ‘You two stay here at home, not fighting and not arguing.’

Xong | HM | Sposato 2015: 498

- (7.330) *wel nhaons beul jeud heut chauk geud.donb*
 1SG with 3 DID help do work
 ‘He and I will help each other out with the work.’

Xong | HM | Sposato 2015: 499

Another available strategy in Xong involves a double classifier construction, with the structure [one CLF V (one)CLF], where the subject is a semantically plural NP, and the classifier is selected as appropriate to the semantic class of the subject (Sposato 2015: 500):

- (7.331) *beul god aod meinl daot/niab/beux/ndaot (aod)meinl*
 3 PL one CLF:person kill/say/hit/curse (one)CLF:person
 ‘Those people are all killing/mockng/hitting/cursing each other.’

Xong | HM | Sposato 2015: 500

- (7.332) *aod hant ob ndaut dox aod zhons giant (aod)zhons*
 one CLF:PL NOM tree that one CLF: stick.to (one)CLF:
 vertical.plant vertical.plant

‘Those trees are all standing close to each other.’

Xong | HM | Sposato 2015: 500

A range of alternative strategies have been described for Iu Mien. In one strategy, two classifiers or other noun phrases (referring to the actor and undergoer in the scene) are linked by an appropriate verb, such as *gan* ‘be with’ or *caux* ‘accompany’, and placed in subject position, as shown between brackets in this example:

- (7.333) *oix.zuqc [laanh gan laanh] daux.gaux*
 must CLF_{human} be.with CLF_{human} pray
 ‘(You/we) must pray for one another.’

Iu Mien | HM | Arisawa 2016: 533

In another pattern, a reciprocal marker *jaax* is used postverbally with a plural subject:

- (7.334) *ninh mbuo nzaeng nzunc jaax mi'aqv*
 3 PL quarrel CLF_{occasion} RECP TELIC
 'They have already had one quarrel.'
 Iu Mien | HM | Arisawa 2016: 533 (citing Purnell *et al.* 2012: 590)

These two patterns may be combined, as shown here:

- (7.335) *[laanh caux laanh] mborqv jaax*
 CLF_{human} accompany CLF_{human} hit RECP
 '(They) hit each other.'

Iu Mien | HM | Arisawa 2016: 532

Finally, a more elaborate and event iconic expression meaning 'I V you, you V me' may be used (cf. examples 7.296 and 7.316, above):

- (7.336) *meih oix.zuqc bangc yie, yie oix.zuqc bangc meih*
 2SG must depend.on 1SG 1SG must depend.on 2SG
 'We have to depend on each other.'

Iu Mien | HM | Arisawa 2016: 534 (citing Purnell *et al.* 2012: 24)

- (7.337) *meih mbuox yie, yie mbuox meih*
 2SG tell 1SG 1SG tell 2SG
 'We told each other.'

Iu Mien | HM | Arisawa 2016: 535

POSTFACE

This concludes our survey of the historical context and structural properties of the languages of mainland Southeast Asia. For a full picture of language in this area, a companion volume to this one would cover the sociology and anthropology of the languages and their speakers, with topics including multilingualism, registers and diglossia, language variation, ideologies of language, language shift and loss, language and identity, verbal art, naming practices, and much more. For now, I close with three points intended to offer some perspective on the picture of MSEA languages and their status that I have outlined in this book. The first point concerns the languages' status in a global typology of language. The other two highlight the need for a nuanced understanding of the social context in which this dynamic area's languages have diversified and continue to develop.

First, while MSEA provides important data for understanding linguistic diversity worldwide, the area shows a relatively low degree of internal structural diversity. So, it appears at first blush not to have the same importance to linguistic typology as more internally diverse areas such as the Amazon or Papua New Guinea. One reason for the relatively low internal diversity of the languages of MSEA is the intensive long term contact between historical communities that has led to convergence in many levels of linguistic organization. Another is that each of the several hundred languages spoken in MSEA belongs to one of just five well established language families (and indeed around three quarters of the languages belong to just two families: Austroasiatic and Tibeto Burman). The languages therefore share many properties due to inheritance from common ancestors. This reduces their value as independent data points for questions in language science most broadly. But as the data presented in this book show, the structural convergence across MSEA languages is by no means total. As data and analysis continue to cumulate, we can expect to discover fine grained yet significant linguistic diversity in the area.

Second, while recent historical processes of colonialism, nationalism, globalization, and economic transformation have disrupted, reshaped, and upended the lives and lifeways of MSEA people, contemporary MSEA languages cannot be properly

understood without a much deeper historical lens than the last few hundred years. Many of us naturally think first of the mainland Southeast Asia area in terms of decidedly modern categories. We think of nation states: Myanmar, Laos, China, Vietnam, Cambodia, and Thailand. We think of the colonialist and nationalist histories that formed those states over the last two centuries. And we think of national languages Burmese, Lao, Chinese, Vietnamese, Khmer, and Thai – with their written forms and their ubiquitous presence in broadcasting, education, and law. But these at hand ideas belie a much more nuanced story of human diversity in MSEA, a product of processes of human movement and contact reaching back thousands of years, to well before the modern era. The extremely recent forces of globalization and industrialized state power have a long way to go before they can undo or erase the many layered logics of human life and language that exist in this area. Their linguistic significance is too easily overstated. The national languages, while prominent, large, and powerful, are not especially representative of the area in linguistic terms. Together, they represent less than two percent of the languages, yet they have attracted the vast majority of scholarly research.

Third, the social fabric of MSEA – the fabric into which its languages are stitched – is considerably more textured than our simple and often used binary distinctions might suggest: modern versus premodern; dominant versus subordinate; urban versus rural; lowland versus upland; civilized versus barbarian; settler versus indigenous. These are terms that I have long thought with, and that I still cannot always avoid. But I have come to understand how inaccurate and misleading they can be. The truth is that any claim about binary, asymmetrical relations between ethnolinguistic groups of MSEA is ultimately on shifting sand. All extant MSEA groups can be understood as diasporas. When we think of settler colonialism, we tend to think of European invasions (the Americas, South Africa, or Australia) or state sponsored internal migration programmes (Christians to Moro areas in the Philippines, Javanese to other Indonesian islands, or Mainland Chinese to Aboriginal Taiwan). These examples are uniquely large in scale and centralized in planning, but they share something in essence with the smaller, self starting waves of migration and inter group contact – reviewed in Chapter 1 – that have taken place across the MSEA area over the last twenty millennia, intensifying over the last 3,000 years, and continuing today. These have often involved displacement or assimilation of *in situ* populations, under pressure from incoming groups in search of land and resources. Viewed in this context, every MSEA born person living today, from a leader of industry in an urban context to an undocumented villager of the most isolated upland setting, is a descendant of migrants and settlers of some era or another. This is obviously true in urban centres, but it is also true in hilltop enclaves like the Nam Noi valley in central Laos. There, within just the last fifty years or so, speakers of Lao (Southwestern Tai, Tai Kadai) have taken control of land dominated by speakers of

varieties of Bru (Katuic, Austroasiatic). But a hundred years before, those Katuic speakers were new arrivals when the area was occupied by speakers of Saek (Northern Tai, Tai Kadai) and various Vietic (Austroasiatic) varieties. And a hundred years before that, Saek speakers were the new arrivals, moving into land inhabited by Vietic speakers. And in turn, those early Vietic speakers were descendants of people who had once moved into the area afresh, displacing or subsuming the Hoabinhian populations who had lived there for tens of thousands of years.

REFERENCES

- Abramson, A. S. and Svastikula, K. 1983. Intersections of tone and intonation in Thai. Haskins Laboratories Status Report on Speech Research, SR-74/75, 143–67.
www.haskins.yale.edu/sr/sr074/SR074_10.pdf
- Abramson, A. S. and Thongkum, T. L. 2009. A fuzzy boundary between tone languages and voice-register languages. In Fant, G., Fujisaki, H. and Shen, J. (eds.) *Frontiers in Phonetics and Speech Science*. 149–55. Beijing: The Commercial Press.
- Adams, K. L. 1989. *Systems of Numeral Classification in the Mon-Khmer, Nicobarese, and Aslian Subfamilies of Austroasiatic*. Canberra: Pacific Linguistics.
- Adams, K. L. 1991. The influence of non-Austroasiatic languages on numeral classification in Austroasiatic. *Journal of the American Oriental Society* 111 (1): 62–81.
- Adelaar, A. 2005a. Malayo-Sumbawan. *Oceanic Linguistics*, 44 (2): 357–88.
- Adelaar, A. 2005b. The Austronesian languages of Asia and Madagascar: A historical perspective. In Adelaar, A. and Himmelmann, N. P. (eds.) *The Austronesian Languages of Asia and Madagascar*. 1–42. Abingdon, Oxon and New York: Routledge.
- Aikhenvald, A. Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Aikhenvald, A. Y. 2006. Serial verbs constructions in a typological perspective. In Aikhenvald, A. Y. and Dixon, R. M. W. (eds.) *Serial Verb Constructions: A cross-linguistic typology*. 1–68. Oxford: Oxford University Press.
- Aikhenvald, A. Y. and Dixon, R. M. W. 2013. *Possession and Ownership: A cross-linguistic typology*. Oxford: Oxford University Press.
- Aikhenvald, A. Y. and Dixon, R. M. W. 2017. *The Cambridge Handbook of Linguistic Typology*. Cambridge: Cambridge University Press.
- Aitken, M. J. and Stokes, S. 1997. Climastratigraphy. In Taylor, R. E. and Aitken, M. J. (eds.), *Chronometric Dating in Archaeology*. 1–30. New York and London: Plenum Press.
- Alves, M. J. 2006. *A Grammar of Pacoh: A Mon-Khmer language of the central highlands of Vietnam*. Shorter Grammars. Canberra: Pacific Linguistics.
- Alves, M. J. 2009. Sino-Vietnamese grammatical vocabulary and sociolinguistic conditions for borrowing. *Journal of the Southeast Asian Linguistics Society* 1: 1–9.
- Alves, M. J. 2015. Morphological functions among Mon-Khmer languages: Beyond the basic. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 532–57. Berlin: De Gruyter Mouton.
- Alves, M. J. 2016. Identifying early Sino-Vietnamese vocabulary via linguistic, historical, archaeological, and ethnological data. *Bulletin of Chinese Linguistics* 9: 264–95.
- Ammerman, A. J. and Cavalli-Sforza, L. L. 1971. Measuring the rate of spread of early farming in Europe. *Man* 6 (4): 674–88.

- Andvik, E. 2003. Tshangla. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 439–55. Abingdon, Oxon and New York: Routledge.
- Ansaldi, U. 1999. Comparative constructions in Sinitic: Areal typology and patterns of grammaticalisation. Dissertation. Stockholm University.
- Arisawa, T. D. 2016. An Iu Mien Grammar: A tool for language documentation and revitalisation. La Trobe University. <http://hdl.handle.net/1959.9/561960>
- Audric, J. 1972. Angkor and the Khmer Empire. London: R. Hale.
- Aung-Thwin, M. A. 2005. The Mists of Ramanna: The legend that was lower Burma. Honolulu: University of Hawai'i Press.
- Banham, D. 2015. The Old English monastic sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 901–10. Berlin: De Gruyter Mouton.
- Banker, E. M. 1964. Bahnar affixation. *Mon-Khmer Studies* 1. 99–117.
- Barram, A. and Glover, I. 2008. Re-thinking Dvaravati. In Pautreau, J.-P. et al. (eds.) *From Homo erectus to the Living Traditions: Choice of papers from the 11th International Conference of the European Association of Southeast Asian Archaeologists*. 175–82. Chiang Mai: Silkworm.
- Barrie, M. 2006. Tone circles and contrast preservation. *Linguistic Inquiry* 37(1): 131–41.
- Bauer, C. 1982. Morphology and syntax of spoken Mon. PhD thesis. School of Oriental and African Studies, University of London.
- Baumgartner, N. I. 1998. A grammar sketch of Western (Cambodian) Cham. *Pacific Linguistics. Series A. Occasional Papers*, 89: 1–20.
- Baxter, W. H. and Sagart, L. 1998. Word formation in Old Chinese. In Packard, J. L. (ed.) *New Approaches to Chinese Word Formation: Morphology, phonology, and the lexicon in Modern and Ancient Chinese*. New York and Berlin: Mouton de Gruyter.
- Baxter, W. H. and Sagart, L. 2014. Old Chinese: A new reconstruction. New York: Oxford University Press.
- Becker, A. L. 1975. A linguistic image of nature: The Burmese numeral classifier system. *Linguistics*, 13 (165): 109–22.
- Bedell, G. 2001. The syntax of deixis in Lai. *Linguistics of the Tibeto-Burman Area*. 24 (2): 157–71.
- Bellwood, P. 1979. Man's Conquest of the Pacific: The prehistory of Southeast Asia and Oceania. New York: Oxford University Press.
- Bellwood, P. 1985. Prehistory of the Indo-Malaysian Archipelago. Sydney: Academic Press.
- Bellwood, P. 1992. Southeast Asia before prehistory. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia*. Vol. 1. Cambridge: Cambridge University Press.
- Bellwood, P. 2007. Prehistory of the Indo-Malaysian Archipelago. Revised edition. Canberra: Australian National University.
- Bellwood, P., Fox, J. J. and Tryon, D. 2006. *The Austronesians: Historical and comparative perspectives*. Canberra: Australian National University.
- Benedict, P. K. 1942. Thai, Kadai, and Indonesian: A new alignment in Southeastern Asia. *American Anthropologist*, 44 (4): 576–601.
- Benedict, P. K. 1972. Sino-Tibetan: A conspectus. In *Princeton-Cambridge Studies in Chinese Linguistics*: Vol. 2. Cambridge: Cambridge University Press.
- Benedict, P. K. 1975. Austro-Thai language and culture, with a glossary of roots. New Haven: HRAF Press.
- Benedict, P. K. 1976a. Austro-Thai and Austroasiatic. In Jenner, P., Thompson, L. C. and Starosta, S. (eds.) *Austroasiatic Studies Part I*. 1–36. Oceanic Linguistics Special Publication No. 13. Honolulu: University of Hawai'i Press.
- Benedict, P. K. 1976b. Sino-Tibetan: Another look. *Journal of the American Oriental Society*, 96 (2): 167.

- Benedict, P. K. 1987. Archaic Chinese initials. Manuscript cited in Ratliff 2010.
- Benedict, P. K. 1990. Japanese/Austro-Tai. *Linguistica Extranea, Studia 20*. Ann Arbor, MI: Karoma Publishers.
- Benjamin, G. 2012. The Aslian languages of Malaysia and Thailand: An assessment. In Austin, P. K. and McGill, S. (eds.) *Language Documentation and Description*. 136 230. London: School of Oriental and African Studies.
- Bentley, R. A. 2006. Strontium isotopes from the Earth to the archaeological skeleton: A review. *Journal of Archaeological Method and Theory* 13. 135–87.
- Bergsland, K. and Vogt, H. 1962. On the validity of glottochronology. *Current Anthropology* 3: 115–53.
- Bertrais-Charrier, Y. 1964. *Dictionnaire Hmong (Mèo Blanc)-Français*. Vientiane: Mission Catholique.
- Bisang, W. 1991. Verb serialization, grammaticalization and attractor positions in Chinese, Hmong, Vietnamese, Thai and Khmer. In Seiler, H. and Premer, W. (eds.) *Partizipation: Das sprachliche Erfassen von Sachverhalten*. 509–62. Tübingen: Gunter Narr Verlag.
- Bisang, W. 1993. Classifiers, quantifiers and class nouns in Hmong. *Studies in language* 17 (1): 1–51.
- Blagden, C. O., 1919. The Myazedi Inscriptions in Ko, T. S. and Duroiselle, C. (eds.) *Epigraphica Birmanica*. 59–68. Rangoon: Government Printing. Cited in Moore 2009.
- Blench, R. 2011. The role of agriculture in the evolution of Mainland Southeast Asian language phyla. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 125–52. Canberra: Pacific Linguistics.
- Blench, R. 2015. The origins of nominal classification markers in MSEA languages: Convergence, contact and some African parallels. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 558–85. Berlin: De Gruyter Mouton.
- Blench, R. and Post, M. W. 2013. Re-thinking Sino-Tibetan phylogeny from the perspective of North East Indian languages. In Hill, N. and Owen-Smith, T. (eds.) *Trans-Himalayan Linguistics*. 71–104. Berlin: Mouton de Gruyter.
- Blench, R. and Post, M. W. 2014. Rethinking Sino-Tibetan phylogeny from the perspective of Northeast Indian languages. *Trans-Himalayan Linguistics* 266: 71–104.
- Blood, D. 1967. Phonological Units in Cham. *Anthropological Linguistics* 9 (8): 15–32.
- Blood, D. 1977 Clause and sentence final particles in Cham. In Thomas, D., Lee, E. W. and Liem, N. D. (eds.) *Papers in South East Asian Linguistics No. 4: Chamic Studies*. 39–51. Pacific Linguistics Series A, No. 48.
- Blood, D. 1980. The script as a cohesive factor in Cham society. In Gergerson, M. and Thomas, D. (eds.) *Notes from Indochina on Ethnic Minority Cultures*. Dallas: Summer Institute of Linguistics.
- Blust, R. 1980. Austronesian etymologies. *Oceanic Linguistics* 19 (1/2): 1–189.
- Blust, R. 1994. The Austronesian settlement of Mainland Southeast Asia. In Adams, K. L. and Hudak, T. J. (eds.) *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*. 25–83. Tempe, AZ: Arizona State University.
- Blust, R. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Zeitoun, E. and Li, P. J. (eds.) *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. 31–94. Taipei: Academia Sinica.
- Blust, R. 2013a. Southeast Asian islands and Oceania: Austronesian linguistic history. In Bellwood, P. (ed.) *The Encyclopedia of Global Human Migration. Vol. 1: Prehistory*. Oxford: Wiley-Blackwell.
- Blust, R. 2013b. *The Austronesian Languages*. Revised edition. Canberra: Asia-Pacific Linguistics.
- Bo, W. 2002. A Study of Mulao. Beijing: The Nationalities Press.
- Boas, F. 1911. Introduction. In Boas, F. (ed.) *Handbook of American Indian Languages*. 1–79. Lincoln, NE: University of Nebraska Press.

- Bodomo, A. 2006. The structure of ideophones in African and Asian languages: The case of Dagaare and Cantonese. In Mugane, J., Hutchison, J.P. and Worman, D.A. (eds.) *Selected Proceedings of the 35th Annual Conference on African Linguistics: African Languages and Linguistics in Broad Perspectives*. pp. 203–13. Cambridge, MA: Harvard University.
- Bon, N. 2014. Une grammaire de la langue stieng, langue en danger du Cambodge et du Vietnam. Dissertation. Université Lumière Lyon 2.
- Bos, K. J. 2009. *Kuy grammar sketch: A basic grammar sketch of the Kuy Ntua language in Cambodia*. SIL Mainland Southeast Asia Group. www.sil.org/resources/archives/63484
- Bradley, D. 1979. *Lahu Dialects*. Canberra: Faculty of Asian Studies.
- Bradley, D. 1997. Tibeto-Burman languages and classification. In D. Bradley (ed.) *Papers in Southeast Asian Linguistics 14: Tibeto-Burman Languages of the Himalayas*. 1–72. Canberra: Australian National University.
- Bradley, D. 2002. The subgrouping of Tibeto-Burman languages. In Beckwith, C. I. (ed.) *Medieval Tibeto-Burman Languages I*. 73–112. Leiden: Brill.
- Bradley, D. 2003. Lisu. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 222–35. Abingdon, Oxon and New York: Routledge.
- Bradley, D. 2007. Language endangerment in China and Mainland Southeast Asia. In Brenzinger, M. (ed.) *Language Diversity Endangered*. Trends in Linguistics: Studies and Monographs. 278–302. Berlin: Mouton de Gruyter.
- Brown, J. M. 1965. From Ancient Thai to modern dialects. In *From Ancient Thai to modern dialects, and other writings on historical Thai linguistics*. 69–254. Bangkok: White Lotus.
- Brown, P. and Levinson, S. C. 1978. Universals in language usage: Politeness phenomena. Question and politeness. In Goody, E. N. (ed.) Cambridge: Cambridge University Press.
- Brunelle, M. 2009a. Contact-induced change? Register in three Cham dialects. *Journal of the Southeast Asian Linguistics Society* 2: 1–22.
- Brunelle, M. 2009b. Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of Southeast Asian Linguistics*, 1, 49–62.
- Brunelle, M. 2015. Vietnamese (Tiếng Việt). In Jenny, M. and Sidwell, P. (eds.) *The Handbook of Austroasiatic Languages*. 909–53. Leiden: Brill.
- Brunelle, M. 2017. Stress and phrasal prominence in tone languages: The case of Southern Vietnamese. *Journal of the International Phonetic Association*, 47 (3): 283–320. DOI:10.1017/S0025100316000402
- Brunelle, M., Dương, N. D. and Hùng, N. K. 2010. A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica* 67: 147–69.
- Brunelle, M., HẠ, K. P. and Grice, M. 2012. Intonation in Northern Vietnamese. *The Linguistic Review* 29: 3–36.
- Brunelle, M., HẠ, K. P. and Grice, M. 2016. Inconspicuous coarticulation: A complex path to sound change in the tone system of Hanoi Vietnamese. *Journal of Phonetics*. 59: 23–39.
- Brunelle, M. and HẮN, P. V. 2019. *Colloquial Eastern Cham*. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 523–58. Berlin: De Gruyter Mouton.
- Brunelle, M. and Kirby, J. 2015. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 82–110. Berlin: De Gruyter Mouton.
- Brunelle, M., and Kirby, J. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass*. 10 (4): 191–207.
- Brunelle, M. and Kirby, J., Michaud, A. and Watkins, J. 2018. Prosodic systems: Mainland Southeast Asia. In Gussenhoven, C. and Chen, A. (eds.) *The Oxford Handbook of Language Prosody*. Oxford: Oxford University Press.
- Buakaw, S. 2012. A phonological study of Palaung dialects spoken in Thailand and Myanmar, with focuses on vowels and final nasals. Doctoral dissertation. Mahidol University, Bangkok.

