Linda Konnerth
A Grammar of Karbi

Mouton Grammar Library

Edited by
Georg Bossong
Bernard Comrie
Patience L. Epps
Irina Nikolaeva

Volume 82

Linda Konnerth

A Grammar of Karbi

DE GRUYTER
MOUTON

ISBN 978-3-11-064885-0
ISSN 0933-7636

Library of Congress Control Number: 2020931427

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

© 2020 Walter de Gruyter GmbH, Berlin/Boston
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Acknowledgments

This grammar is a revised version of my doctoral dissertation, which I completed at the University of Oregon in 2014. I remain grateful to everybody I acknowledge in my dissertation. In addition, I am most grateful to my wonderful editor Pattie Epps, who had excellent comments and suggestions, which in many ways improved this grammar. The revisions I have carried out were also influenced by the feedback from the ALT's Panini award committee, headed by Nick Enfield, as well as by Scott DeLancey, whose research never ceases to inspire me. My colleague and partner Pavel Ozerov has supported my work for the publication of this grammar as well as all my other research in all possible ways. In the final stages of proof-reading and getting the manuscript ready for print, I was lucky to receive tremendous help from Raphael Kraut, and I am also very thankful to the professional and friendly de Gruyter Mouton team.

For the original research that led to the first version of this grammar and my dissertation, my deepest gratitude goes to my mentor Scott DeLancey. In addition to sharing his knowledge, Scott has always supported me, inspired me, and led me back to see the big picture when I was lost in details. The Karbi descriptive grammar project was born in 2007 when members of the *Karbi Lammet Amei* (Karbi Literary Association) contacted Scott. This was serendipitous, as Scott had just weeks earlier given me the late Karl-Heinz Grüßner's 1978 grammar of Karbi phonology and morphology to look at; I happily took on this exciting project. Before my first trip to India, I was able to get in touch with Karl-Heinz and visited him in Tübingen, Germany, where he let me into the magical realm of his attic. I spent hours perusing his notes, transcriptions, and other Karbi materials he had collected. Meeting and getting to know Karl-Heinz was a very special experience and I will always remember and be grateful for his generosity, encouragement, and the friendly welcome I received from him and his family, both in Tübingen and in Shillong. Karl-Heinz suddenly passed away in 2014 but he lives on in his invaluable work on the Karbi language, as well as in the memories we have of his warm and friendly personality and the jokes he was never short of.

During my first trip to India I met Sikari Tisso, who was to become my main language consultant and collaborator, and who made everything possible. His dedication and tireless efforts make him a true hero for all who value the wonderful language and culture of the Karbi people. While this project was ongoing, his son passed away much too young; this grammar is also dedicated to the memory of Sarpo Rongkhelan Tisso.

I am deeply grateful to UV Jose. His enormous knowledge of the languages in the area and of the people speaking them as well as his generosity and advice inspired and helped me during all this time. Many thanks also to Brother Benjamin Ingti Kathar, who shared his language and warm friendship with me.

This project would not have existed without the effort of the *Karbi Lammet Amei* to preserve their beautiful and rich language and inherent culture and knowledge for

future generations. Khorsing Teron as well as Dharamsing Teron (not actually a KLA member) volunteered much time and effort to the project. Thanks also to Hokursing Rongpi, and to Budheswar Timung (Nowgong KLA). Due to space limitations, other Karbi speakers that have made this project possible are acknowledged in §2.3.1.

I would like to thank Amphu Rongpipi, Klirdap Teronpi, Serdihun Beypi, Larshika Tissopi, and Sarlomet Tisso, who all greatly contributed to the project by preparing transcriptions, struggling through translations, and helping with the analysis.

None of my research would have been possible without a place for me to stay. In Diphu, *Kro hem* always had their doors open for me: Sashikola Hansepi, Member Kro, Ruplin, Sintu, Rasinza, and Sampri. Jirsong Asong in Diphu also offered me a warm and friendly place to stay - many thanks to Fr. Vinod! In Umswai, I stayed with the wonderful Hanse family: Albina Teronpi, Holiwel Hanse, Hunmily, Basapi, John, and Platinum.