- Bulbeck, D. 2011. Biological and cultural evolution in the population and culture history of Homo sapiens in Malaya. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 207–55. Canberra: Pacific Linguistics.
- Bunopas, S., and Vella, P. 1983. Tectonic and geologic evolution of Thailand. *Proceedings of a Workshop on Stratigraphic Correlation of Thailand and Malaysia*. 1: 307–22.
- Burenhult, N. 2002. A grammar of Jahai. PhD dissertation. Lund University.
- Burenhult, N. 2003. Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton. *Pragmatics*. 13: 363–79.
- Burenhult, N. 2005. *A Grammar of Jahai*. Canberra: Pacific Linguistics Research School of Pacific and Asian Studies.
- Burenhult, N. 2006. Body part terms in Jahai. *Language Sciences*. 28: 162–80.
- Burenhult, N. and Majid, A. 2011. Olfaction in Aslian ideology and language. *The Senses and Society*. 6 (1): 19–29.
- Burenhult, N. and Wegener, C. 2009. Preliminary notes on the phonology, orthography and vocabulary of Semnam (Austroasiatic, Malay Peninsula). *Journal of the Southeast Asian Linguistics Society*. 1: 283–312.
- Burenhult, N., Kruspe, N. and Dunn, M. 2011. Language history and culture groups among Austroasiatic-speaking foragers of the Malay Peninsula. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 257–75. Canberra: Pacific Linguistics.
- Burling, R. 1967. *Proto Lolo-Burmane*. Bloomington, IN: Indiana University.
- Burling, R. 1983. The Sal languages. *Linguistics of the Tibeto-Burman Area*. 7(2): 1–32.
- Burling, R. 2004. *The Language of the Modhipur Mandi (Garo)*. Vol. I: Grammar. New Delhi: Bibliophile South Asia.
- Burnay, J. and Coëdès, G. 1927. Note sur les tons et les initiales du vieux-siamois à l'époque de Sukhodaya. *Journal of the Siam Society*. 21(2): 103–18.
- Butler, B. 2015. Approaching a phonological understanding of the sesquisyllable with phonetic evidence from Khmer and Bunong. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 443–99. Berlin: De Gruyter Mouton.
- Campbell, L. 2010. Language Isolates and Their history, or, What's Weird Anyway? In *Proceedings of Berkeley Linguistics Society* 36. 16–31. <https://doi.org/10.3765/bls.v36i1.3900>
- Cao, J. 2002. Hán yǔ Shēngdiào yǔ Yúdiào de Guānxī. [The relation between Tone and Intonation in Chinese]. *Studies of The Chinese Language* (3): 195–202.
- Capell, A. 1979. Further typological studies in Southeast Asian languages. In Liem, N. D. (ed.) *South-East Asian Linguistic Studies*. Vol. 3. 1–42. Canberra: Pacific Linguistics.
- Cavalli-Sforza, L., Menozzi, P. and Piazza, A. 1993. Demic expansions and human evolution. *Science*. 259 (5095): 639–46.
- Chamberlain, J. R. 1997. Tai-Kadai Arthropods: A preliminary biolinguistic investigation. In Edmondson, J. A. and Solnit, D. B. (eds.) *Comparative Kadai: The Tai branch*. 291–326. Dallas: SIL International.
- Chandler, D. 1992. *A History of Cambodia*. Boulder, CO: Westview Press.
- Chang, J. 2009. A reference grammar of Yuanjiang Kucong. PhD. Minzu University of China, Beijing.
- Chang, K. 1947. Miaoyaoyu shengdiaowen ti [On the tone system of the Miao-Yao languages]. *Bulletin of the Institute of History and Philology* 16: 93–110.
- Chang, K. 1953. On the tone system of the Miao-Yao languages. *Language* 29: 374–8.
- Chang, K. 1966. A comparative study of the Yao tone system. *Language* 42: 303–10.
- Chang, K. 1972. The reconstruction of Proto-Miao-Yao tones. *Bulletin of the Institute of History and Philology* 44: 541–628.
- Chang, K. 1976. Proto-Miao initials. *Bulletin of the Institute of History and Philology*. 47: 155–218.

- Chao, Y. R. 1930. A system of tone letters. *Le Maître Phonétique* 45:24 7.
- Chao, Y. R. 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology, Academia Sinica* 4: 121 34.
- Chao, Y. R. 1941. Distinctions within Ancient Chinese. *Harvard Journal of Asiatic Studies* 5: 303 22.
- Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chapman, C. 2011. Subordination and coordination in Karen Sgaw. In Brunelle, M. (ed.) *Sgaw Karen Papers: Presented to Nimrod Andrew*. Papers written for the course LIN 4370/7911 Field Methods, Fall 2010. 85 96. Ottawa: University of Ottawa.
- Chappell, H. 2001. Language contact and areal diffusion in Sinitic languages. In Aikhenvald, A. Y. and Dixon, R. M. W. (eds.) *Areal Diffusion and Genetic Inheritance: Problems in comparative linguistics*. 328 57. Oxford: Oxford University Press.
- Chappell, H. 2019. Southern Min. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 176 233. Berlin: De Gruyter Mouton.
- Chappell, J. and Shackleton, N. J. 1986. Oxygen isotopes and sea level. *Nature* 324(6093): 137 40.
- Chen, G. 2005. *Kemieyu yanjiu* [A study of Kemie]. Beijing: Minzu chubanshe.
- Chen, M. 1987. The syntax of Xiamen Tone Sandhi. *Phonology Yearbook*, 4: 109 49.
- Chen, P. 2007. China. In Simpson, A. (ed.) *Language and National Identity in Asia*. 141 67. Oxford: Oxford University Press.
- Chen, W. 2011. The Southern Min dialect of Hui'an: Morphosyntax and grammaticalization. PhD. University of Hong Kong.
- Cheng, R. 1974. Causative constructions in Taiwanese. *Journal of Chinese Linguistics*. 2 (3): 279 324.
- Cheung, Y.-S. (1969). 'Xianggang Yueyu yinping-diao ji biandiao wenti' [A study on the upper even tone and tone sandhi in the Cantonese dialect as spoken in Hongkong]. *Journal of the Institute of Chinese Studies of the Chinese University of Hongkong*. 2 (2): 81 105.
- Chinese Academy of Social Sciences (CASS). 2012. *A Dictionary of Modern Chinese*. Sixth edition. Beijing: The Commercial Press.
- Church, P. (ed.) 2003. *A Short History of South-East Asia*. Third edition. Singapore: Wiley.
- Clark, A. L., King, C. L., Buckley, H. R., Collins, C. J., Dhavale, N., Elliott, G. E., Gosling, A., Halcrow, S. E., Ivory, B., Matisoo-Smith, E., McDonald, S. K., Stantis, C., Tromp, M., Ward, S. and West, K. 2017. Biological anthropology in the indo-pacific region: New approaches to age-old questions. *Journal of Indo-Pacific Archaeology*. 41: 78 94.
- Clark, M. 1985. Asking questions in Hmong and other southeast Asian languages. *Linguistics of the Tibeto-Burman Area*, 8 (2): 60 7.
- Clark, M. 1989. Hmong and areal South-East Asia. In Bradley, D. (ed.) *Papers in Southeast Asian Linguistics No.11, Southeast Asian Syntax*. 175 230. Canberra: Pacific Linguistics.
- Clark, M. 1996. Where do you feel? Stative verbs and body-part terms in Mainland Southeast Asia. In Chappell, H. and McGregor, W. (eds.) *The Grammar of Inalienability*. 529 64. Berlin: De Gruyter Mouton.
- Clark, M. and Prasithrathsint, A. 1985. Synchronic lexical derivation in Southeast Asian languages. In Ratanakul, S. and Premsrirat, S. (eds.) *Southeast Asian Linguistic Studies Presented to André-G. Haudricourt*. 34 81. Bangkok: Mahidol University.
- Clynes, A. and Deterding, D. 2011. Standard Malay (Brunei). *Journal of the International Phonetic Association*. 41, 259 68.
- Cœdès, G. 1925. *Tamnan akson Thai* [History of the Thai writing]. Phra Nakhon: Sophonphiphatthanakon.
- Cœdès, G. 1968. *The Indianized States of Southeast Asia*. Hawaii: East-West Center Press.
- Comrie, B. 1985. *Aspect*. Cambridge: Cambridge University Press.
- Comrie, B. (ed.) 1990. *The Major Languages of East and South-East Asia*. London: Routledge.

- Comrie, B. 2007. Areal typology of mainland Southeast Asia: What we learn from the WALS maps. In Kullavanijaya, P. (ed.) *Trends in Thai Linguistics [Manusya Special Issue 13]*, 18–47. Bangkok: Chulalongkorn University.
- Comrie, B. and Zamponi, R. 2019 Subgrouping and lexical distance in the Great Andamanese family. In Kotin, M. L. (ed.) *Wortschäfte und Sprachwelten: Beiträge zu Sprachtypologie, kontrastiver Wort- bzw. Wortschatzforschung und Pragmatik*. 35–57. Berlin: Lang.
- Cooke, J. R. 1968. *Pronominal Reference in Thai, Burmese and Vietnamese*. Berkeley, CA: University of California Press.
- Cooper, J. S. 1966. Haläng verb phrase. In Thomas, D. D. (ed.) *Papers on four Vietnamese languages*. 28–34. Auckland: Linguistic Society of New Zealand.
- Coupe, A. R. 2003. *A Phonetic and Phonological Description of Ao: a Tibeto-Burman Language of Nagaland, North-East India*. Canberra: Pacific Linguistics.
- Coupe, A. R. 2007. *A Grammar of Mongsen Ao*. Mouton Grammar Library 39. Berlin: Mouton de Gruyter.
- Coupe, A. R. 2017. Mongsen Ao. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 2nd ed. 277–301. Abingdon, Oxon and New York: Routledge.
- Court, C. 1971. A fleeting encounter with the Moken (the Sea Gypsies) in Southern Thailand: Some linguistic and general notes. *Journal of the Siam Society* 59: 83–96.
- Culas, C. and Michaud, J. 2004. A contribution to the study of Hmong (Miao) migrations and history. 61–96. In Tapp, N. et al. (eds.) *Hmong/Miao in Asia*. Chiang Mai: Silkworm Books.
- Darcause, K. N., Chan, C. W., Arunotai, N. H. and Lum, J. K. 2009. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *Journal of Human Genetics* 54: 86–93.
- Davis, J. E. 2010. *Hand Talk: Sign Language among American Indian Nations*. Cambridge: Cambridge University Press.
- Davis, J. E. 2015. North American Indian Sign Language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiungi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 911–931. Berlin: De Gruyter Mouton.
- DeFrancis, J. 1977. *Colonialism and Language Policy in Viet Nam*. The Hague: Mouton.
- DeFrancis, J. 1984. *The Chinese Language: Fact and fantasy*. Honolulu: University of Hawai‘i Press.
- DeLancey, S. 1986. Toward a history of Tai classifier systems. In Craig, C. (ed.) *Noun Classes and Categorization*. 437–52. Amsterdam: John Benjamins.
- DeLancey, S. 2011. Notes on verb agreement prefixes in Tibeto-Burman. *Himalayan Linguistics*, 10 (1). <https://doi.org/10.5070/H910123565>
- Demeter, F., Shackelford, L., Westaway, K., Barnes, L., Duringer, P., Ponche, J.-L., Dumoncel, J., Sénégas, F., Sayavongkhamdy, T., Zhao, J.-X., Sichanthongtip, P., Patole-Edoumba, E., Dunn, T., Zachwieja, A., Coppens, Y., Willerslev, E., Bacon, A.-M. 2017. Early modern humans from Tam Pà Ling, Laos: Fossil review and perspectives. *Current Anthropology* 58 (17): S527–S538. DOI:10.1086/694192
- Dempwolff, O. 1934–1938. *Vergleichende Lautlehre des Austronesischen Wortschatzes*. 3 vols. Beihefte Zeitschrift für Eingeborensprachen 15, 17, 19. Berlin: Dietrich Reimer Verlag.
- Denny, J. P. 1978. Locating the universals in lexical systems for spatial deixis. *Papers from the Parasession on the Lexicon*. 71–84. Chicago: Chicago Linguistic Society.
- DiCanio, C. T. 2007. The phonetics of register in Takhian Thong Chong. *Phonology Lab Annual Report*. 459–82. University of California, Berkeley. <http://linguistics.berkeley.edu/phonlab/documents/2007/ChristianDiCanio.pdf>
- Diffløth, G. 1972. Notes on expressive meaning. In *Papers from the 8th Regional Meeting of the Chicago Linguistic Society*. 440–7. Chicago: Chicago Linguistic Society.

- Diffloth, G. 1994. 'I: Big, a: Small'. In Hinton, L. and Leanne, J. (eds) *Sound Symbolism*. Cambridge: Cambridge University Press.
- Diffloth, G. 2005. The contribution of linguistic palaeontology to the homeland of Austro-Asiatic. In Sagart, L., Blench, R. and Sanchez-Mazas, A. (eds.) *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*. 77–80. London: RoutledgeCurzon.
- Diffloth, G. 2011. Austroasiatic word histories: boat, husked rice and taro. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 295–313. Canberra: Pacific Linguistics.
- Diffloth, G. and Zide, N. H. 1992. Austro-Asiatic languages. In Bright, W. (ed.) *International Encyclopedia of Linguistics*. Vol. 1. 137–42. Oxford: Oxford University Press.
- Diller, A. 2007. Lexical innovation and the 'genius of the Thai Language'. *International Journal of the Sociology of Language*. 186: 13–28.
- Diller, A. 2008. Introduction. In Diller, A., Edmondson, J. A. and Luo, Y. (eds.) *The Tai-Kadai Languages*. Abingdon, Oxon and New York: Routledge.
- Diller, A., Edmondson, J. A. and Luo, Y. (eds.) 2008. *The Tai-Kadai Languages*. Abingdon, Oxon and New York: Routledge.
- Dingemanse, M. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*. 6: 654–72. DOI:10.1002/lnc3.361
- Dingemanse, M. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *STUF*. 70.2: 363–84. DOI:10.1515/stuf-2017_0018
- Dixon, R. M. W. 2008. Complement clauses and complementation strategies in typological perspective. In Dixon, R. M. W. and Aikhenvald, A. Y. (eds.) *Complementation: A cross-linguistic typology*. Oxford: Oxford University Press.
- Dixon, R. M. W. 2010. *Basic Linguistic Theory*. Oxford: Oxford University Press.
- Dixon, R. M. W. and Aikhenvald, A. Y. (eds.) 2002. *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. and Aikhenvald, A. Y. (eds.) forthcoming. *Word*. Oxford: Oxford University Press.
- Đỗ, T. D., Trần, T. H. and Boulakia, G. 1998. Intonation in Vietnamese. In Hirst, D. and Cristo, A. D. (eds.) *Intonation Systems: A survey of twenty languages*. 395–416. Cambridge: Cambridge University Press.
- Do-Hurinville, D. T. and Dao, H. L. 2019. Vietnamese. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 384–431. Berlin: De Gruyter Mouton.
- Doan, T. T. and Mai, N. C. 1992. *Tiếng Dao* [Yao language]. Translated into Chinese by Meihua Pan, 2013, Publishing House of Minority Nationalities, China. Hanoi: Xuất Bản Khoa Học Xã Hội.
- Domeier, M. 2018. Early Paleozoic tectonics of Asia: Towards a full-plate model. *Geoscience Frontiers*. 9 (3): 789–862. <https://doi.org/10.1016/j.gsf.2017.11.012>
- Donegan, J. and Stampe, D. 1983. Rhythm and the holistic organization of language structure. In Richardson, J. F., Marks, M. and Chukerman, A. (eds.) *The Interplay of Phonology, Morphology, and Syntax*. 337–53. Chicago: Chicago Linguistic Society.
- Donegan, J. and Stampe, D. 2002. South-East Asian Features in the Munda Languages: Evidence for the Analytic-to-synthetic Drift of Munda. In *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society, Special Session on Tibeto-Burman and Southeast Asian Linguistics, in Honor of Prof. James A. Matisoff*. 111–29. University of California, Berkeley.
- Donohue, M., Hetherington, R., McElvenny, J. and Dawson, V. 2013. *World Phonotactics Database*. Canberra: Australian National University.
- Downer, G. B. 1963. Chinese, Thai, and Miao-Yao. In Shorto, H. L. (ed.), *Linguistic Comparison in South East Asia and the Pacific*. 133–9. London: School of Oriental and African Studies.
- Downer, G. B. 1967. Tone-change and tone-shift in White Miao. *BSOAS* 30: 589–99.

- Dryer, M. S. 2003. Word order in Sino-Tibetan languages from a typological and geographical perspective. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 43–55. Abingdon, Oxon and New York: Routledge.
- Dryer, M. S. and Haspelmath, M. (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info>
- Dudley, S. 2000. Displacement and identity. Karen refugees in Thailand. DPhil dissertation. University of Oxford.
- Dunn, M., Burenhult, N., Kruspe, N., Tufvesson, S. and Becker, N. 2011. Aslian linguistic prehistory: A case study in computational phylogenetics. *Diachronica* 28 (3): 291–323.
- Dunn, M., Kruspe, N. and Burenhult, N. 2013. Time and place in the prehistory of the Aslian languages. *Human Biology* 85: 383–99.
- Edmondson, J. A. and Esling, J. H. 2006. The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology* 23 (2): 157–91.
- Edmondson, J. A., and Gregerson, K. J. 1993. Western Cham as a register language. In Edmondson, J. A. and Gregerson, K. J. (eds.) *Tonality in Austronesian Languages*. 61–74. Honolulu: University of Hawai'i Press.
- Edmondson, J. A. and Gregerson, K. J. 2007. The languages of Vietnam: Mosaics and expansions. *Language and Linguistics Compass* 1 (6): 727–49.
- Edmondson, J. A., and Solnit, D. B. 1988. Introduction. In Edmondson, J. A. and Solnit, D. B. (eds.) *Comparative Kadai: Linguistic studies beyond Tai*. 1–26. SIL International.
- Edmondson, J. A., and Solnit, D. B. 1997. Introduction. In Edmondson, J. A. and Solnit, D. B. (eds.) *Comparative Kadai: The Tai branch*. 1–32. SIL International.
- Ehrman, M. E. 1972. *Contemporary Cambodian: Grammatical sketch*. Washington: Foreign Service Institute.
- Emiliani, C. 1955. Pleistocene temperatures. *Journal of Geology* 63: 538–78.
- Emmorey, K. 2001. Space on hand: The exploitation of signing space to illustrate abstract thought. In Gattis, M. (ed.) *Spatial Schemas and Abstract Thought*. 147–74. Cambridge, MA: MIT Press.
- Enfield, N. J. 1999. *Lao as a national language*. In Evans, G. (ed.) *Laos: Culture and society*. 258–90. Chiang Mai: Silkworm Books.
- Enfield, N. J. 2001. ‘Lip-pointing’: A discussion of form and function with reference to data from Laos. *Gesture*, 1 (2): 185–211.
- Enfield, N. J. 2003. *Linguistic Epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London: RoutledgeCurzon.
- Enfield, N. J. 2004. On linear segmentation and combinatorics in co-speech gesture: A symmetry-dominance construction in Lao fish trap descriptions. *Semiotica* 149 (1/4): 57–123.
- Enfield, N. J. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34: 181–206.
- Enfield, N. J. 2006a. Languages as historical documents: The endangered archive in Laos. *South East Asia Research* 14 (3): 471–88.
- Enfield, N. J. 2006b. Heterosemy and the grammar-lexicon trade-off. In: Ameka, F., Dench, A. and Evans, N. (eds.) *Catching Language*, 297–320. Berlin: Mouton de Gruyter.
- Enfield, N. J. 2007. *A Grammar of Lao*. Berlin: Mouton de Gruyter.
- Enfield, N. J. 2009. Review of *Serial Verb Constructions: A cross-linguistic typology* edited by Alexandra Y. Aikhenvald and R. M. W. Dixon. *Language* 85: 445–51. DOI:10.1353/lan.0.0124
- Enfield, N. J. (ed.) 2011a. *Dynamics of Human Diversity in Mainland Southeast Asia*. Canberra: Pacific Linguistics.
- Enfield, N. J. 2011b. Linguistic diversity in mainland Southeast Asia. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 63–80. Canberra: Pacific Linguistics.

- Enfield, N. J. 2013. *Relationship Thinking: Agency, enchrony, and human sociality*. New York: Oxford University Press.
- Enfield, N. J. 2015. *The Utility of Meaning: What words mean and why*. Oxford: Oxford University Press.
- Enfield, N. J. 2017. Linguistic expression of commands in Lao. In Aikhenvald, A. Y. and Dixon, R. M. W. (eds.) *Commands: A cross-linguistic typology*: 189–205. Cambridge: Cambridge University Press.
- Enfield, N. J. 2020. Word in Lao. In Aikhenvald, A. Y., Dixon, R. M. W. and White, N. M. (eds.) *Phonological Word and Grammatical Word*. 176–212. Oxford: Oxford University Press.
- Enfield, N. J. and Diffloth, G. 2009. Phonology and sketch grammar of Kri, a Vietic language of Laos. *Cahiers de Linguistique Asie Orientale* 38 (1): 3–69.
- Enfield, N. J. and Wierzbicka, A. 2002. Introduction: The body in description of emotion. *Pragmatics and Cognition*, 10 (1): 1–24.
- Enwall, J. 1994. *A Myth Become Reality: History and development of the Miao written language*. East Asia Monographs. Stockholm: Institute of Oriental Languages, Stockholm University.
- Farnell, B. 1995. *Do you see what I mean? Plains Indian sign talk and the embodiment of action*. Austin: University of Texas Press.
- Fedden, S., and Corbett, G. G. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A journal of general linguistics*, 2(1).
- Ferlus, M. 1992. Essai de phonétique historique du Khmer (Du milieu du premier millénaire de notre ère à l'époque actuelle). *Mon-Khmer Studies* 21: 57–89.
- Ferlus, M. 2014. Arem, a Vietic language. *Mon-Khmer Studies* 43 (1): 1–15.
- Finot, L. 1959. Laotian writings. In de Berval, R. et al. (eds) *Kingdom of Laos*. Saigon: France-Asie.
- Fix, A. 2011. Origin of genetic diversity among Malaysian Orang Asli: An alternative to the demic diffusion model. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 277–91. Canberra: Pacific Linguistics.
- Foley, W. A., and Olson, M. 1985. Clausehood and verb serialisation. In Nichols, J. and Woodbury, A. C. (eds.) *Grammar inside and outside the clause: Some approaches to theory from the field*. 17–60. Cambridge: Cambridge University Press.
- Forestier, H., Sophady, H. and Celiberti, V. 2017. Le techno-complexe hoabinhien en Asie du Sud-est continentale: L'histoire d'un galet qui cache la forêt [The Hoabinhian techno-complex in Mainland Southeast Asia: The history of a pebble which hides the forest]. *Journal of Lithic Studies*, 4 (2): 305–49. <https://doi.org/10.2218/jls.v4i2.2545>
- Fuller, D. Q. 2011. Pathways to Asian Civilizations: Tracing the origins and spread of rice and rice cultures. *Rice*, 4 (3): 78–92.
- Fuller, J. W. 1985. Topic and comment in Hmong. PhD dissertation. University of Minnesota.
- Gage, W. W. 1987. Vietnamese in Mon-Khmer perspective. In Ratanakul, S., Thomas, D. and Premisirat, S. (eds.) *Southeast Asian Studies presented to Andre-G. Haudricourt*. 493–524. Bangkok: Mahidol University.
- Gai, X. 2002. *A Study of Tanglang*. Beijing: The Nationalities Press.
- Gao, Y. 2004. *Buxing yu yan jiu / 布兴语研究* [A study of the Buxing language]. Beijing: Min zu chu ban she.
- Gedney, W. J. 1947. Indic loanwords in spoken Thai. Doctoral dissertation. Yale University.
- Gedney, W. J. 1972. A Puzzle in Comparative Tai Phonology. In Harris, J. G. and Noss, R. B. (eds.) *A Conference on Tai Phonetics and Phonology*. 52–7. Bangkok: Mahidol University.
- Gedney, W. J. 1989. A checklist for determining tones in Tai dialects. In Bickner, R. J., Hartmann, J., Hudak, T. J. and Peyasantiwong, P. (eds.) *Selected Papers on Comparative Tai Studies*. 191–206. Center for South and Southeast Asian studies, the University of Michigan.

- Giaphong, S. 2004. Plang Grammar as spoken in Huay Namkun village, Chiang Rai province. MA thesis. Bangkok: Mahidol University.
- Gil, D. 2005. Word order without syntactic categories: How Riau Indonesian does it, in Carnie, A., Harley, H. and Dooley, S. A. (eds.) *Verb First: On the syntax of verb-initial languages*. 243–63. Amsterdam: John Benjamins.
- Gil, D. 2015. The Mekong-Mamberamo linguistic area. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 266–355. Berlin: De Gruyter Mouton.
- Goddard, C. 2005. *The Languages of East and Southeast Asia*. Oxford: Oxford University Press.
- Goldin-Meadow, S. 2003a. *Hearing Gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- Goldin-Meadow, S. 2003b. *The Resilience of Language: What gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press.
- Gomi, T. 1989. *An Illustrated Dictionary of Japanese Onomatopoeic Expressions*. Translated by J. Turrent. Tokyo: Japan Times.
- Goody, J. 1977. *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.
- Gordon, M. K. 2016. *Phonological Typology*. Oxford: Oxford University Press.
- Gorman, C. 1969. Hoabinhian: A pebble tool complex with early plant associations in Southeast Asia. *Science* 163: 671–3.
- Gorman, C. 1970. Excavations at Spirit cave, North Thailand: Some interim interpretations. *Asian Perspectives*. 13: 79–107.
- Gorman, C. 1971. The Hoabinhian and after: Subsistence patterns in Southeast Asia during the late Pleistocene and early recent periods. *World Archaeology* 2: 300–20.
- Grant, A. and Sidwell, P. 2005. *Chamic and Beyond: Studies in mainland Austronesian languages*. Canberra: Pacific Linguistics.
- Gray, R. D. and Jordan, F. M. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405: 1052–5.
- Green, J., and Wilkins, D. P. 2015. Arandic alternate sign language(s). In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign languages of the World: A comparative handbook*. 843–69. Berlin: De Gruyter Mouton.
- Greenberg, J. H. 1970. Some generalizations concerning glottallic consonants, especially implosives. *International Journal of American Linguistics*, 36 (2): 123–45.
- Greenhill, S. J. and Gray, R. D. 2009. Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In Adelaar, A. and Pawley, A. (eds.) *Austronesian Historical Linguistics and Culture History: A festschrift for Robert Blust*. Canberra: Pacific Linguistics.
- Gregerson, K. 1976. Tongue-root and register in Mon-Khmer. In Jenner, P. (ed.) *Austroasiatic Studies I*. 323–66. Honolulu: University of Hawai'i Press.
- Gregerson, K. 1979. *Predicate and argument in Rengao grammar*. Dallas, TX: University of Texas at Arlington, SIL.
- Gregerson, K. and Smith, K. D. 1973. The development of Todrah register. *Mon-Khmer Studies* 4: 143–84.
- Grierson, G. A. 1904. *Linguistic Survey of India (II): Môn-Khmér and Siamese-Chinese families (including Khassi and Tai)*. Calcutta: Office of the Superintendent of Government Printing.
- Grinevald, C. 2000. A morphosyntactic typology of classifiers. In Senft, G. (ed.) *Systems of Nominal Classification*, 50–92. Cambridge: Cambridge University Press.
- Guion, S. G., and Wayland, R. P. 2004. Aerodynamic coarticulation in sound change or how onset trills can condition a falling tone. In Agwuele, A., Warren, W. and Park, S.-H. (eds.) *Proceedings of the 2003 Texas Linguistic Society Conference*. 107–15. Cascadilla Proceedings Project.

- Guo, J. 1993. *Hànyǔ Shēngdiào Yúdiào Chǎnyào yú Tānsuo* [The study and exploration of Chinese tone and intonation]. Beijing: Beijing Language and Culture University Press.
- Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Hạ, K. P. 2010. Prosody of Vietnamese from an interactional perspective: ò, ù and vâng in backchannels and requests for information. *Journal of the Southeast Asian Linguistics Society* 3 (1): 56–76.
- Hạ, K. P. 2012. *Prosody in Vietnamese: Intonational form and function of short utterances in conversation*. Asia-Pacific Linguistics (SEAMLES).
- Hạ, K. P. 2013. Prosodic means in repair initiation as an activity in Northern Vietnamese conversation. In Hole, D. and Löbel, E. (eds.) *Linguistics of Vietnamese: An international survey*. 35–54. Berlin and New York: De Gruyter Mouton.
- Hạ, K. P. and Grice, M. 2010. Modelling the interaction of intonation and lexical tone in Vietnamese. Paper presented at Speech Prosody 2010, Chicago.
- Haas, M. R. 1942 The use of numeral classifiers in Thai, *Language* 18: 201–5.
- Haas, M. R. 1969. Sibling terms as used by marriage partners. *Southwestern Journal of Anthropology* 25 (3): 228–35.
- Haiman, J. 1983. Iconic and economic motivation. *Language* 59 (4): 781–819.
- Haiman, J. 1985. *Natural Syntax: Iconicity and erosion*. Cambridge: Cambridge University Press.
- Haiman, J. 2011. *Cambodian: Khmer*. London Oriental and African Language Library 16. Amsterdam: John Benjamins.
- Haiman, J. 2019. Khmer. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 320–83. Berlin: De Gruyter Mouton.
- Halcrow, S. E. and Tayles, N. 2011. Human diversity in Mainland Southeast Asia: The contribution of bioarchaeology. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 47–61. Canberra: Pacific Linguistics.
- Halcrow, S. E., Tayles, N., and King, C. L. 2016. Infant and child health and disease with agricultural intensification in mainland Southeast Asia. In Oxenham, M. and Buckley, H. (eds.) *Routledge Handbook of Bioarchaeology in Southeast Asia and the Pacific Islands*. 158–86. Abingdon, Oxon: Routledge.
- Hall, D. G. E. 1966. *A History of Southeast Asia*. New York: St Martin's Press.
- Hall, K. R. 1992. Economic history of early Southeast Asia. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia*. Vol. 1. 183–275. Cambridge: Cambridge University Press.
- Hall, R. 1996. Reconstructing Cenozoic SE Asia. In: Hall, R. and Blundell, D. (eds.) *Tectonic Evolution of SE Asia*. Special Publication No. 106. 153–84. London: Geological Society of London.
- Halliday, M. A. K. and Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Hanihara, T. 2006. Interpretation of craniofacial variation and diversification of East and Southeast Asians. In Oxenham, M. O. and Tayles, N. (eds.) *Bioarchaeology of Southeast Asia*. 91–111. Cambridge: Cambridge University Press.
- Hansson, I.-L. 2003. Akha. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 236–51. Abingdon, Oxon and New York: Routledge.
- Harris, R. 1980. *The Language-makers*. Ithaca, NY: Cornell University Press.
- Hartmann, H. 1987. Time expressions in Daai Chin. *Linguistics of the Tibeto-Burman Area*, 10 (1): 147–51.
- Hartmann, J. 1986. The spread of South Indic scripts in Southeast Asia.' *Crossroads: An Interdisciplinary Journal of Southeast Asian Studies*, 3 (1) 1986: 6–20.
- Hartmann, J. 1998. A linguistic geography and history of Tai Meuang-Fai (Ditch-Dike) techno-culture. *Language and Linguistics* 16 (2): 67–101.

- Haudricourt, A.-G. 1954. De l'origine des tones en Vietnamien. *Journal Asiatique*, 242: 69–82.
- Haudricourt, A.-G. 1965. Les mutations consonantiques et les occlusives initiales en mon-khmer. *Bulletin de la Société Linguistique de Paris* 60: 160–72.
- Haudricourt, A.-G. 1966. The limits and connections of Austroasiatic in the Northeast. In Zide N. H. (ed.) *Studies in Comparative Austroasiatic Linguistics*. 44–56. The Hague: Mouton.
- Haudricourt, A.-G. 2010. The origin of the peculiarities of the Vietnamese alphabet. *Mon-Khmer Studies* 39: 89–104. Translated from ‘L'origine Des Particularités de L'alphabet Vietnamien’ (1949) *Dân Việt-Nam* 3: 61–8.
- Hayes, L. V. H. 1984. The register systems of Thavung. *Mon-Khmer Studies*. 12: 91–122.
- He, Diyong. 1993. *Kèjiā Fāngyán Yúfǎ Yánjiū* [A study of Hakka grammar]. Xiamen: Xiamen University Press.
- Headley, R. K. 1998. Cham evidence for Khmer sound changes. *Pacific Linguistics*. Series A. Occasional Papers. 89: 21–9.
- Heder, S. R. 2007. Cambodia. In Simpson, A. (ed.) *Language and National Identity in Asia*. 288–311. Oxford: Oxford University Press.
- Heimbach, E. E. 1979. *White Hmong-English Dictionary*. Revised edition. Linguistic Series IV, Data Paper no. 75, Southeast Asia Program, Cornell University. Ithaca, NY: Southeast Asia Program.
- Heine, B. 1997. *Possession: Cognitive Sources, Forces and Grammaticalization*. Cambridge: Cambridge University Press.
- Heine-Geldern, R. 1956. *Conceptions of State and Kingship in Southeast Asia*. Ithaca, NY: Cornell University.
- Henderson, E. J. A. 1952. The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14. 149–74.
- Henderson, E. J. A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15. 400–34.
- Henderson, E. J. A. 1967. Grammar and tone in South East Asian languages. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig* 16 (1/2): 171–8.
- Hickman, J. R. 2019. Writing a New World Into Being: Orthographic ideology and millenarian religion in Hmong Society. Paper presented at The Anthropology of Language in Mainland Southeast Asia. Sydney, 19 August.
- Higbie, J. and Thinsan, S. 2003. *Thai Reference Grammar: The structure of spoken Thai*. Bangkok: Orchid Press.
- Higham, C. 1989. *The Archaeology of Mainland Southeast Asia: From 10,000 BC to the fall of Angkor*. Cambridge: Cambridge University Press.
- Higham, C. 1996. *The Bronze Age of Southeast Asia*. Cambridge: Cambridge University Press.
- Higham, C. 2001. *The Civilization of Angkor*. Berkeley, CA: University of California Press.
- Higham, C. 2002. *Early Cultures of Mainland Southeast Asia*. Bangkok: River Books.
- Hill C., Soares P., Mormina M., Macaulay V., Meehan W. et al. 2006 Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol* 23: 2480–91.
DOI:10.1093/molbev/msl124. PubMed: 16982817
- Himmelmann, N. P. and Schultze-Berndt, E. (eds.) 2005. *Secondary Predication and Adverbial Modification*. Oxford: Oxford University Press.
- Hnin Tun, S. S. 2014. *Colloquial Burmese: The complete course for beginners*. Routledge.
- Hoang, T., Lợi, N. V., Phúc, D. V. and Thành, P. X. 1986a. *Sách Học Tiếng Bru Văn Kiều* [A text to study the Bru language]. Hanoi: Ủy Ban Nhân Dân.
- Hoang, T., Thắng, L. T., Thông, T. V., K'bren and K'broh. 1986b. *Ngữ Pháp Tiếng Koho* [A Grammar of Koho]. Lam Dong, Vietnam: Sở Văn Hóa và Thông Tin Lâm Đồng.
- Hoàng, V. H. 1998. Từ Láy Đôi trong Tiếng Mường [Reduplication in Muong]. In *Từ Láy: Những Vấn Đề Còn Đề Ngó* [Reduplication: Remaining Issues], 165–93. Hanoi: Institute of Linguistics.

- Hoàng, V. M. and Tạ, V. T. 1998. *Tiếng Bru-Vân Kiêu* [The Bru-Van Kieu language]. Hanoi: Nhà Xuất Bản Khoa Học Xã Hội.
- Hoijer, H. 1956. Lexicostatistics: A critique. *Language* 32 (1): 49–60.
- Holmer, A. 2009. Atayal clitics and sentence structure. *Working Papers in Linguistics* 40: 71–94.
- Hombert, J.-M., Ohala, J. J., and Ewan, W. G. 1979. Phonetic explanations for the development of tones. *Language* 55: 37–58.
- Hope, G. 2005. The Quaternary in Southeast Asia. In Gupta, A. (ed.) *The Physical Geography of Southeast Asia*. 24–37. New York: Oxford University Press.
- House, D., Karlsson, A., Svantesson, J.-O. and Tayanin, D. 2009. The phrase-final accent in Kammu: Effects of tone, focus and engagement. Paper presented to INTERSPEECH, 2009.
- Hsiu, A. 2018. Preliminary classification of Mienic languages. East Asian Language Phyla Project. <https://sites.google.com/site/eastasiaphyla/hm/mienic>
- Huang, Y. 2010. A grammar of Jingxi Zhuang. Master's thesis. Guangxi University.
- Hudak, T. J. 2008. *William J. Gedney's Comparative Tai Sourcebook*. Honolulu: University of Hawai'i Press.
- Hudson, B., Nyein, L., Maung, W. (Tanpawady). 2001. The origins of Bagan: New dates and old inhabitants. *Asian Perspectives* 40 (1): 48–74.
- Huffman, F. E. 1970. *Modern Spoken Cambodian*. Ithaca, NY: Cornell University, Southeast Asia Program.
- Huffman, F. E. 1973. Thai and Cambodian – A case of syntactic borrowing? *Journal of the American Oriental Society*, 93 (4): 488–509.
- Huffman, F. E. 1976. The relevance of lexicostatistics to Mon-Khmer languages. In P. N. Jenner, P., Thompson, L. C. and Starosta, S. (eds.) *Austroasiatic studies*. 539–74. Honolulu: University of Hawai'i Press.
- Hyman, L. M. 1978. Historical tonology. In Fromkin, V. A. (ed.) *Tone: A linguistic survey*. 257–69. New York: Academic Press. https://doi.org/10.1016/B978-0_12-267350_4.50013_3
- Hyman, L. M. 2001. Tone systems. In Haspelmath, M. et al. (eds) *Language Typology and Language Universals: An international handbook*. Vol. 2. 1367–80. Berlin and New York: Walter de Gruyter.
- Hyman, L. M. and Monaka, K. C. 2011. Tonal and non-tonal intonation in Shekgalagari. In Frota, S., Elordieta, G. and Prieto, P. (eds) *Prosodic Categories: Production, perception and comprehension*. 267–90. Dordrecht: Springer.
- Hyman, L. M. and VanBik, K. 2002. Tone and syllable structure in Hakha-Lai. Berkeley Linguistics Society, Special Session: Tibeto-Burman/Southeast Asian Linguistics, 15 February. <https://bit.ly/39JxbkX>
- Hyman, L. M. and VanBik, K. 2004. Directional rule application and output problems in Hakha Lai tone. *Language and Linguistics* 5 (4): 821–61.
- Hymes, D. H. 1960. Lexicostatistics so far. *Current Anthropology* 1: 3–44.
- Hyslop, G. 2009. Kurtöp tone: A tonogenetic case study. *Lingua*, 119 (6): 827–45. <https://doi.org/10.1016/j.lingua.2007.11.012>
- Iletó, R. 1992. Religion and anti-colonial movements. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia*. Vol. 2. 197–248. Cambridge: Cambridge University Press.
- Iwasaki, S. and Horie, P. I. 2005. *A Reference Grammar of Thai*. Cambridge: Cambridge University Press.
- Jacques, R. 2002. *Portuguese Pioneers of Vietnamese Linguistics Prior to 1650* [Pionniers portugais de la linguistique vietnamienne jusqu'en 1650]. In English and French. Bangkok: Orchid Press.
- Jacquesson, F. 2004. *Le Deuri: Langue Tibéto-Birman d'Assam*. Leuven, Paris and Dudley, MA: Peeters.
- Jaisser, A. 1984. Complementation in Hmong. MA thesis. San Diego State University.