At the University of Oregon, my dissertation committee offered great feedback and plenty of food for thought: thank you to Spike Gildea, Doris Payne, and Zhuo Jing-Schmidt! More generally, I feel very fortunate and grateful about my time at the Linguistics Department at the University of Oregon, where I received support, friendship, and encouragement from many sides. Thanks in particular also to my then fellow grad students, Shahar Shirtz, Amos Teo, Krishna Boro, Danielle Barth, Anna Pucilowski, Dan Wood, and especially Gwen Hyslop.

I am grateful to a number of other linguists who have given me feedback on my work and have inspired me, in particular, Northeast Indianists Mark Post, Stephen Morey, and Jyotiprakash Tamuli; Tibeto-Burmanists David Bradley, Christina Willis, Dave Peterson, Stephen Watters, and Robert Schikowski; as well as Antoine Guillaume.

This research was funded in part by National Science Foundation Grant # BCS-0951749. At the UO, further support was provided by the Center for the Study of Women in Society; the Center for Asian and Pacific Studies; and the Oregon Humanities Center. I was working through the revisions while in residence at the Martin Buber Society of Fellows at the Hebrew University of Jerusalem, whose support and engaging research environment is also gratefully acknowledged.

Finally, I have always had incredible support from my family and friends, and none of my work would have been possible without them. I have received love, encouragement, and advice from Roswitha, Arthur, Arne, Sascha, Irmi, Juli, Franzi, and everybody else in my extended family and of my wonderful friends; they all sustained me when I was writing my dissertation, and they all sustained me when I, with new challenges and new priorities, worked through the revisions. The revisions were supposed to be finished before Yaniv arrived, but they did not – and consequently, with much less time on my hands for the best reason possible, they could not have been finished without Pasha, my partner in life in every sense.

Contents

Acknowledgments —— V

List of Figures —— XXIII

List of Tables —— XXV

1 Introduction —— 1
1.1 Karbi people, language, and culture —— **1**
1.1.1 Names and ISO codes for the Karbi language —— **1**
1.1.2 Number of speakers and geographical spread of Karbi —— **2**
1.1.3 Aspects of Karbi culture —— **3**
1.1.3.1 Traditional culture and social organization —— **3**
1.1.3.2 Oral literature —— **5**
1.1.3.3 Song language —— **7**
1.1.4 The Karbi Lammet Amei (KLA) —— **8**
1.1.5 The Karbi Anglong and West Karbi Anglong districts —— **8**
1.2 Karbi in Tibeto-Burman —— **9**
1.2.1 Tibeto-Burman languages of Northeast India —— **9**
1.2.2 Karbi in Tibeto-Burman classification proposals —— **10**
1.3 On the role of contact in the development of Karbi —— **12**
1.3.1 Linguistic evidence for contact with Austroasiatic, Indo-Aryan, and other Tibeto-Burman languages —— **12**
1.3.2 Non-linguistic evidence for contact with populations speaking Austroasiatic and Indo-Aryan languages —— **14**
1.3.3 A creolization account for grammatical characteristics of Karbi —— **16**
1.3.3.1 Areal ethnolinguistic profile —— **17**
1.3.3.2 Aspects of a “creoloid” grammatical profile —— **18**
1.3.3.3 Non-obligatoriness of grammatical markers —— **19**
1.3.4 Summary —— **22**
1.4 Varieties of the Karbi language —— **23**
1.4.1 Plains Karbi (“Amri Karbi”) and Hills Karbi —— **24**
1.4.2 Relationships between the Hills Karbi ‘dialects’ —— **25**
1.4.3 Hills Karbi: Differences between Rongkhang (Hills Karbi; Diphu) and Amri (Hills Karbi; West Karbi Anglong) dialects —— **26**
1.4.4 Variation in lexemes —— **27**
1.5 Sociolinguistic profile of language endangerment —— **28**
1.5.1 Setting factors —— **28**
1.5.2 Domains/vitality/attitudes —— **29**
1.5.3 Other factors —— **30**
1.5.4 Summary of degree of endangerment —— **30**