- Jaisser, A. 1990. DeLIVERing an introduction to psycho-collocations with SIAB in Hmong. *Linguistics of the Tibeto-Burman Area*, 13 (1): 159–78.
- Jakobson, R. 1971. Shifters, verbals categories, and the Russian verb. In Jakobson, R., *Selected Writings II: Word and language*. The Hague and Paris: Mouton.
- Jarkey, N. 2006. Complement clause types and complementation strategies in White Hmong. In Dixon, R. M. W. and Aikhenvald Alexandra Y. (eds.) *Complementation: A cross-linguistic typology*. 115–36. Oxford: Oxford University Press.
- Jarkey, N. 2015. *Serial Verbs in White Hmong*. Brill.
- Jenks, R. D. 1994. *Insurgency and Social Disorder in Guizhou: The “Miao” rebellion, 1854–1873*. Honolulu: University of Hawai‘i Press.
- Jenny, M. 2015. The far west of Southeast Asia: ‘Give’ and ‘get’ in the languages of Myanmar. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 155–208. Berlin: De Gruyter Mouton.
- Jenny, M. 2019a. Mon. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 277–319. Berlin: De Gruyter Mouton.
- Jenny, M. 2019b. Thai. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 559–608. Berlin: De Gruyter Mouton.
- Jenny, M. and Hnin Tun, S. S. 2016. *Burmese: A comprehensive grammar*. Abingdon, Oxon and New York: Routledge.
- Jenny, M. and Sidwell, P. (eds.) 2015. *Handbook of the Austroasiatic Languages*. Leiden: Brill.
- Jenny, M., Weber, T. and Weymuth, R. 2015. The Austroasiatic Languages: A typological overview. In Jenny, M. and Sidwell, P. (eds.) *The Handbook of Austroasiatic Languages*. 13–143. Leiden: Brill.
- Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) 2015. *Sign Languages of the World: A comparative handbook*. Berlin: De Gruyter Mouton.
- Ji, X., Kuman, K., Clarke, R. J., Forestier, H., Li, Y., Ma, J., Qiu, K., Li, H. and Wu, Y. 2016. The oldest Hoabinhian technocomplex in Asia (43.5 ka) at Xiaodong rockshelter, Yunnan province, southwest China. *Quaternary International*, 400 (2): 166–74. <https://doi.org/10.1016/j.quaint.2015.09.080>
- Jiang, P. 2000. The semantics of noun phrases in Chinese reciprocal constructions. *Studies in Language and Linguistics* 20 (1): 41–50.
- Jinam, T. A., Phipps, M. E., Aghakhanian, F., Majumder, P. P., Datar, F., Stoneking, M., Sawai, H., Nishida, N., Tokunaga, K., Kawamura, S., Omoto, K. and Saitou, N. 2017. Discerning the origins of the Negritos, First Sundaland people: Deep divergence and archaic admixture. *Genome Biology and Evolution* 9 (8): 2013–22. DOI: <https://doi.org/10.1093/gbe/evx118>
- Johnson, C. 1985. *Dab neeg Hmoob: Myths, legends and folk tales from the Hmong of Laos*. St Paul, MN: Linguistics Department, Macalester College.
- Jones, R. B. 1970. Classifier constructions in Southeast Asia. *Journal of the American Oriental Society* 90 (1): 1–12.
- Jonsson, H. 2011. Ethnology and the issue of human diversity in Mainland Southeast Asia. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 109–22. Canberra: Pacific Linguistics.
- Jonsson, H. 2014. *Slow Anthropology: Negotiating difference with the Iu Mien*. Ithaca, NY: Cornell Southeast Asia Program Publications.
- Jonsson, N. L. 1991. Proto-Southwestern Tai. PhD dissertation. State University of New York at Albany.
- Kamei, N. 2004. Sign language in Myanmar: One language or two? In Foreign Sign Department, Japan Institute for Sign Language Studies (ed.) *Sign Languages of Asia [Asia no shuwa]*, Vol. 5. 10–19. Tokyo: Japan Institute for Sign Language Studies.

- Kang, Z. 2009. Judu Gelao Reference Grammar. PhD thesis. Beijing: Minzu University of China.
- Karlgren, B. 1915. *Etudes sur la phonologie Chinoise*. Leiden: E.-J. Brill.
- Karlgren, B. 1954. Compendium of phonetics in Ancient and Archaic Chinese. *Bulletin of the Museum of Far Eastern Antiquities* 26: 211–367.
- Karlsson, A., House, D. and Svantesson, J.-O. 2012. Intonation adapts to lexical tone: The case of Kammu. *Phonetica* 69: 28–47.
- Karlsson, A., House, D., Svantesson, J.-O. and Tayanin, D. 2007. Prosodic phrasing in tonal and non-tonal dialects of Kammu. Paper presented to the 16th International Congress of Phonetic Sciences, 2007.
- Karlsson, A., House, D., Svantesson, J.-O. and Tayanin, D. 2010. Influence of lexical tones on intonation in Kammu. Paper presented to the INTERSPEECH, 2010.
- Kato, A. 2003. Pwo Karen. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 632–48. Abingdon, Oxon and New York: Routledge.
- Kato, A. 2019. Pwo Karen. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 131–75. Berlin: De Gruyter Mouton.
- Keenan, E. L. and Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic inquiry* 8 (1): 63–100.
- Keller, C. 1976. *A Grammatical Sketch of Brao, a Mon-Khmer Language*. Grand Forks, ND: Summer Institute of Linguistics.
- Kelly, P. 2018. The art of not being legible. *Terrain* 70 <https://journals.openedition.org/terrain/17103>
- Kendon, A. 1988. *Sign Languages of Aboriginal Australia: Cultural, semiotic and communicative perspectives*. Cambridge: Cambridge University Press.
- Kendon, A. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kerr, A. D. 1972. *Lao-English Dictionary*. Washington, DC: Catholic University of America Press.
- Keyes, C. 1977. *The Golden Peninsula: Culture and adaptation in mainland Southeast Asia*. New York: Macmillan.
- King, C. L., Halcrow, S. E., Tayles, N. and Shkrum, S. 2017. Considering the palaeoepidemiological implications of socioeconomic and environmental change in Southeast Asia. *Archaeological Research in Asia* 11: 27–37. <https://doi.org/10.1016/jара.2017.05.003>
- Kirby, J. P. 2014a. Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics* 43: 69–85.
- Kirby, J. P. 2014b. Incipient tonogenesis in Phnom Penh Khmer: Computational studies. *Laboratory Phonology-Journal of the Association for Laboratory Phonology* 5 (1): 195–230.
- Kirby, J., and Brunelle, M. 2017. Southeast Asian tone in areal perspective. In Hickey, R. (ed.) *The Cambridge Handbook of Areal Linguistics*. 703–31. Cambridge: Cambridge University Press.
- Kölver, U. 1991. Local prepositions and serial verb constructions in Thai. In Seiler, H. and Prempér, W. (eds.) *Partizipation: Das sprachliche Erfassen von Sachverhalten*. 485–508. Tübingen: Gunter Narr Verlag.
- Konnert, L. 2014. A Grammar of Karbi. Dissertation. University of Oregon.
- Kosaka, R. 2000. A descriptive study of the Lachi language: Syntactic description, historical reconstruction and genetic relation. Doctoral dissertation. Tokyo University of Foreign Studies.
- Kratochvíl, Paul. 1968. *The Chinese Language Today: Features of an emerging standard*. London: Hutchinson.
- Kratoska, P., and Batson, B. 1992. Nationalism and modernist reform. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia*. Vol. 2. 249–324. Cambridge: Cambridge University Press.
- Kruspe, N. 2004. *A Grammar of Semelai*. Cambridge: Cambridge University Press.
- Kunyot, T. 1984. General characteristics of Hmong Njua grammar. MA thesis. Mahidol University, Bangkok.

- Kurabe, K. 2017. Jinghpaw. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 2nd ed. 993–1010. Abingdon, Oxon and New York: Routledge.
- Kutanan, W., Kampuansai, J., Srikuammol, M., Kangwanpong, D., Ghirotto, S., Brunelli, A. and Stoneking, M. 2017. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. *Human Genetics* 136 (1): 85–98.
- Kwok, Helen. 1984. *Sentence Particles in Cantonese*. Hong Kong: Centre of Asian Studies, University of Hong Kong.
- Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- LaPolla, R. J. 1992. ‘Anti-ergative’ marking in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 15 (1): 1–9.
- LaPolla, R. J. 1994. On the change to verb-medial word order in proto-Chinese: Evidence from Tibeto-Burman. In Kitamura, H., Nishida, T. and Nagano, Y. (eds.), *Current Issues in Sino-Tibetan Linguistics*. 98–104. Osaka: Organizing Committee of the 26th International Conference on Sino-Tibetan Languages and Linguistics.
- LaPolla, R. J. 1995. “Ergative” marking in Tibeto-Burman. In Nishi, Y., Matisoff, J. A. and Nagano, Y. (eds.) *New Horizons in Tibeto-Burman Morphosyntax*. 189–228. Osaka: National Museum of Ethnology.
- LaPolla, R. J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In Aikhenvald, A. Y. and Dixon, R. M. W. (eds.) *Areal Diffusion and Genetic Inheritance: Problems in comparative linguistics*. 225–54. Oxford, New York: Oxford University Press.
- LaPolla, R. J. 2003. Overview of Sino-Tibetan syntax. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 22–42. Abingdon, Oxon and New York: Routledge.
- LaPolla, R. J. 2012. Comments on methodology and evidence in Sino-Tibetan comparative linguistics. *Language and Linguistics* 13 (1): 117–32.
- LaPolla, R. J. 2013. Subgrouping in Tibeto-Burman: Can an individual-identifying standard be developed? How do we factor in the history of migrations and language contact? In Bickel, B., Grenoble, L. A., Peterson, D. A., Timberlake, A. (eds.) *Language Typology and Historical Contingency: In honor of Johanna Nichols*. 463–76. Amsterdam: John Benjamins.
- Larish, M. D. 1993. Who are the Moken and Moklen on the Islands and Coasts of the Andaman Sea? In *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics*, 1305–19. Bangkok: Chulalongkorn University.
- Larish, M. D. 1997 Moklen-Moken phonology: Mainland or insular Southeast Asian typology? In Odé, C. and Stokhof, W. (eds.) *Proceedings of the Seventh International Conference on Austronesian Linguistics*. 125–50. Amsterdam: Rodopi.
- Larish, M. D. 1999. The position of Moken and Moklen within the Austronesian language family. PhD, University of Hawai‘i.
- Larish, M. D. 2005. Moken and Moklen. In Adelaar, A. and Himmelmann, N. P. (eds.) *The Austronesian Languages of Asia and Madagascar*, 513–33. Abingdon, Oxon and New York: Routledge.
- Larson, G. et al. 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *PNAS* 107 (17): 7686–91. <https://doi.org/10.1073/pnas.0912264107>
- Lê, M.-H. and O’Harrow, S. 2007. Vietnam. In Simpson, A. (ed.) *Language and National Identity in Asia*. 415–41. Oxford: Oxford University Press.
- Leach, E. 1954. *Political Systems of Highland Burma*. London: The Athlone Press.
- Lee, E. W. 1966. Proto-Chamic phonologic word and vocabulary. PhD dissertation. Indiana University.
- Lee, E. W. 1996. Bipartite negatives in Chamic. *Mon-Khmer Studies* 26: 291–317.

- Levinson, S. C., Cutfield, S., Dunn, M., Enfield, N. J. and Meira, S. 2018. *Demonstratives in Cross-linguistic Perspective*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/9781108333818>
- Lexicography Office, Institute of Linguistics, Chinese Academy of Social Sciences. 2012. *A Dictionary of Modern Chinese*. 6th edition. Beijing: The Commercial Press.
- Li, C. N. and Thompson, S. A. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, Los Angeles and London: University of California Press.
- Li, D. 2003. *A Study of Geman*. Beijing: The Nationalities Press.
- Li, D. 2004. *Sulong Yu Yan Jiu* [A study of Sulong]. Beijing: Minzu Chubanshe.
- Li, F. K. 1977. *A Handbook of Comparative Tai*. Honolulu: University of Hawai'i Press.
- Li, J. 2008. Chadong, a newly-discovered Kam-Sui language in northern Guangxi. In Diller, A., Edmondson, J. A. and Luo, Y. (eds.) *The Tai-Kadai Languages*. 509–84. Abingdon, Oxon and New York: Routledge.
- Li, X. 2013. *Numerical Classifiers in Chinese: The syntax-semantics interface*. Berlin: De Gruyter Mouton.
- Li, Y. 2003. *A Study of Sangkong*. Beijing: The Nationalities Press.
- Li, Y. C. and Yip, M. 1979. Ba Construction and Ergativity in Chinese. In Plank, F. (ed.), *Ergativity: Towards a theory of grammatical relations*. London and New York: Academic Press, 103–14.
- Liang, M. and Zhang, J. 1996. *Dong Tai yu zu gai lun* [An introduction to the Kam-Tai languages]. Beijing: China Social Sciences Academy Press.
- Lidz, L. 2010. A descriptive grammar of Yongning Na (Mosuo). Doctoral dissertation. University of Texas at Austin.
- Lidz, L. 2019. Yongning Na (Mosuo). In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 234–76. Berlin: De Gruyter Mouton.
- Liem, N. D. 1970. *Four-syllable Idiomatic Expressions in Vietnamese*. Canberra: Australian National University.
- Lin, P. T. 2015. *Taiwanese Grammar: A concise reference*. Greenhorn Media.
- Liu, H. 2016. *Yuèbēi Rúyuán Guòshān Miǎnyù Yánjiū* [A study of Yuebei Ruyuan Guoshan Mien]. Beijing: Culture and Art Publishing House.
- Liu, L. 2013. Analysis of ba marker in Chinese. *Theory and Practice in Language Studies*. 3 (12): 2227.
- Liu, L., Wang, H., and Bo, Y. 2007. *Xiāndài Hanyu fāngyán héxīnci, tèzhèngcì jí* [Core vocabulary and diagnostic vocabulary of Modern Chinese Dialects]. Nanjing: Feng huang chu ban she.
- Liu, Y. 2012. Thailand Mien reference grammar. PhD thesis. Minzu University of China, Beijing.
- Long, Y. and Zheng, G. 1998. *The Dong Language in Guizhou Province, China*. Publications in Linguistics, 126. Translated by Norman Geary. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.
- Longley, I. M. 1997. The tectonostratigraphic evolution of SE Asia. In Fraser, A. J., Matthews, S. J. and Murphy, R. W. (eds.) *Petroleum Geology of Southeast Asia*. 311–39. Geological Society Special Publication 126.
- Lu, T. Q. 2008. *A Grammar of Maonian*. Boca Raton, FL: Universal Publishers.
- Luce, G. H. 1985. *Phases of Pre-Pagan Burma: Languages and history*. Oxford: Oxford University Press.
- Lucey-Weinhold, Meredith. 2011. Serial verb constructions in Sgaw Karen: a comparison of Karenic VP structures. In Brunelle, M. (ed.) *Sgaw Karen papers: Presented to Nimrod Andrew*. 76–84. Ottawa: University of Ottawa.
- Luke, K. K. 1990. *Utterance Particles in Cantonese Conversation*. Amsterdam and Philadelphia: John Benjamins.

- Luksameeyanawin, S. 1998. Intonation in Thai. In Hirst, D. and Cristo, A. D. (eds.) *Intonation Systems*. 376–94. Cambridge: Cambridge University Press.
- Luo, Y. 2008. Zhuang. In Diller, A., Edmondson, J. A. and Luo, Y. (eds.) *The Tai-Kadai Languages*. 317–77. Abingdon, Oxon and New York: Routledge.
- Luong, H. V. 1988. Discursive practices and power structure: Person-referring forms and sociopolitical struggles in colonial Vietnam. *American Ethnologist* 15 (2): 239–53.
- Luong, H. V. 1990. *Discursive Practices and Linguistic Meanings: The Vietnamese system of person reference*. Amsterdam: John Benjamins.
- Lutz-Hughes, M. 2016. Defining serial verb constructions in Kri. Honours dissertation. University of Sydney.
- Lyman, T. A. 1974. *Dictionary of Mong Njua, a Miao (Meo) Language of Southeast Asia*. The Hague and Paris: Mouton.
- Lyman, T. A. 1979. *Grammar of Mong Njua (Green Miao): A descriptive linguistic study*. Sattley, CA: Blue Oak Press.
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D. et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034–6. DOI:10.1126/science.1109792. PubMed: 15890885
- Maddieson, I. 2013. Consonant inventories. In Dryer, M. S. and Haspelmath, M. (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/1>
- Majid, A. and Burenhult, N. 2014. Odors are expressible in language, as long as you speak the right language. *Cognition* 130 (2): 266–70.
- Manley, T. M. 1972. *Outline of Sre Structure*. Oceanic Linguistics Special Publication No. 12. Honolulu: University of Hawai‘i Press.
- Manson, K. 2011. The subgrouping of Karen. Paper presented at the 21st Annual Meeting of the Southeast Asian Linguistics Society, Kasetsart University, Bangkok, 11–13 May.
- Mao, Z. 2004. Yaozu mianyu fangyan yanjiu [A study of the Mien language of the Yao ethnicity]. In *China Ethnic Minority Language and Dialect Studies*. Beijing: Publishing House of Minority Nationalities.
- Mao, Z. and Li, Y. 2002. *A Study of Jiongnai*. Beijing: Central Nationalities University Press.
- Mao, Z. and Li, Y. 2007. *Younuoyu Yanjiu* [A study of Younuo]. Beijing: Minzu University of China Publishing House.
- Marrison, G. E. 1975. The early Cham language and its relationship to Malay. *Journal of the Malaysian Branch of the Royal Asiatic Society* 48 (2): 52–9.
- Martinsson, D. G., Pisias, N. G., Hays, J. D., Imbrie, J., Moore, T. C. and Shackleton, N. J. 1987. Age dating and orbital theory of the Ice Ages: Development of a high resolution 0–300 000-year chronostratigraphy. *Quaternary Research* 27: 1–29.
- Masica, C. 1991. *Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Maspéro, H. 1911. Contribution d l'étude du système phonétique des langues thai. *Bulletin de l'École Française d'Extrême-Orient* 11: 153–69.
- Maspéro, H. 1912. Phonétique historique de la langue Annamite: les initiales. *Bulletin de l'École Française d'Extrême-Orient* 12: 1–124.
- Maspéro, H. 1952. L'Annamite. In Meillet, A. and Cohen, M. *Les langues du monde*. 581–7. Paris: H. Champion.
- Matisoff, J. A. 1972. *The Loloish Tonal Split Revisited*. Vol. 7. Center for South and Southeast Asia Studies, University of California.
- Matisoff, J. A. 1973a. *The Grammar of Lahu*. University of California Press.
- Matisoff, J. A. 1973b. Tonogenesis in Southeast Asia. In Hyman, L. M. (ed.) *Southern California Occasional Papers in Linguistics*, No. 1. 72–95. Los Angeles: University of Southern California.

- Matisoff, J. A. 1986. Hearts and minds in South-East Asian languages and English: An essay in the comparative lexical semantics of psycho-collocations. *Cahiers de linguistique Asie orientale* 15 (1): 5–57. <https://doi.org/10.3406/clao.1986.1191>
- Matisoff, J. A. 1991a. Endangered languages of Mainland Southeast Asia. In Robins, R. J. and Uhlenbeck, E. M. (eds.) *Endangered Languages*. 189–228. London: Berg.
- Matisoff, J. A. 1991b. Areal and universal dimensions of grammaticalization in Lahu. In Traugott, E. C. and Heine, B. (eds.) *Approaches to Grammaticalization*. 383–453. Amsterdam: John Benjamins.
- Matisoff, J. A. 1991c. Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20 (1): 469–504. <https://doi.org/10.1146/annurev.an.20.100191.002345>
- Matisoff, J. A. 1994. Tone, intonation, and sound symbolism in Lahu: Loading the syllable canon. In Hinton, L., Nichols, J. and Ohala, J. J. (eds.) *Sound Symbolism*, 115–29. Cambridge: Cambridge University Press.
- Matisoff, J. A. 1999. Tibeto-Burman tonology in an areal context. In Kaji, S. (ed.) *Proceedings of the Symposium ‘Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics’*. 3–32. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Matisoff, J. A. 2000a. On ‘Sino-Bodic’ and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies* 63 (3): 356–69.
- Matisoff, J. A. 2000b. Tibeto-Burman tonology in an areal context. In *The Fifth International Symposium on Languages and Linguistics*. 85–151. Ho Chi Minh City: Vietnam National University, Ho Chi Minh City University of Social Sciences and Humanities.
<http://sealang.net/sala/archives/pdf8/matisoff2000tibeto-burman.pdf>
- Matisoff, J. A. 2001. Genetic vs. contact relationship: Prosodic diffusibility in South-East Asian languages. In Aikhenvald, A. Y. and Dixon, R. M. W. (eds.) *Areal Diffusion and Genetic Inheritance: Problems in comparative linguistics*. 291–327. Oxford: Oxford University Press.
- Matisoff, J. A. 2003a. *Handbook of Proto-Tibeto-Burman*. Berkeley, CA: University of California Press.
- Matisoff, J. A. 2003b. Aslian: Mon-Khmer of the Malay Peninsula. *Mon-Khmer Studies* 33: 1–58.
- Matisoff, J. A. 2003c. Lahu. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 208–21. Abingdon, Oxon and New York: Routledge.
- Matisoff, J. A. 2009. Stable roots in Sino-Tibetan/Tibeto-Burman. *Senri Ethnological Studies*, 75, 291–318.
- Matisoff, J. A. 2015a. *The Sino-Tibetan Etymological Dictionary and Thesaurus*. Berkeley, CA: The Regents of the University of California.
- Matisoff, J. A. 2015b. Re-examining the genetic position of Jingpho: Putting flesh on the bones of the Jingpho/Luish relationship. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 111–51. Berlin: De Gruyter Mouton.
- Matisoff, J. A. 2019. Preface. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. v–xvi. Berlin: De Gruyter Mouton.
- Matsumura, H., and Hudson, M. J. 2005. Dental perspectives on the population history of Southeast Asia. *American Journal of Physical Anthropology*. 127 (2): 182–209.
<https://doi.org/10.1002/ajpa.20067>
- Matsumura, H. and M. Oxenham, 2013. Population dispersal from East Asia into Southeast Asia: Evidence from cranial and dental morphology. In *Bioarchaeology of East Asia*, K. Pechenkina, M. Oxenham, eds. (Univ. Press of Florida), pp. 179–209.
- Matthews, S. and Yip V. 1994. *Cantonese: A comprehensive grammar*. London: Routledge.
- Mayuree, T. 2006. Gong: An endangered language of Thailand. Doctoral dissertation. Mahidol University, Bangkok.

- McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J. V., van Driem, G., Wilken, U. G., Seguin-Orlando, A., De la Fuente Castro, C., *et al.* 2018. The prehistoric peopling of Southeast Asia. *Science*, 361 (6397): 88–92.
- Meissner, M. and Philpott, S. B. 1975. The sign language of sawmill workers in British Columbia. *Sign Language Studies* 9: 291–308.
- Meyerhoff, A. A. *et al.* 1996. The tectonic evolution of Southeast Asia: A regional application of the surge-tectonics hypothesis. In Hull, D. M. (ed.) *Surge Tectonics: A new hypothesis of global geodynamics*. Solid Earth Sciences Library, Vol. 9. Dordrecht: Springer.
- Michaud, J. 2010. Editorial – Zomia and beyond. Special issue. *Journal of Global History* 5 (2): 187–214.
- Milne, L. 1921. *An Elementary Palaung Grammar*. Oxford: The Clarendon Press.
- Mithun, M. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.
- Mohr, S. 2015. Tshaukak’ui – hunting signs of the Ts’ixa in Northern Botswana. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 933–54. Berlin: De Gruyter Mouton.
- Moore, E. 2009. Place and space in early Burma: A new look at ‘Pyu Culture’. *Journal of the Siam Society*, 97, 101–28.
- Morey, S. 2005. *The Tai Languages of Assam: A grammar and texts*. Canberra: Pacific Linguistics.
- Morey, S. 2010. *Turung: A variety of Singpho language spoken in Assam*. Canberra: Australian National University.
- Morey, S. 2013. The marking of noun phrases: Some observations on the languages of North East India. In Hyslop, G., Morey, S. and Post, M. W. (eds.) 2013. *North East Indian Linguistics. Vol. 5*. 171–91. New Delhi: Cambridge University Press India.
- Morey, S. and Post, M. W. (eds.) 2008. *North East Indian Linguistics. Vol. 1*. New Delhi: Cambridge University Press India.
- Morey, S. and Post, M. W. (eds.) 2010. *North East Indian Linguistics. Vol. 2*. New Delhi: Cambridge University Press India.
- Mortensen, D. 2019. Hmong (Mong Leng). In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 609–52. Berlin: De Gruyter Mouton.
- Mottin, J. 1978. *Éléments de grammaire Hmong Blanc*. Bangkok: Don Bosco Press.
- Mukdawijitra, Y. 2019. Ethnic inclusive ICT policy: The case of digitising ethnic Tai scripts in Vietnam. Paper presented at the 10th EuroSEAS conference, Humboldt-Universität zu Berlin, September.
- Murphy, E. M. 1998. An osteological and palaeopathological study of the Scythian and Hunno-Sarmatian period populations from the cemetery complex of Aymyrlyg, Tuva, South Siberia. PhD. Queen’s University, Belfast.
- Murphy, S. A. 2013. Buddhism and its relationship to Dvaravati period settlement patterns and material culture in northeast Thailand and central Laos c. Sixth eleventh centuries ce: A historical ecology approach to the landscape of the Khorat Plateau. *Asian Perspectives* 52 (2): 300–26.
- Murphy, S. A. 2016. The case for proto-Dvāravatī: A review of the art historical and archaeological evidence. *Journal of Southeast Asian Studies*, 47 (3): 366–92. DOI:10.1017/s0022463416000242
- Nagaraja, K. S. 2011. Austroasiatic Languages: An introduction. In Nagaraja, K. S. (ed.) *Austro-Asiatic Linguistics: In memory of R. Elangaiyan*. 1–32. Mysore: CIIL (India).
- Namkung, J. 1996. Phonological inventories of Tibeto-Burman languages. *Sino-Tibetan Etymological Dictionary and Thesaurus Project*, Center for Southeast Asia Studies, University of California, Berkeley.
- Neng, Y. 1987. Hmoob Thaib Teb. *Haiv Hmoob* 3 (2): 36–7.
- Nespor, M. and Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.

- Nettle, D. 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
- Nguyen, A.-T., and Ingram, C. L. 2007. Stress and tone sandhi in Vietnamese reduplications. *The Mon-Khmer Studies* 37: 15–39.
- Nguyễn, H. H. and Nguyễn V. L. 1998. *Tiếng Katu* [The Katu language]. Hanoi: Nhà Xuất Bản Khoa Học Xã Hội.
- Nguyen, T. 2013. A grammar of Bih. PhD. University of Oregon.
- Nguyễn, T. T. H. and Boulakia, G. 1999. Another look at Vietnamese intonation. *Proceedings of the 14th International Congress of Phonetic Sciences*, 2399–402. Linguistics Department, University of California, Berkeley.
- Nguyễn, V. L. 1993. *Tiếng Ruc* (The Ruc language). Hanoi: Nhà Xuất Bản Khoa Học Xã Hội.
- Nguyễn, V. L., Hoành, N. H. and Thông, T. V. 2008. *Tiếng Mảng* [The Mang language]. Hanoi: Nhà Xuất Bản Khoa Học Hà Nội.
- Nguyễn, V. L., Phúc, Đ. V. and Thành, P. X. 1986. *Sách học tiếng Pakôh-Taôih* [Text for studying Pakoh-Taoih]. Hanoi: Ủy Ban Nhân Dân, Tỉnh Bình Trị Thiên.
- Nichols, J. 1992. *Linguistic Diversity in Space and Time*. Chicago: Chicago University Press.
- Niederer, B. 1998. *Les langues Hmong-Mjen (Miáo-Yáo): Phonologie historique*. Munich: Lincom Europa.
- Niederer, B. 2004. Pa-hng and the classification of the Hmong-Mien languages. In Tapp, N., Michaud, J., Culas, C. and Lee, G. Y. (eds.) *Hmong/Miao in Asia*. 129–46. Chiang Mai: Silkworm Books.
- Nonaka, A. M. 2007. Emergence of an indigenous sign language and a speech/sign community in Ban Khor, Thailand. PhD thesis. University of California, Los Angeles.
- Nonaka, A. M. 2009. Estimating size, scope, and membership of the speech/sign communities of undocumented indigenous/village sign languages: The Ban Khor case study. *Language and Communication* 29: 210–29.
- Nonaka, A. M. 2010. Interrogatives in Ban Khor sign language: A preliminary description. In Mathur, G. and Napoli, D. J. (eds.) *Deaf around the World: The impact of language*. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/9780199732548.003.0011
- Nonaka, A. M. 2011. Interrogatives in Ban Khor Sign Language: A Preliminary Description. In Mathur, G. and Napoli, D. J. (eds.) *Deaf around the World: The impact of language*. 194–220. Oxford: Oxford University Press.
- Nonaka, A. M. 2014. (Almost) everyone here spoke Ban Khor sign language – until they started using TSL: Language shift and endangerment of a Thai village sign language. *Language and Communication* 38: 54–72.
- Nonaka, A. M. 2015. Toponyms in Ban Khor sign language. *Learning Communities*. Special Issue: Indigenous Sign Languages.
- Norman, J. 1988. *Chinese*. Cambridge and New York: Cambridge University Press.
- Norquest, P. K. 2016. A phonological reconstruction of Proto-Hlai. In *The Languages of Asia Series*. Leiden and Boston: Brill.
- O'Connor, R. A. 1995. Agricultural change and ethnic succession in Southeast Asian states: A case for regional anthropology. *The Journal of Asian Studies* 54 (4): 968–96.
- O'Connor, R. A. 2003. Founders' cults in regional and historical perspective. In Tannenbaum, N. and Kammerer, C. A. (eds.) *Founders' Cults in Southeast Asia: Polity, and identity*. 269–311. New Haven: Yale University Press.
- Okell, J. 1969. *A Reference Grammar of Colloquial Burmese*. London: Oxford University Press.
- Ong, W. J. 1982. *Orality and Literacy: The technologizing of the word*. London: Methuen.
- Oota, H., Kurosaki, K., Pookajorn, S., Ishida, T. and Ueda, S. 2001. Genetic study of the paleolithic and neolithic Southeast Asians. *Human Biology* 73(2): 225–31.

- Oppenheimer, S. 2011. MtDNA variation and southward Holocene human dispersals within Mainland Southeast Asia. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 81–108. Canberra: Pacific Linguistics.
- Orlandi, Giorgio. 2018. Evaluating the Sino-Tibeto-Austronesian hypothesis. *Journal of Language Relationship* 16 (1): 1–18.
- Ostapirat, W. 2000. Proto-Kra. *Linguistics of the Tibeto-Burman Area*, 23 (1): 1–251.
- Ostapirat, W. 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary distribution. In Sagart, L., Blench, R. and Sanchez-Mazas, A. (eds.) *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*. 107–31. London: RoutledgeCurzon.
- Ostapirat, W. 2008. The Hlai language. In Diller, A., Edmondson, J. A. and Luo, Y. (eds.) *The Tai-Kadai Languages*. 623–52. Abingdon, Oxon and New York: Routledge.
- Ostapirat, W. 2013. Austro-Tai revisited. 23rd Annual Meeting of the Southeast Asian Linguistic Society, Chulalongkorn University.
- Ostapirat, W. 2018. Macrophylectic trees of East Asian languages re-examined. In Kikusawa, R. and Reid, L. A. (eds.) *Let's Talk about Trees*. Osaka: Semri Ethnological Studies, Minpaku.
- Ouyang, J. and Zheng, Y. 1983. The Huihui speech (Hainan Cham) of Hui nationality in Ya County, Hainan. *Minzu Yuwen* 1: 30–40.
- Owen, R. W. 2012. A tonal analysis of contemporary Tai Khuen varieties. *Journal of the Southeast Asian Linguistics Society (JSEALS)* 5: 12–31.
- Oxenham, M. and Tayles, N. 2006. *Bioarchaeology of Southeast Asia*. New York: Cambridge University Press.
- Parkin, R. 1991. *A Guide to Austroasiatic Speakers and their Languages*. Honolulu: University of Hawai'i Press.
- Payne, T. E. 2017. Morphological typology. In Aikhenvald, A. and Dixon R. M. W. (eds.) *The Cambridge Handbook of Linguistic Typology*. 39–77. Cambridge: Cambridge University Press.
- Pe, H. 1965. A re-examination of Burmese ‘classifiers’. *Lingua* 15: 163–85.
- Peiros, Ilia. 1998. *Comparative Linguistics in Southeast Asia*. Canberra: Pacific Linguistics.
- Perkins, R. D. 1992. *Deixis, Grammar, and Culture*. Amsterdam: John Benjamins.
- Peterson, D. A. 2003. Hakha Lai. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 409–26. Abingdon, Oxon and New York: Routledge.
- Peterson, D. A. 2019. Bangladesh Khumi. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 12–55. Berlin: De Gruyter Mouton.
- Pfau, R., Steinbach, M., and Woll, B. (eds.). 2012. *Sign Language: An international handbook*. Berlin: De Gruyter Mouton.
- Pham, A. 2008. Is there a prosodic word in Vietnamese? *Toronto Working Papers in Linguistics* 29. <https://bit.ly/31b6HVu>
- Phattharathanit, S. 2012. Identity maintenance in Lanna (Northern Thai). *Journal of the Southeast Asian Linguistics Society* 5: 67–84.
- Phillips, R. 1973. A Mnong pedagogical grammar: The verb phrase and constructions with two or more verbs. *Mon-Khmer Studies* 4: 129–38.
- Pietruszewsky, M. 2005. The physical anthropology of the Pacific, East Asia and Southeast Asia: a multivariate craniometric analysis. In Sagart, L., Blench, R. and Sanchez-Mazas, A. (eds.) *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*. 201–29. London: RoutledgeCurzon.
- Pinnow, H.-J. 1954. Grundzüge einer Phonetik der Kharia Sprache. Unpublished Manuscript, Berlin.
- Pinnow, H.-J. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.
- Pinnow, H.-J. 1960. Über der Ursprung der voneinander abweichenden strukturen der Munda- und Khmer-Nikobar-Sprachen. *Indo-Iranian Journal* 4: 81–103.

- Pinnow, H.-J. 1963. The position of the Munda languages within the Austroasiatic language family. In Shorto, H. L. (ed.) *Linguistic Comparison in Southeast Asia and the Pacific*. 140–52. London: School of Oriental and African Studies.
- Pittayaporn, P. 2005. Moken as a mainland Southeast Asian language. In Grant, A. and Sidwell, P. (eds.) *Chamic and Beyond: Studies in mainland Austronesian languages*. 189–210. Canberra: Australian National University. DOI:10.15144/PL-569.189
- Pittayaporn, P. 2007. Prosody of final particles in Thai: interaction between lexical tones and boundary tones. Paper presented to the International Workshop on ‘Intonation Phonology: Understudied or Fieldwork Languages’, Saarbrücken.
- Pittayaporn, P. 2009. Proto-Southwestern-Thai: A new reconstruction. *Journal of the Southeast Asian Linguistics Society* 2: 119–43.
- Pittayaporn, P. 2015. Typologizing sesquisyllabicity: The role of structural analysis in the study of linguistic diversity in Mainland Southeast Asia. In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 500–28. Berlin: De Gruyter Mouton.
- Ploykaew, P. 2001. Samre grammar. Doctoral dissertation. Mahidol University, Bangkok.
- Post, M. W. 2007. A grammar of Galo. Dissertation. La Trobe University, Melbourne.
- Post, M. W. 2008. Adjectives in Thai: Implications for a functionalist typology of word classes. *Linguistic Typology* 12 (3): 339–81.
- Post, M. W. 2009. The phonology and grammar of Galo ‘words’: A case study in benign disunity. *Studies in Language. International Journal Sponsored by the Foundation ‘Foundations of Language’* 33 (4): 934–74.
- Post, M. W. 2011. Prosody and typological drift in Austroasiatic and Tibeto-Burman: Against ‘Indosphere’ and ‘Sinosphere’. In Srichampa, S., Sidwell, P., and Gregerson, K. J. (eds.) *Austroasiatic Studies: Papers from International Conference on Austroasiatic Linguistics (ICAAL4)*. Mon-Khmer Studies Special Issue 3. 198–211. Canberra: Pacific Linguistics.
- Post, M. W. 2015. Morphosyntactic reconstruction in an areal-historical context: A pre-historical relationship between North East India and Mainland Southeast Asia? In Enfield, N. J. and Comrie, B. (eds.) *The Languages of Mainland Southeast Asia: The state of the art*. 209–65. Berlin: De Gruyter Mouton.
- Preecha S. 1993. Correlates of the register complex in Kuay. *Mon-Khmer Studies*, 22: 245–51.
- Premsrirat, S. 1987. *Khmu, a Minority Language of Thailand*. Canberra: Australian National University.
- Premsrirat, S. 1998a. So (Thavung)-English-Thai glossary Part I. *Mon-Khmer Studies* 28: 189–218.
- Premsrirat, S. 1998b. Language maintenance and language shift in minority languages of Thailand. In Matsumura, K. (ed.) *Studies in Endangered Languages*. 191–211. Tokyo: Hituzi Syobo.
- Premsrirat, S. 1999. So (Thavung)-English-Thai glossary Part II. *Mon-Khmer Studies* 29: 107–32.
- Premsrirat, S. 2007. Endangered languages of Thailand. *International Journal of the Sociology of Language* 186: 75–93.
- Premsrirat, S. 2008. Orthography development: A tool for revitalizing and maintaining ethnic minority languages. *Journal of Language and Culture* 26: 18–34.
- Premsrirat, S. 2011. Pearic, a dying branch of Austroasiatic languages and its struggle for survival. In Srichampa, S. and Sidwell, P. (eds.) *Austroasiatic studies: Papers from International Conference on Austroasiatic Linguistics (ICAAL4)*. Mon-Khmer Studies Special Issue 2. 138–53. Canberra: Pacific Linguistics.
- Premsrirat, S. and Rojanakul, Nattamon. 2015. Chong. In Jenny, M. and Sidwell, P. (eds.) *The Handbook of Austroasiatic Languages*. 603–40. Leiden: Brill.

- Provencher, R. 1975. *Mainland Southeast Asia: An anthropological perspective*. Pacific Palisades, CA: Goodyear Pub. Co.
- Pulleyblank, E. G. 1970. Late Middle Chinese, Part I, *Asia Major* 15: 197–239.
- Pulleyblank, E. G. 1971. Late Middle Chinese, Part II, *Asia Major* 16: 121–66.
- Pulleyblank, E. G. 1984. *Middle Chinese: A study in historical phonology*. Vancouver: University of British Columbia Press.
- Pulu, J. 1978. *Idu Phrase Book*. Shillong: Directorate of Information and Public Relations, Arunachal Pradesh.
- Purnell, H. C. 1970. Toward a reconstruction of Proto-Miao-Yao. PhD dissertation, Cornell University.
- Purnell, H. C. 2012. *An Iu-Mienh English Dictionary with Cultural Notes*. Chiang Mai: Silkworm Books.
- Purnell, H. C., Zanh, G.-F., Burgess, V. A., and Aumann, G. 2012. *An Iu-Mienh-English Dictionary: With cultural notes*. Chiang Mai: Silkworm Books, Center for Lao Studies.
- Quay, S. 2015. Monastic sign language from medieval to modern times. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 871–900. Berlin: De Gruyter Mouton.
- Ramsey, S. R. 1987. *The Languages of China*. Princeton, NJ: Princeton University Press.
- Rastorfer, J.-M. 1994. *On the Development of Kayah and Kayan National Identity: A Study and Bibliography*. Bangkok: Southeast Asian Publishing House.
- Ratanakul, S. 1986. *Thai-Sgaw Karen Dictionary*. 2 vols. Bangkok: Institute of Language and Culture for Rural Development.
- Ratanakul, S. 1998. Numeral classifiers in Sgaw Karen. *Mon-Khmer Studies* 28: 101–13.
- Ratanakul, S. 2001. Adverbs in Sgaw Karen. *Mon-Khmer Studies* 31: 127–34.
- Ratcliff, M. 1987. Tone sandhi compounding in White Hmong. *Linguistics of the Tibeto-Burman Area* 10 (2): 71–105.
- Ratcliff, M. 1990. The influence of geographical change on grammar: The case of Hmong spatial deictics. Paper presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, University of Texas at Arlington, 3–7 October.
- Ratcliff, M. 1992. *Meaningful tone: A study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong*. DeKalb, IL: Northern Illinois University, Center for Southeast Asian Studies.
- Ratcliff, M. 1998. Ho Ne (She) is Hmongic: One final argument. *Linguistics of the Tibeto-Burman Area* 21 (2): 97–109.
- Ratcliff, M. 2010. Hmong-Mien Language History. *Studies in Language Change* 8. Canberra: Pacific Linguistics.
- Ratcliff, M. 2015. Tonogenesis, and tone change. In *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press.
- Reardon, W., Toriello, H., Downs, C., 2004. Epidemiology, etiology, genetic patterns, and genetic counseling. In Toriello, H., Reardon, W., Gorlin, R. (eds.) 2004. *Heredity Hearing Loss and Its Syndromes*. 2nd edition. 8–16 Oxford: Oxford University Press.
- Reid, L. A. 2005. The current status of Austroasiatic. In Sagart, L., Blench, R. and Sanchez-Mazas, A. (eds.) *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*. 132–60. London: RoutledgeCurzon.
- Riddle, E. M. 1993. The relative marker uas in Hmong. *Linguistics of the Tibeto-Burman area* 16 (2): 57–68.
- Ring, H. and Anderson, G. D. S. 2018. On prosodic structures in Austroasiatic diachrony: ‘Rhythmic Holism’ revisited in light of preliminary acoustic studies. In Ring, H. and Rau, F. (eds.) *Papers from the Seventh International Conference on Austroasiatic Linguistics*. 1–35. Honolulu: University of Hawai‘i at Mānoa.

- Rischel, J. 1995. *Minor Mlabri: A hunter-gatherer language of northern Indochina*. Copenhagen: Museum Tusculanum Press.
- Rousset, I. 2004. Structures syllabiques et lexicales des langues du monde. PhD dissertation. University of Grenoble.
- Ruey, Y.-F. and Kuan, T.-K. 1962. *Chuannan Yaqiao Miao de hun sang lisu* [Marriage and mortuary customs of the Magpie Miao, Southern Szechuan, China]. Academia Sinica, the Institute of History and Philology Monographs Series A, no. 23. Taipei: Academis Sinica.
- Sadock, J. M. and Zwicky, A. M. 1985. Speech acts distinctions in syntax. In Shopen, T. (ed.) *Language Typology and Syntactic Description*. 155–96. Cambridge: Cambridge University Press.
- Sagart, L. 1990. Chinese and Austronesian are genetically related. Paper presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, University of Texas at Arlington, 3–7 October.
- Sagart, L. 1993. Chinese and Austronesian: Evidence for a genetic relationship. *Journal of Chinese Linguistics* 21 (1): 1–63.
- Sagart, L. 1994. Old Chinese and Proto-Austronesian. *Oceanic Linguistics* 33: 271–308.
- Sagart, L. 1999. *The Roots of Old Chinese*. Amsterdam: John Benjamins.
- Sagart, L. 2011. The Austroasiatics: East to west or west to east? In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 345–59. Canberra: Pacific Linguistics.
- Sagart, L., Blench, R. and Sanchez-Mazas, A. (eds.) 2005. *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*. London: RoutledgeCurzon.
- Sagart, L., Jacques, G., Lai, Y., Ryder, R. J., Thouzeau, V., Greenhill, S. J., and List, J.-M. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences*. 116 (21): 10317–22. <https://doi.org/10.1073/pnas.1817972116>
- Samarin, W. J. 1952. Intonation in tone languages. *African Studies* 11 (2): 80–2. DOI:10.1080/00020185208706870
- Samarin, W. J. 1970. Inventory and choice in expressive language. *Word* 26: 153–69.
- Samarina, I. 2011. *Jazyki gelao: Materialy k sopostavitel'nomu slovarju kadajskich jazykov* [Gael Languages: Materials for a Comparative Dictionary of Kadai Languages]. Moscow: Academia.
- Sapir, E. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace.
- Sarawit, M. E. 1973. The Proto-Tai vowel system. PhD dissertation. University of Michigan, Ann Arbor.
- Saul, J. E. 1964. Classifiers in Nung. *Lingua* 13: 278–90.
- Schiering, R., Bickel, B. and Hildebrandt, K. A. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics*, 46 (3): 657–709. <https://doi.org/10.1017/S002226710000216>
- Schmidt, W. 1906. *Die Mon-Khmer Völker: Ein bindeglied zwischen Völkern Zentralasiens und Austronesiens*. Braunschweig: Friedrich Vieweg und Sohn.
- Scott, J. C. 2009. *The Art of Not Being Governed: An anarchist history of upland Southeast Asia*. New Haven, CT: Yale University Press.
- Searle, J. R. 2010. Making the social world: The structure of human civilization. Oxford: Oxford University Press.
- Sebeok, T. A. 1942. An examination of the Austroasiatic language family. *Language*, 18 (3): 206–17.
- Seiler, H. 1983. *Possession as an Operational Dimension of Language*. Tübingen: Gunter Narr Verlag.
- Seng Mai, M. 2012. A descriptive grammar of Wa. MA thesis. Payap University, Chiang Mai.
- Shackelford, L., Demeter, F., Westaway, K., Duringer, P., Ponche, J.-L., Sayavongkhamdy, T., Zhao, J.-X., Barnes, L., Boyon, M., Sichanthongtip, P., Sénégas, F., Patole-Edoumba, E., Coppens, Y., Dumoncel, J., and Bacon, A.-M. 2018. Additional evidence for early modern human morphological diversity in Southeast Asia at Tam Pa Ling, Laos. *Quaternary International* 466 (Part A): 93–106. <https://doi.org/10.1016/j.quaint.2016.12.002>
- Shafer, R. 1966. *Introduction to Sino-Tibetan*. Part I. Wiesbaden: Harrassowitz.

- Shafer, R. 1974. *Introduction to Sino-Tibetan*. Wiesbaden: Harrassowitz.
- Sharifian, F., Dirven, R., Yu, N. and Niemeier, S. 2008. Culture, body, and language conceptualizations of internal body organs across cultures and languages. Berlin: De Gruyter Mouton.
- Sharp, L. 1962. Cultural continuities and discontinuities in Southeast Asia. *The Journal of Asian Studies* 22 (1): 3–11.
- Shee, N. H. 2008. A descriptive grammar of Geba Karen. MA thesis. Payap University, Chiang Mai.
- Shen, X. 1989. Interplay of the four citation tones and intonation in Mandarin Chinese. *Journal of Chinese Linguistics* 17 (1): 61–74.
- Shintani, T. 2008. *The Palaung Language: The comparative lexicon of its southern dialects*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCCA).
- Shorto, H. L. 1960. Word and syllable pattern in Palaung. *Bulletin of the School of Oriental and African Studies* 23 (3): 544–57.
- Shorto, H. L. 2006. *A Mon-Khmer Comparative Dictionary*. Edited by P. Sidwell, D. Cooper and C. Bauer. Canberra: Pacific Linguistics.
- Sidnell, J. and Shohet, M. 2013. The problem of peers in Vietnamese interaction. *Journal of the Royal Anthropological Institute* 19 (3): 618–38.
- Sidwell, P. 2006. Dating the separation of Acehnese and Chamic by etymological analysis of the Aceh-Chamic lexicon. *Mon-Khmer Studies*, 36, 187–206.
- Sidwell, P. 2008. The Khom script of the Kommodam Rebellion. *International Journal of the Sociology of Language* 192: 15–25. DOI 10.1515/IJSL.2008.032
- Sidwell, P. 2009a. *Classifying the Austroasiatic languages: History and state of the art*. Munich: Lincom Europa.
- Sidwell, P. 2009b. Proto-Mon-Khmer vocalism: Moving on from Shorto's "Alternances." *Journal of the Southeast Asian Linguistics Society* 1. 205–14.
- Sidwell, P. 2015. Austroasiatic classification. In Jenny, M. and Sidwell, P. (eds.) *The Handbook of Austroasiatic Languages*. 144–220. Leiden: Brill.
- Sidwell, P. and Blench, R. 2011. The Austroasiatic urheimat: The southeastern riverine hypothesis. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 315–43. Canberra: Pacific Linguistics .
- Sidwell, P. and Rau, F. 2015. Austroasiatic comparative-historical reconstruction: An overview. In Jenny, M. and Sidwell, P. (eds.) *The Handbook of Austroasiatic Languages*. 221–363. Leiden: Brill.
- Simpson, A. (ed.) 2007. *Language and National Identity in Asia*. Oxford: Oxford University Press.
- Simpson, A. and Thammasathien, N. 2007. Thailand and Laos. In Simpson, A. (ed.) *Language and National Identity in Asia*. 391–414. Oxford: Oxford University Press.
- Smalley, W. A., Vang, C. K. and Yang, G. Y. 1990. *Mother of Writing: The origin and development of a Hmong messianic script*. Chicago: University of Chicago Press.
- Smith, K. D. 1979. *Sedang Grammar*. Canberra: Pacific Linguistics.
- Smith, K. D. 1981. A lexico-statistical study of 45 Mon-Khmer languages. In Gonzalez, A. B. and Thomas, D. D. (eds.), *Linguistics Across Continents*. 180–205. Manila: SIL International.
- Snyder, W. C. and Lu, T. 1997. Wuming Zhuang tone sandhi: A phonological, syntactic, and lexical investigation. In Edmondson, J. A. and Solnit, D. B. (eds.) *Comparative Kadai: The Tai branch*. 107–37. Dallas: SIL International.
- So-Hartmann, H. 2009. *A Descriptive Grammar of Daai Chin*. STEDT Monograph 7. Berkeley, CA: University of California, Berkeley.
- Söderberg, S. 2006. The semantic domains of numeral classifiers in Kammu. Thesis. Lund University.
- Solnit, D. B. 1985. Introduction to the Biao Min Yao language. *Cahiers de Linguistique Asie Orientale* 14 (2): 175–91.

- Solnit, D. B. 1988. The position of Lakkia within Tai. In Edmondson, J. A. and Solnit, D. B. (eds.) *Comparative Kadai: Linguistic studies beyond Tai*. 219–38. Dallas: SIL International.
- Solnit, D. B. 1997. *Eastern Kayah Li: Grammar, texts, glossary*. Honolulu: University of Hawai‘i Press.
- Song, J. J. (ed.) 2011. *The Oxford Handbook of Linguistic Typology*. New York: Oxford University Press.
- Sposato, A. 2012. Relative clauses in Xong (Miao-Yao). *Journal of the Southeast Asian Linguistics Society* 5: 49–66.
- Sposato, A. 2015. A grammar of Xong. PhD. State University of New York.
- Sproat, R. 2005. Language use and policy in a linguistically fragmented refugee community. Master’s dissertation. Macquarie University.
- Srichampa, S. 1997. Serial verb constructions in Vietnamese. *Mon-Khmer Studies* 27: 137–44.
- Srichampa, S. 2008. Patterns of polite expressions in Vietnamese. *Mon-Khmer Studies* 38: 117–47.
- Srisakorn, P. 2008. So (Thavung) grammar. Dissertation. Mahidol University, Bangkok.
- Stargardt, J. 2000. *Tracing Thoughts through Things: The oldest Pali texts and the early Buddhist archaeology of India and Burma*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Starosta, S. 2005. Proto-East Asian and the origin and dispersal of the languages of East and Southeast Asia and the Pacific. In Sagart, L., Blench, R. and Sanchez-Mazas, A. (eds.) *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*. 182–97. London: RoutledgeCurzon.
- Stocking, G. W. 1992. *The Ethnographer’s Magic and Other Essays in the History of Anthropology*. Madison, WI: University of Wisconsin Press.
- Stockwell, A. J. 1992. Southeast Asia in War and Peace: The end of European colonial empires. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia*. Vol. 2.325–86. Cambridge: Cambridge University Press.
- Stojanowski, C. M. and Schillaci, M. A. 2006. Phenotypic approaches for understanding patterns of intracemetery biological variation. *Yearbook of Physical Anthropology*, 131 (S43): 49–88.
- Stokoe, William. C. 1960. *Sign Language Structure: An outline of the visual communication systems of the American Deaf*. Studies in Linguistics: Occasional Papers, 8. Buffalo, NY: University of Buffalo Department of Anthropology and Linguistics.
- Stoneking, M. and Delfin, F. 2010. The human genetic history of East Asia: Weaving a complex tapestry. *Current Biology* 20 (4): R188–R193.
- Sun, H. 1991. *Zangmianyu yuyin he cihui [Tibeto-Burman phonology and lexicon]*. Beijing: Chinese Social Sciences Press.
- Sun, H. and Liu, G. 2009. *A Grammar of Anong: Language death under intense contact*. Translated and edited by Thurgood, G., Li, F. and Thurgood, E. Leiden: Brill.
- Swanarat, M. and Reilly, C. (eds.) 1986. *The Thai Sign Language Dictionary*. Vol. 1. Bangkok: National Association of the Deaf. www.language-archives.org/item/oai:refdb.wals.info:5707
- Svantesson, J.-O. 1989. Tonogenetic mechanisms in Northern Mon-Khmer. *Phonetica* 46: 60–79.
- Svantesson, J.-O. and Holmer, A. 2015. Kammu. In Jenny, M. and Sidwell, P. (eds.) *The Handbook of Austroasiatic Languages*. 955–1002. Leiden: Brill.
- Svantesson, J.-O., Kam, R., Lindell, K. and Lundstrom, H. 2013. *Dictionary of Kammu Yuan Language and Culture*. Honolulu: University of Hawai‘i Press.
- Swadesh, M. 1955. Chemakum lexicon compared with Quileute. *International Journal of American Linguistics* 21 (1): 60–72. <https://doi.org/10.1086/464309>
- Swastham, P. 1982. A description of Moklen: A Malayo-Polynesian language. MA thesis. Mahidol University, Bangkok.

- Taguchi, Y. 2013. On the phylogeny of Hmongic languages. Oral presentation presented at the 23rd Annual Meeting of the Southeast Asian Linguistics Society, Bangkok. www.jseals.org/seals23/taguchi2013phylogenyp.pdf
- Tang, G. 2007. *Hong Kong Sign Language: A trilingual dictionary with linguistic descriptions.* Hong Kong: Chinese University Press.
- Tang, G. 2015. Hong Kong sign language. In Wang, W. S-Y. and Sun, C. (eds.) *The Oxford Handbook of Chinese Linguistics.* 710–28. Oxford: Oxford University Press.
- Tang, C. 2018. Dialects of Chinese. In Boberg, C., Nerbonne, J. A. and Watt, D. J. L. (eds.) *The Handbook of Dialectology.* 547–58. Hoboken, NJ: Wiley-Blackwell.
- Tarling, N. (ed.) 1992a. *The Cambridge History of Southeast Asia.* Vol. 1. Cambridge: Cambridge University Press.
- Tarling, N. (ed.) 1992b. *The Cambridge history of Southeast Asia.* Vol. 2. Cambridge: Cambridge University Press.
- Tarling, N. 2004. *Nationalism in Southeast Asia: If the people are with us.* Abingdon, Oxon and New York: RoutledgeCurzon.
- Taylor, K. 1992. The early kingdoms. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia.* Vol. 1. 137–182. Cambridge: Cambridge University Press.
- Thangaraj, K., Chaubey, G., Reddy, A. G., Singh, V. K., Singh, L. 2006. Unique origin of Andaman Islanders: insight from autosomal loci. *Journal of Human Genetics* 51: 800–4. DOI:10.1007/s10038-006-0026-0. PubMed: 16924390
- Thoij, C. 1981. *Kev tsim neej tshiab hauv Asmeslivkas.* Washington, D.C: Center for Applied Linguistics.
- Thomas, D. D. 1962. On defining the ‘word’ in Vietnamese. *Van-Hóa Nguyêt-San* 11 (5): 519–23.
- Thomas, D. D. 1966. Mon-Khmer subgroupings in Vietnam. In Zide, N. H. (ed.) *Studies in Comparative Austroasiatic Linguistics.* 194–202. The Hague: Mouton.
- Thomas, D. D. 1967. Chrau grammar: A Mon-Khmer language of Vietnam. PhD. University of Pennsylvania.
- Thomas, D. D. 1971. *Chrau Grammar.* Honolulu: University of Hawai‘i Press.
- Thomas, D. D. 1992. On sesquisyllabic structure. *Mon-Khmer Studies* 21: 207–10.
- Thomas, D. D. and Headley, R. K. 1970. More on Mon-Khmer subgroupings. *Lingua* 25: 398–418. [https://doi.org/10.1016/0024-3841\(70\)90048-3](https://doi.org/10.1016/0024-3841(70)90048-3)
- Thomas, G. 2008. An analysis of the Xiamen tone circle. In Abner, N. and Bishop, J. (eds.) *Proceedings of the 27th West Coast Conference on Formal Linguistics.* 422–30. Somerville, MA: Cascadilla Proceedings Project.
- Thompson, L. 1965. *A Vietnamese Grammar.* Seattle: University of Washington Press.
- Thompson, L. 1987. *A Vietnamese Reference Grammar.* Mon Khmer Studies 13–14. Honolulu: University of Hawai‘i Press.
- Thongkum, T. L. 1988. *Phonation types in Mon-Khmer languages.* In Fujimura, O. (ed.) *Vocal Physiology: Voice production, mechanisms and functions.* 319–33. New York: Raven Press.
- Thongkum, T. L. 1989. An acoustic study of the register complex in Kui (Suai). *Mon-Khmer Studies* 15: 1–19.
- Thongkum, T. L. 1992. A preliminary reconstruction of Proto-Lakkja (Chan Shan Yao). *Mon-Khmer Studies* 20: 57–89.
- Thongkum, T. L. 1993. A view on Proto-Mjuenic (Yao). *Mon-Khmer Studies* 22: 163–230.
- Thurgood, G. 1974. Lolo-Burmese rhymes. *Linguistics of the Tibeto-Burman Area*, 1 (1): 98–107.
- Thurgood, G. 1977. Lisu and Proto-Lolo-Burmese. *Acta Orientalia* 38: 147–207.
- Thurgood, G. 1982. Subgrouping on the basis of shared phonological innovations: A Lolo-Burmese case study. In Macaulay, M., Gensler, O. D., Brugman, C., Čivkulis, I., Dahlstrom, A., Krile, K. and Sturm, R. (eds.) *Proceedings of the 8th Annual Meeting of the Berkeley Linguistics Society.* 251–60.

- Thurgood, G. 1998. The development of the Chamic vowel system: The interaction of inheritance and borrowing. In Thomas, D. (ed.) *Papers in Southeast Asian Linguistics No. 15: Further Chamic Studies*, 61–90. Canberra: Pacific Linguistics.
- Thurgood, G. 1999. *From Ancient Cham to Modern Dialects: Two thousand years of language contact and change*. Oceanic Linguistics Special Publication vol. 28. Honolulu: University of Hawai'i Press.
- Thurgood, G. 2002. Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, 19 (2): 333–63.
- Thurgood, G. 2003a. A subgrouping of the Sino-Tibetan languages: The interaction between language contact, chance, and inheritance. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 3: 21. Abingdon, Oxon and New York: Routledge.
- Thurgood, G. 2003b. Hrê contact and the origins of the Haroi restructured register system: A case of shared sound changes. In Adams, K. L. et al. (eds.) *Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society, Tempe, Arizona*. 199–207. Tempe, AZ: Arizona State University.
- Thurgood, G. 2005. Phan Rang Cham. In Adelaar, A. and Himmelmann, N. P. (eds.) *The Austronesian Languages of Asia and Madagascar*. 489–512. Abingdon, Oxon and New York: Routledge.
- Thurgood, G. 2010. Hainan Cham, Anong, and Eastern Cham: Three languages, three social contexts, three patterns of change. *Journal of Language Contact VARIA* 3: 39–61.
- Thurgood, G. 2017. Sino-Tibetan: Genetic and areal subgroups. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 2nd ed. 3: 39. Abingdon, Oxon and New York: Routledge.
- Thurgood, G., Edmondson, J. A. and Solnit, D. B. 1988. Notes on the reconstruction of Proto-Kam-Sui. In Edmondson, J. A. and Solnit, D. B. (eds.) *Comparative Kadai: Linguistic studies beyond Tai*. 179–218. Dallas: SIL International.
- Thurgood, G., and LaPolla, R. J. (eds.) 2017. *The Sino-Tibetan Languages*. (2nd ed.). Abingdon, Oxon and New York: Routledge.
- Thurgood, G., Thurgood, E. and Fengxiang, L. 2014. *A Grammatical Sketch of Hainan Cham: History, contact, and phonology*. Berlin: De Gruyter Mouton.
- Tooley, M. J. and Shennan, I. (eds.) 1987. *Sea Level Changes*. Oxford: Blackwell.
- Trần, H. M. 1967. *Tones and Intonation in South Vietnamese*. Papers in Southeast Asian Linguistics No. 1, edited by Đ. L. Nguyễn, H. M. Trần and D. Dellinger. Canberra: Linguistics Circle of Canberra.
- Tun, Win Naing. 2015. Prehistory to proto-history of Myanmar: A perspective of historical geography. Paper presented at the International Conference on Burma/Myanmar Studies Burma/Myanmar in Transition: Connectivity, Changes and Challenges University Academic Service Centre (UNISERV), Chiang Mai University, Thailand, 24–25 July.
- Turnbull, C. M. 1992. Regionalism and nationalism. In Tarling, N. (ed.) *The Cambridge History of Southeast Asia*. Vol. 2. 585–646. Cambridge: Cambridge University Press.
- Umaporn, S. 2007. Backchannel response in Mon conversation. *Mon-Khmer Studies* 37: 67–85.
- van Breugel, S. 2014. *A Grammar of Atong*. Leiden: Brill.
- van Driem, G. 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies*, 60 (3): 455–88.
- van Driem, G. 2001. *Languages of the Himalayas: An ethnolinguistic handbook of the Greater Himalayan region, containing an introduction to the symbiotic theory of language*. Vols. 1–2. Leiden: Brill.
- van Driem, G. 2002. The father tongue hypothesis: Sexually dimorphic dispersals in prehistory. Paper presented at the 17th Conference of the Indo-Pacific Prehistory Association, Taipei.

- van Driem, G. 2005. Sino-Austronesian vs. Sino-Causasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In Yadava, Y., Bhattacharai, G., Lohani, R. R., Prasain, B. and Parajuli, K. (eds.) *Contemporary Issues in Nepalese Linguistics*.285 338. Kathmandu: Linguistic Society of Nepal.
- van Driem, G. 2007. Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies. *Mon-Khmer Studies* 37: 1–14.
- van Driem, G. 2011. Rice and the Austroasiatic and Hmong-Mien homelands. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of mainland Southeast Asia*. 361–90. Canberra: Pacific Linguistics.
- van Driem, G. 2013. Trans-Himalayan. In Hill, N. and Owen-Smith, T. (eds.) *Trans-Himalayan Linguistics*. Berlin: Mouton De Gruyter. <https://doi.org/10.1515/9783110310832.11>
- van Liere, W. J. 1980. Traditional water management in the Lower Mekong Basin. *World Archaeology*, 11 (3): 265–80.
- van Schendel, W. 2002. Geographies of knowing, geographies of ignorance: Jumping scale in Southeast Asia. *Environment and Planning D: Society and Space*, 20 (6): 647–68. <https://doi.org/10.1068/d16s>
- VanBik, K. 1998. Lai psycho-collocations. *Linguistics of the Tibeto-Burman Area*, 21 (1): 201–33.
- Vittrant, A. 2007. Serial verb constructions in Burmese. Paper presented at the Fédération de Typologie et Universaux Linguistiques Trajectoire Project meeting, Lyon, 9 November.
- Vittrant, A. 2019. Burmese. In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 56–130. Berlin: De Gruyter Mouton.
- Vittrant, A. and Watkins, J. (eds.) 2019. *Mainland South East Asia Linguistic Area*. Berlin: De Gruyter Mouton.
- Voris, H. K. 2000. Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography* 27 (5): 1153–67.
- Vossen, F. 2011. The Jespersen Cycle in South-East Asia, Oceania and Australia. Studies van de BKL Travaux du CBL, 6. <https://bit.ly/3dLYVaA>
- Wang, F. 1979. Miáoyǔ fāngyán shēngyùnmǔ bǐjiào [The comparison of initials and finals of Miao dialects]. Beijing: Nationalities Research Institute of the CASS.
- Wang, F. 1994. Reconstruction of the Proto-Miao language. *The Research Institute for Languages and Cultures of Asia and Africa*. Tokyo University of Foreign Studies.
- Wang, F. and Mao, Z. 1995. Miáoyáoyǔ gǔyīn gòuni [Reconstruction of Proto-Hmong-Mien phonology]. Beijing: China Social Sciences Press.
- Wang, J. and Zheng, G. 1980. *Mulauyü chienchi*.
- Wang, J. and Zheng, G. 1993. *An Outline Grammar of Mulao*. Translated by Y. Luo. Canberra: Australian National University.
- Wangkumhang P., Shaw P. J., Chaichoompu K., Ngamphiw C., Assawamakin A., et al. 2013. Insight into the peopling of Mainland Southeast Asia from Thai population genetic structure. *PLoS ONE* 8 (11): e79522. DOI:10.1371/journal.pone.0079522
- Watkins, J. 2002. *The Phonetics of Wa: Experimental phonetics, phonology, orthography and sociolinguistics*. Canberra: Pacific Linguistics.
- Watkins, J. 2007. Burma/Myanmar. In Simpson, A. (ed.) *Language and National Identity in Asia*. 263–87. Oxford: Oxford University Press.
- Watkins, J. 2010. Sign language in Burma: Two Burmese finger spelling systems used in Yangon. Paper presented at Burma Studies Conference 2010: Burma in the Era of Globalization, 6–9 July, Université de Provence, Marseille. <https://eprints.soas.ac.uk/10275>
- Watkins, J. 2013a. “A first account of tone in Myebon Sumtu Chin.” *Linguistics of the Tibeto-Burman Area* 36 (2): 97–127.
- Watkins, J. 2013b. *Dictionary of Wa*. 2 vols. Leiden: Brill.
- Watkins, J. 2019. Wa (Paraok). In Vittrant, A. and Watkins, J. (eds.) *The Mainland Southeast Asia Linguistic Area*. 432–74. Berlin: De Gruyter Mouton.

- Watson, R. 2011. A case for clitics in Pacoh. In Srichampa, S., Sidwell, P. and Gregerson, K. (eds.) *Austroasiatic Studies: Papers from International Conference on Austroasiatic Linguistics (ICAAL4)*. Vol. 2. Mon-Khmer Studies Special Issue 3, 222–32.
- Wayesha, A. J. 2010. A phonological description of Leinong Naga. MA thesis. Payap University, Chiang Mai.
- Wei, M. 2012. Xia'ao Zhuang reference grammar. PhD thesis. Shanghai Normal University.
- Weidert, A. K. 1987. *Tibeto-Burman Tonology: A comparative account*. Amsterdam and Philadelphia: John Benjamins.
- Westaway, K. E., Louys, J., Awe, R. D., Morwood, M. J., Price, G. J., Zhao, J. X., Aubert, M., Joannes-Boyau, R., Smith, T. M., Skinner, M. M., Compton, T., Bailey, R. M., van den Bergh, G. D., de Vos, J., Pike, A. W. G., Stringer, C., Sapomo, E. W., Rizal, Y., Zaim, J., Santoso, W. D., Trihascaryo, A., Kinsley, L. and Sulistyanto, B. 2017. An early modern human presence in Sumatra 73,000–63,000 years ago. *Nature* 548: 322–5.
- White, J. C. 2011. Cultural diversity in Mainland Southeast Asia: A view from prehistory. In Enfield, N. J. (ed.) *Dynamics of Human Diversity: The case of Mainland Southeast Asia*. 9–46. Canberra: Pacific Linguistics.
- Wilkinson, E. L. 2009. Typology of signed languages: Differentiation through kinship terminology. PhD thesis. University of New Mexico, Albuquerque.
- Wnuk, E. 2016. Semantic specificity of perception verbs in Maniq. PhD thesis. Radboud University, Nijmegen.
- Wnuk, E. and Majid, A. 2014. Revisiting the limits of language: The odor lexicon of Maniq. *Cognition* 131 (1): 125–38.
- Wolfenden, S. N. 1929. *Outlines of Tibeto-Burman linguistic morphology: With special reference to the prefixes, infixes and suffixes of Classical Tibetan and the languages of the Kichin, Bodo, Nāgā, Kuki-Chin and Burma groups*. London: Royal Asiatic Society.
- Wolters, O. W. 1967. *Early Indonesian Commerce: A study of the origins of Sri Vijaya*. Ithaca: Cornell University Press.
- Woodward, J. 1993. Lexical evidence for the existence of South Asian and East Asian language families. *Journal of Asian Pacific Communication* 4 (2): 91–106.
- Woodward, J. 1996. Modern standard Thai Sign Language, influence from ASL, and its relationship to original Thai sign varieties. *Sign Language Studies* 92: 227–52.
- Woodward, J. 2000. Sign languages and sign language families in Thailand and Viet Nam. In Emmorey, K. and Lane, H. (eds.) *The Signs of Language Revisited: An anthology to Honor Ursula Bellugi and Edward Klima*. 23–47. Mahwah, NJ, and London: Lawrence Erlbaum Associates.
- Woodward, J. 2003. Sign languages and deaf identities in Thailand and Viet Nam. In Monaghan, L., Schmalzing, C., Nakamura, K. and Turner, G. H. (eds.) *Many Ways to Be Deaf: International variation in deaf communities*. 283–301. Washington, D.C.: Gallaudet University Press.
- Woodward, J. 2015. Hai Phong sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 351–60. Berlin: De Gruyter Mouton.
- Woodward, J., and Suwanarat, K. 2015. Original Bangkok sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 677–86. Berlin: De Gruyter Mouton.
- Woodward, J. and Wongchai, T. 2015. Original Chiang Mai sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 687–700. Berlin: De Gruyter Mouton.
- Woodward, J., Bradford, A., Sokchea, C. and Samath, H. 2015a. Cambodian sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 159–76. Berlin: De Gruyter Mouton.

- Woodward, J., Nguyen, T. H., Nguyen, D. M. G., Le, T. T. H., Luu, N. T., and Ho, T. V. 2015b. Ho Chi Minh City sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 401–18. Berlin: De Gruyter Mouton.
- Woodward, J., Nguyen, T. H., Nguyen, H. L., Nguyen, T. L., Nguyen, M. N., and Nguyen, T. T. T. 2015c. Ha Noi sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 335–50. Berlin: De Gruyter Mouton.
- Woodward, J., Suksiri Danthanavanich, S. and Janyawong, P. 2015d. Modern Thai sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 629–47. Berlin: De Gruyter Mouton.
- Wright, R. and Ladefoged, P. 1994. A phonetic study of Tsou. *UCLA Working Papers in Phonetics* 87: 67–89.
- Wrigley, O., Suwanarat, M., Ratanasint, A., Rungsritthong, V., and Anderson, L. B. 1990. *The Thai Sign Language Dictionary*. Revised and expanded edition. Bangkok: National Association of the Deaf in Thailand.
- Wu, M. 2015. A grammar of Sanjiang Kam. PhD. University of Hong Kong. <http://hdl.handle.net/10722/221219>
- Wulff, K. 1934. Chinesisch und Tai: Sprachvergleichende Untersuchungen. Copenhagen: Levin and Munksgaard.
- Yang, J. 2015. Chinese sign language. In Jepsen, J. B., de Clerck, G., Lutalo-Kiingi, S. and McGregor, W. B. (eds.) *Sign Languages of the World: A comparative handbook*. 177–94. Berlin: De Gruyter Mouton.
- Yang, T. and Edmonson, J. A. 2008. Kam. In Diller, A., Edmondson, J. A. and Luo, Y. (eds.) *The Tai-Kadai Languages*. 509–84. Abingdon, Oxon and New York: Routledge.
- Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.
- Yu, D. 2007. *Aspects of Lisu Phonology and Grammar; a Language of Southeast Asia*. Canberra: Pacific Linguistics.
- Yu, J. 2010. Aizhai Miao reference grammar. PhD thesis. Minzu University of China, Beijing.
- Zee, E. and Xu, L. 2003. Shanghai. In Thurgood, G. and LaPolla, R. J. (eds.) *The Sino-Tibetan Languages*. 131–45. Abingdon, Oxon and New York: Routledge.
- Zeshan, U. 2003. Indo-Pakistani sign language grammar: A typological outline. *Sign Language Studies* 3 (2): 157–212. <https://doi.org/10.1353/sls.2003.0005>
- Zeshan, U. 2004. Hand, head, and face: Negative constructions in sign languages. *Linguistic Typology* 8 (1): 1–58. <https://doi.org/10.1515/lity.2004.003>
- Zeshan, U. (ed.) 2006. *Interrogative and Negative Constructions in Sign Languages*. Nijmegen: Ishara Press.
- Zeshan, U. and de Vos, C. (eds.) 2012. *Sign Languages in village communities: Anthropological and linguistic insights*. Berlin: De Gruyter Mouton.
- Zhang, M. and Gong, T. 2016. How many is enough? Statistical principles for lexicostatistics. *Frontiers in Psychology* 7 (1916) www.frontiersin.org/article/10.3389/fpsyg.2016.01916
- Zhang, J., Lai, Y. and Turnbull-Sailor, C. 2006. Wug-testing the ‘tone circle’ in Taiwanese. Baumer, D., Montero, D. and Scanlon, M. (eds.) *Proceedings of the 25th West Coast Conference on Formal Linguistics*. 453–61. Somerville, MA: Cascadilla Proceedings Project.
- Zheng, Y. 1997. *Huihui Yu Yanjiu* [A study of Cham]. Shanghai: Yuandong Chuban She.
- Zhou, G. 2000. The origin and development of the ‘disposal’ form in Bouyei language. In Burusphat, S. (ed.) *Proceedings of the International Conference on Tai Studies*. 443–56. Bangkok: Institute of Language and Culture for Rural Development.
- Zhou, Z. and Yan, Q. 2006. 佤语简志 *Wa yu jianzhi* [An outline of the Wa language]. Beijing: Min zu chu ban she.

AUTHOR INDEX

- Adelaar, A., 70
Aikhenvald, A. Y., 283, 342
Andvik, E., 134
Aung-Thwin, M. A., 21, 22, 30

Bellwood, P., 1, 13, 14, 15, 16, 26, 28, 30 1, 66
Benedict, P. K., 55, 74, 76, 82, 85, 94
Benjamin, G., 47
Bisang, W., 300
Blench, R., 4, 9, 82, 89, 96
Blood, D. E., 117
Blust, R., 66, 67, 69, 93, 94
Bradley, D., 86, 87, 93, 154
Brunelle, M., 204, 205
Burenhult, N., 270, 289, 324 5
Burling, R., 79

Chao, Y. R., 84
Comrie, B., 1, 155
Coupe, A. R., 127
Culas, C., 24, 72

DeLancey, S., 90
DiCanio, C. T., 173, 174
Diffloth, G., 93, 167, 169, 170, 192
Diller, A. V. N., 55, 106

Edmondson, J. A., 51, 53, 55, 58

Fix, A., 41

Gedney, W. F., 60, 61, 62, 63
Grinevald, C., 283
Gussenhoven, C., 203

Haiman, J., 281
Hall, K. R., 16, 19, 23

Hartmann, H., 110, 111, 112, 117
Hartmann, J. F., 108
Haudricourt, A.-G., 184, 186, 188
Headley, R. K., 49, 49, 172, 172
Henderson, E. J. A., 174
Hickman, J. R., 116
Higham, C., 16, 17, 18, 19, 31
Hnin Tun, S. S., 389
Hyman, L. M., 200, 201

Jakobson, R., 330
Jenny, M., 183, 282, 287, 389
Jepsen, J.B., 100, 101

Kato, A., 348, 349
Kelly, P., 114
Kendon, A., 100
Kirby, J. P., 204, 206, 207, 208, 209
Kratochvíl, P., 209, 210, 218
Kruspe, N., 126, 236, 270, 288, 321, 326, 344, 345
Kurabe, K., 127

LaPolla, R. J., 9, 35, 80, 82, 95, 96
Larish, M. D., 70, 72, 159, 193, 193,
206, 208
Leach, E., 7, 30
Li, C. N., 308
Li, F. K., 59, 60, 65, 111, 112
Liem, N. D., 151
Lu, T., 201, 202
Lyman, T. A., 292

Mao, Z., 77
Masica, C., 108
Matisoff, J. A., 32, 35, 36, 38, 82, 86, 87, 89,
167, 168, 222, 308
Matthews, S., 144, 265, 337, 372

- McColl, H., 41, 42
Michaud, J., 24, 27, 72
Moore, E., 22

Nettle, D., 3
Nguyen, T., 387
Nonaka, A. M., 98, 99, 106
Norman, J., 81, 84, 190
Norquest, P. K., 58

O'Connor, R. A., 16, 24, 25, 26, 31
Ostapirat, W., 55, 65, 66
Owen, R. W., 55

Payne, T. E., 124
Peterson, D. A., 273
Pham, A., 205
Pittayaporn, P., 59, 72, 208
Post, M. W., 9, 39, 82, 89, 96, 206,
 211 13
Preecha, S., 168
Premsrirat, S., 155, 286

Ratanakul, S., 159
Ratliff, M., 38, 72, 73, 74, 75, 76, 77, 197, 197,
 252 6
Rau, F., 44, 46, 47
Reardon, W., 97
Reid, L. A., 94

Sagart, L., 89
Sapir, E., 124
Schmidt, W., 93

Scott, J. C., 27, 28
Searle, J. R., 8
Shafer, R., 85
Sidwell, P., 44, 46, 49, 50, 51, 115, 243
Snyder, W. C., 201, 202, 203
Solnit, D. B., 51, 53, 55, 58, 59,
 198, 199
Stargardt, J., 106
Starosta, S., 93

Taylor, K., 16, 17, 18, 19
Thomas, D. D., 49, 49, 150
Thompson, L. C., 241, 327
Thompson, S. A., 308
Thongkum, L., 59, 74, 77, 78
Thurgood, G., 58, 70, 71, 72, 79, 82, 88,
 90, 91, 95, 277, 300

van Driem, G., 88
van Liere, W. J., 19
van Schendel, W. 27
VanBik, K., 200, 201

Wang, F., 74, 76, 77, 279
Watkins, J., 4, 6, 33, 34
White, J. C., 11, 15
Woodward, J., 98, 101, 102, 103,
 105, 106

Yip, V., 144, 265, 337, 372

Zeshan, U., 97, 106
Zheng, G., 279

LANGUAGE INDEX

- Abo-Miri-Dafla, 85
Acehnese, 69, 70
Achino-Cham, 69
Ahom, 55, 111
Aizhai Miao, 293, 300 3
Akha, 90, 91
Anong, 36
Ao, 36
Ao Yao. *See* Muen
Arem, 36
Aslian, 36, 41, 43
Atayalic, 66
Atong, 36
Austric, 93
Austroasiatic, 3, 21, 28, 30, 38, 43, 46 8, 49, 74, 92, 93, 124, 242, 259
Austronesian, 3, 28, 43, 66, 66, 92, 93
Austro-Tai, 74, 93, 94
- Bahasa Malaysia, 165
Bahiing-Vayu, 85
Bahmar, 46
Bahnaric, 49, 243
Bai, 87, 88
Balinese, 70
Bangladesh Khumi, 223
Baoding, 55
Baric, 85
Barish. *See* Bodo-Garo
Biao Min, 77
Bih, 378 9
Black Tai, 55, 111
Bodic, 85
Bodish, 88
 East Bodish, 88
Bodish-Himalayish. *See* Tibetan-Kanauri
Bodo, 89
- Bodo-Garo, 85
Brou, 129
Bugan, 251
Bunun, 66
Burmese, 10, 30, 79, 89, 90, 105, 110, 226, 380 1, 389 90
 Burmese-Lolo, 85, 87
 See also Lolo-Burmese
Burmic, 85
Burmish, 85
Buyang, 55
Bwe Karen, 155
- Cambodian, 36
Cantonese, 6, 162, 168, 185, 290 1, 355 6, 379 80
Central, 86
Chamic, 17, 69 70
Chepang, 88
Chinese, 39, 79, 87, 88, 94, 124, 129, 176
Chrau, 206
Chru, 69, 71
Chuanqiandian. *See* West Hmongic
Colloquial Eastern Cham, 226
- Daai-Chin, 206
Dànnánshān Miao, 161, 168
deaf languages. *See* signed languages
Deuri, 36
Dhimal, 87
Dioi, 55
Dong, 152, 179
Dravidian, 129
- East Formosan, 66
Eastern Cham. *See* Phan Rang Cham
Eastern Kayah Li, 264, 382

- Eastern Kiranti, 90
 English, 33, 34, 120, 121, 124, 138, 143, 211, 214, 216, 267, 277, 300, 318, 326, 368
 Garo, 36
 Gelao, 55
Gelao-Hlai. See Hlai
 Hainan Cham, 36, 276 7
 Hakha Lai, Hakka Lai, 175 6, 273 4
 Haroi, 69
 Heitu, 55
 Himalayan, 87
 Hlai, 55, 155
 Hmong, 6, 24, 116, 225, 364
 Green Hmong, 116
 Hmong East, 74
 Hmong North, 74
 Hmong West, 74
 Hmong White. *See* White Hmong
 Hmongic, 72
 East Hmongic, 76
 West Hmongic, 76
 Hmong-Mien, 3, 8, 24, 28, 38, 72 8, 92, 93, 129, 155, 168
 Hmoob Dawb. *See* White Hmong
 Ho Ne. *See* She
 Idu, 134
 Idu-Digaru, 88
 Indic, 87, 108, 129
 Indo-Chinese, 94
 Indo-Iranian, 129
 Inuit-Aleut, 123
 Iu Mien, 36, 391 2
 Iu-men, 74
 Jahai, 36, 270, 289, 313, 324 5
 Japanese, 129
 Jarai, 69, 71
 Jinghpaw, Jingphaw, Jingpho, 38, 127, 327
 Jingxi Zhuang, 276
 Judu Gelao, 294 5
 Kachin, 85, 89
 Kadai. *See* Tai-Kadai
 Kaman-Meyor, 88
 Kamengic, 88
 Kammu, 142
 Kammu Yùan, 36
 Kam-Sui, 55
 Kam-Tai, 55, 59
 Karbi, 36
 Karen, 34, 190
 Karenic, 85, 88, 129 30
 Katu, 46
 Katuic, 49, 243
 Kham, 88
 Kham-Magar, 90
 Khamti, 55, 111
 Khasi, 218
 Khmer, 10, 21, 43, 124, 129, 226, 329 30, 375
 Khmu, 125, 155, 158, 163, 286 8
 Kim Mun, 74, 77
 Kiong Nai, 76
 Kiranti, 85, 87, 88
 Kmhmau, 124
 Koho, 250
 Korean, 129
 Koro-Milang, 88
 Kra, 55
Kra-Dai. See Tai-Kadai
 Kri, 10, 36, 124, 154, 155, 166, 167, 169, 170, 175, 190, 216, 314 15, 331 2
 Kuay, 168, 169
 Kucong, 381 2
 Kui, 169
 Kuki-Chin, 89, 90
 Kuki-Naga, 85
Kukish. See Kuki-Naga
 Kuy, 375, 385
 Lachi, 55
 Laha, 55
 Lahu, 10, 90, 126, 297 9, 309
 Lakkia, Lakkja, 58, 59
 Lao, 6, 10, 36, 55, 112, 124, 126, 147 50, 152, 164, 180, 182, 282, 296 7, 345, 384 5
 Lepcha, 87
 Lisu, 36, 38, 90
 Lolo-Burmese, 88
 See also Burmese-Lolo
 Loven, 115
Luoxiang. See Muen
 Magar, 88
 Malacca, 49
 Malay, 70
 Malayic, 69
 Malayo-Chamic, 69

- Malayo-Polynesian, 66
 Malayo-Sumbawan, 66
 Mandarin, 6, 151, 162, 180, 217 18, 271, 291 2, 305 8, 379
Miao. *See Hmongic*
 Middle Chinese, 184
 Mien, 10, 77
 Mienic, 72
 Min, 84
Mirish. *See Abo-Miri-Dafla, Tani*
 Mo Piu, 155
 Modern Western Cham, 165
See also Western (Cambodian) Cham
 Moken, 10, 70, 165, 206, 208, 208
 Moklen, 70, 165, 193, 206, 208
 Moklenic, 70 2
 Mon, 44, 130, 226
 Mong Njua, 292 3
 Mongolian, 129
 Mongsen Ao, 36
 Mon-Khmer, 8, 38, 49 51
 Moob Ntsuab. *See Green Hmong*
 Muen, 77
 Mulao, 263, 279 80, 304 5
 Munda, 38, 44, 47 9
 Muong, 44
- Naga, 89
 national languages
 Burmese, 33, 35, 107, 108, 111
 Cambodian, 43
 Chinese, 108
 Khmer, 33, 35, 107, 108, 111
 Lao, 33, 35, 52, 107, 108, 111
 Mandarin Chinese, 33, 35
 Thai, 33, 35, 52, 107, 108, 111
 Vietnamese, 33, 35, 43,
 107, 108
See also individual entries.
 Naxi, 87
 Newar, 88
 Ngeq, 132
 Nicobarese, 49
 North Hmongic, 76
 Northeastern, 86
 Northeastern Border, 88
 Northern Vietnamese, 178
 Northwest Formosan, 66
 Nung, 55, 90, 303 4
 Nungish, 87, *See Rawang*
- Old Chinese, 39
 Pa Hng, Pa-hng, PaHng, 74, 76, 77
 Pacoh, 36, 370 1
 Paha, 55
 Paiwan, 66
 Palaung, 133
 Pali, 106, 110, 111
 Pearic, 49
 Pepcha, 87
 Phan Rang Cham, 69, 220
 Proto-Austroasiatic, 38
 Proto-Austronesian, 66
 Proto-Chamic, 69
 Proto-Hmong-Mien, 73, 74, 75, 77
 Proto Hmongic. *See Proto-Miao*
 Proto Lolo-Burmese, 90, 91
 Proto-Miao, 74
 Proto-Mienic, 74
 Proto-Southwestern-Tai, 38
 Proto-Tai, 38, 60
 Proto-Tibeto-Burman, 90
 Pubiao, 55
 Puyama, 66
 Pwo Karen, 223, 347 50
 Pyū, 23, 88
- Qiandong. *See East Hmongic*
 Qiangic, 88
- Rade, 69, 71
 Raji, 88
 Raute, 88
 Rawang, 88
 Rengao, 171
 rGalryong, 90
 Riau Indonesian, 125
 Roglai, 69
 Rukai, 66
 Rung, 88
 Ruyuan Mien, 142
- Saek, 5, 10
 Sal, 86, 88, 89
 Sanskrit, 106, 109, 111
 Santali, 43, 44
 Sasak, 70
 Sedang, 170, 251
 Semelai, 10, 36, 124, 126, 129, 236, 270 1,
 288 9, 335 6, 343, 358 9

- Sgaw Karen, 146, 158
 Shan, 53, 55, 111
 Shanghainese, 6, 162
 She, 76
 Siamese, 55, 111
 signed languages, 96 106
 - American Sign Language, 102, 105
 - Ban Khor Sign Language, 98, 99
 - Cambodian Sign Language, 101, 105
 - Chinese Sign Language, 101
 - French Sign Language, 103, 105
 - Ha Noi Sign Language, 101, 103, 104
 - Hai Phong Mai Sign Language, 104
 - Hai Phong Sign Language, 101, 103, 104, 105
 - Ho Chi Minh City Sign Language, 101, 103, 104
 - Laos Sign Language, 101
 - Modern Thai Sign Language, 101, 102, 103, 105
 - See also* Thai Sign Language.
 - Original Bangkok Sign Language, 101, 103
 - Original Chiang Mai Sign Language, 101, 103
 - Plains Indian Sign Language, 100
 - Thai Sign Language, 99
 - See also* Modern Thai Sign Language
 - Yangon Sign Language, 105
- Sinitic, 81 4, 89, 129
 Sino-Bodic, 88
 Sino-Tibetan, 3, 38, 39, 43, 73, 79 92, 93, 94, 168, 259
 Southeastern, 86
 Southern Min, 224
 Southwestern Mandarin, 84
 Sumbawa, 70
- Tai, 20, 28, 36, 55, 94, 111, 124, 130
 Tai Dam. *See* Black Tai
 Tai-Kadai, 3, 38, 51 66, 73, 74, 92, 93, 158
 Tamangic, 88
 Tani, 88
 Thai, 6, 10, 21, 36, 110, 124, 203, 213, 225, 343, 363
 Thailand Mien, 293 4
- Tho, 55
 Tibetan, 79, 88
 Tibetan-Kanauri, 85
 Tibeto-Burman, 8, 9, 28, 76, 81, 87, 128, 129, 134, 242
 Trans-Himalayan, 9, 88
 Tsat, 69
 Tshangla, 88
 Tsou, 155
 Tsouic, 66
 Tujia, 87, 88
 Turung, 36
- Uralic, 129
- Vietnamese, 10, 43, 124, 151 2, 155, 175, 183 4, 204, 224, 236 41, 327 8
- Wa, 36, 224
 West Himalayan, 88
 Western, 86
 Western (Cambodian) Cham, 69, 329
 - See also* Modern Western Cham
- Western Plains, 66
 Western-Malayo-Polynesian, 66
 White Hmong, 116, 178, 218, 252, 268 70, 343
- White Tai, 55
 Wuming Zhuang, 201 3
- Xia'ao Zhuang, 295 6, 377
 Xiangxi. *See* Xiong
 Xiong, 76, 77
 Xong, 382 3, 391
- Yao. *See* Mienic
 Yongning Na (Mosuo), 223
 Yuanmen, 55
 Yue, 84
- Zao Min, 77
 Zhuang, 53
 Zhuang-Dong. *See* Tai-Kadai

SUBJECT INDEX

- adjectives, 211 14
adpositions, 138 9, 214 16
affixes
 reconstruction, 90
agreement, 120, 326
agriculture, 18, 24, 26, 28, 31, 41, 42
 early, 12, 13 15, 18
 millet, 95
 rice, 1, 15, 19, 26, 28, 55
 taro, 15
ambiguity, 125, 128, 353
Angkor, 19, 20, 21, 23, 24
animacy, 127,
Annamite range, 11
Arakanese coast, 23
argument structure, 120
Assam basin, 23
avoidance relationship, 100
Ayutthaya, 20
- Ban Khor, 98,
Bengali coast, 23
bioinformatic techniques, 43
Boloven Plateau, 115
borrowing, linguistic, 39, 94, 279
Brahmaputra, 1
breathy voice. *See* phonation type
Burman, 23
- Cambodia, 1, 19, 43, 110
case marking, 120, 126, 127, 326
 lack of, 120
causative constructions, 368
 indirect causation, 369
 with ‘give’, 369
 with ‘take’, 373
Champa, 17
- Chaophraya River, 1, 20, 21
Chenla, 16, 17
China, 1, 17, 20, 24, 72, 86
clitics, 205, 206
colonialism, 6, 28, 32, 33, 34, 113, 394
comparative constructions, 139 41
complementizer, 366 8
compounds, 209, 216 22, 241, 252
computational phylogenetic methods, 50,
 78, 89
consonant clusters, 156, 166
constituent order, 125, 129 30, 259, 326, 348
 and adjectives, 139 41
 and adpositions, 138 9
 flexible, 120, 125, 128, 343, 359
 head initial, 260
 in noun phrases, 132 5
 in numeral classifiers, 284, 288, 295
 in possession, 137 8, 278
 in relative clauses, 136
 verb and object, 131 2
contact
 linguistic, 87, 137, 262, 267, 275,
 318, 393
 contact-induced language change, 40, 71, 74,
 87, 129, 130, 208, 275, 278
core MSEA, 1, 3, 53, 111, 134
coverbs, 271, 372, 379, 380
cultural diffusion, 31
- deafness. *See* signed languages, in Language Index
demic diffusion, 31
depictive constructions, 129, 360
derivational morphology, 124
dictionaries, 33, 36, 40, 106, 113
disposal construction, 374

- diversity
 linguistic, 3, 10, 32, 118, 393
 domestication, 13, 95
 dual number, 310
 Dvāravatī, 22
- Easter Island, 66
 ecology, linguistic, 5
 ellipsis, 125, 330, 343
 endangerment, 3, 36, 99, 105
 ergativity, 121
 anti-ergative, 126, 127
 expatriate communities, 43
- French Indochina, 28
 Funan, 16, 17
- genetics, 21, 31, 40, 41, 42
 gestures, 96, 97
 globalization, 394
 Golden Pali Text, 106
 grammatical categories, lack of, 122, 124, 128, 308, 331
 grammatical distinctions,
 lack of, 327
 Great Silver Reliquary, 106
 greater MSEA, 1, 134
 grounding, 120
 Guangxi, 53, 58, 84, 136
 Guizhou, 72
- Hainan, 84
 Hainan Island, 58, 69
 Himalayas, 11
 historical comparative linguistics, 15, 38, 39, 50, 55, 74, 78, 86, 89, 92
 Hoabinhian, 13, 28, 41
 Human, 72
 hunter-gatherers, 13, 28
- idioms, 97
 implosive stops, 155
 inclusory constructions, 316
 incorporation, 372
 index of synthesis, 123, 124
 India, 23, 43, 86, 111, 134
 infixation, 243, 248
 interrogatives, 141–4
 intonation, 203–4
 in multi-verb constructions, 344
- iron, 26, 95
 Irrawaddy, 1, 21, 23, 30
 irrigation, 19, 25
 Is Npis Mis Nus, 116
- juxtaposition, 277
- Kashmir, 79
 Khmer, 16, 19, 20, 21
 kinship, 99
- Laos, 1, 20, 72, 116
 lexical reconstruction, 39, 66, 74, 81, 82, 92, 93
 Luang Prabang, 20
- macro-family proposals, 93, 94
 Madagascar, 66
 maize, 24
 Malacca, 17
 Malayo-Polynesians, 17
 Malaysia, 41, 43, 70
 maritime silk route, 16
 maritime trade, 16, 17, 18, 19, 23
 measure words, 251
 Mekong, 1, 16, 17, 18, 20, 20, 43
 migration, 40, 42, 72, 394
 ancient, 11
 and agriculture, 31
 and ethnolinguistic diversity, 28, 30, 32, 39, 41, 53
 bringing languages, 31
 early, 40, 42, 66, 394
 into MSEA, 15
 out-migration, 3
 waves of, 28, 41, 72, 87, 95
- minority languages, 34, 105, 111, 114, 325
 Mon, 16, 20, 22, 23
 Mongol, 24
 Mongolia, 79
 multilingualism, 87
 mutual intelligibility, 6
 Myanmar, 1, 43, 70, 111, 134
- national languages. *See* Language Index
 nationalism, 394
 necessity, marking of, 340
 New Zealand, 66
 Nicobar Islands, 28
 nomadism, 17
 Northeast India, 1

- Orang Asli, 41
 orthography, 6, 107 17
 alphabetic, 108
 Brāhmī, 108
 Chữ Hán, 112
 Chữ Nôm, 112
 Chữ Quốc Ngữ, 112, 113
 Hmong Romanized Popular Alphabet (RPA), 107, 116
 Indian, 18
 Indic-based scripts, 108, 109, 112
 Khom, 115
 lack of, 107, 114
 Latin-based, 113, 116
 loss of, 116
 Mon, 23
- Pagan, 20, 24
 palm leaf inscription, 59, 110
 Pearl River, 1
 Peninsular Malaysia, 1
 phonation type, 154, 173, 175, 194, 195
 breathy, 37, 119, 168, 169, 172, 174, 178, 194, 254, 255
 phonological reconstruction
 PMP, 69
 Proto Austroasiatic, 46, 47
 Proto Austronesian, 71
 Proto Tibeto-Burman, 82
 Proto-Hlai, 58
 Proto-Hmong-Mien, 75
 Proto Lolo-Burmese, 90 2
 Proto-Tai, 60
 politeness, 293, 308, 312, 313, 314
 political entities, 17, 18, 19, 20, 22, 23, 32
 political independence, 34
 political instability, 40
 political power, centralisation of, 32
 population density, 18
 possession, 135 6, 257, 277 82
 possibility, marking of, 340, 341 2
 postcolonialism, 7, 29, 33, 35
 prosody, 203, 204 6, 365
 proto-languages. *See* individual entries in
 Language Index
 Pyū, 16, 21, 23, 106
- quantification, 257
 quotation, direct and indirect, 365
- reconstructed tonal systems, 65
 Red River, 1
 reduplication, 124, 227 42
 register, 37, 119, 167, 168 74, 175, 185, 195
 relative clauses, 135 6, 266 77
 religion, 19, 21, 23, 106, 116
 animism, 33
 Buddhism, 22, 23, 33, 106
 Christianity, 33
 Hinduism, 18, 19, 20, 106
 Islam, 33
 missionaries, 110, 112
 resumptive pronouns, 269, 271
 retroflex stops, 154
 rimes, 75, 166, 168, 178, 179, 194
 Rohingya, 33
- Salween, 1
 sentence-final particles, 123, 141, 144 5, 204, 210, 331
 Shan, 30
 shift language, 3, 20, 33, 72, 99, 193, 207, 208, 393
 social dynamism, 7, 11, 28
 South East Asia Linguistic Society (SEALS), 35
 speech event, 120
 Śrī Kṣetra, 106
 standardisation, 3
 stress, 72, 166, 205 6
 subject, grammatical, 121, 126, 249, 267, 270, 328, 346, 363, 364, 378
 subordination, 129, 250, 363, 364
 Sukhothai, 20
 Sumatra, 11, 28
 Suphanburi, 20
 syllable structure, 71, 76, 112, 154, 161, 177
- Tai, 20
 Taiwan, 66
 Tang era, 18
 tertiary dispersion, 26
 Thai Royal Institute, 106
 Thailand, 1, 20, 69
 tone, 59, 112, 113, 119, 154, 252 4
 definition of, 175
 grammatical description of, 174 83
 tonation, 154
 tone sandhi, 195 203
 tonogenesis, 183 95

- Tonle Sap, 1
transitivity. *See* valency
- valency, 119
ambitransitivity, 121
bivalent, 119, 120
monovalent, 119, 120
valency-changing derivations, 249, 250
- Vientiane, 20
- Vietnam, 1, 43, 72
- voice, 121
antipassive, 121
grammatical, 327
passive, 121, 327, 328, 329, 330
- vowel mutation, 228
- warfare, 17, 18, 19, 24, 30, 87
- word
grammatical, 205 8, 337, 342
iambic structure, 207
phonological, 205 8
- word structure
isolating, 90, 118, 124, 125, 205, 207, 209 10
monosyllabic, 208
sesquisyllabic, 166, 167 8, 183, 206
- Yangtze River, 1
- Yunnan, 21, 72, 84
- zero anaphora, 125, 128, 259
- Zomia, 27, 43, 115