1.6 Previous study of Karbi grammar — 31
1.6.1 Early work on Karbi — 31
1.6.2 Karl-Heinz Grüßner's work — 31
1.6.3 Longkam Teron's work — 32
1.6.4 Other resources on Karbi grammar and lexicon — 33
1.7 Writing in Karbi and conventions followed in this grammar — 33
1.8 Organization of this grammar — 35

2 Methodology and data — 36
2.1 Approach and theoretical framework — 36
2.1.1 Functional-typological framework — 36
2.1.2 Collaboration with the language community — 37
2.1.3 Historical-comparative perspective — 38
2.2 Corpus — 38
2.2.1 Data types — 38
2.2.2 Access to the corpus — 39
2.3 Primary data collection — 39
2.3.1 Acknowledgment of Karbi speakers that facilitated data collection — 40
2.3.2 Data collecting team — 41
2.3.3 Recorded speakers: Ethical considerations — 41
2.3.4 Recorded genres — 42
2.3.5 Recording procedures and settings — 43
2.4 Data annotation and presentation — 45

3 Phonology — 48
3.1 Consonants — 48
3.1.1 Consonant onsets — 48
3.1.1.1 Stop onset minimal sets — 51
3.1.1.2 Fricative onset minimal set — 52
3.1.1.3 Sonorant onset minimal sets — 52
3.1.1.4 Marginal onset consonants and consonant clusters — 52
3.1.2 Consonant codas — 53
3.1.2.1 Stop coda minimal set — 54
3.1.2.2 Sonorant coda minimal set — 54
3.1.2.3 Marginal coda consonants — 55
3.2 Vowels — 55
3.2.1 Hills Amri Karbi sixth vowel — 56
3.2.2 Marginal vowels and phonetic variation — 57
3.3 Syllable — 59
3.4 Palatal glide coda versus diphthongs — 60
3.5 Tone — 61
3.5.1 Tone minimal sets: Monosyllabic roots — 63

3.5.2 Tone minimal sets: Disyllabic roots — **65**
3.5.3 Phonological basis of the tone system — **65**
3.5.4 Low functional load — **67**
3.5.4.1 Native speakers' difficulty to identify tone categories — **68**
3.5.4.2 Low contrastiveness — **68**
3.5.4.3 Large amount of contextuality — **68**
3.5.4.4 Speaker differences in realization of mid versus high tones — **69**
3.5.4.5 Production-perception mismatch — **70**
3.5.4.6 Prosodic tone changes and indeterminability — **70**
3.5.5 Word-internal tone sandhi effects — **72**
3.5.6 Tone and phonological and morphological units — **73**
3.5.7 Tone (changes) in compounds — **75**
3.5.8 Strategies used for identifying tones of roots and suffixes — **76**
3.5.8.1 Syllable under investigation is the last syllable of the word and non-stopped — **76**
3.5.8.2 Syllable under investigation followed by *-lò* 'realis' — **76**
3.5.8.3 Syllable under investigation followed by *-Cē* 'negative' — **77**
3.5.9 Representation of tone in this grammar — **77**
3.5.10 Remarks on grammatical functions of tone — **78**
3.6 Stress — **80**
3.7 Characteristics of hypoarticulated speech — **81**
3.7.1 /t/ → Ø / ___ v, l — **81**
3.7.2 Imperative onsets: /n/ → [α nasal] / [α nasal] ___ — **81**
3.7.3 *-jí* 'IRR2?', *-lò* '?', *-si* '?' → a / ___ *-lāng* 'still', *-nāng* 'hortative' — **82**
3.8 Phonological shapes of morpheme types — **82**
3.8.1 Roots — **82**
3.8.2 Suffixes — **82**
3.8.3 Prefixes — **82**
3.8.4 Clitics — **83**
3.8.5 Interjections — **85**
3.8.6 Reduplication — **85**
3.8.6.1 Reduplication of stems (without vowel change) — **85**
3.8.6.2 Reduplication of stems with vowel change — **86**
3.8.6.3 Negation: Onset reduplication of last stem syllable — **87**
3.9 Morphophonemics — **88**
3.9.1 Tone changes — **88**
3.9.1.1 Stem tone change induced by *che-* 'reflexive/reciprocal' and *cho-* 'auto-benefactive/malefactive' — **88**
3.9.1.2 Stem tone change induced by possessive prefixes — **89**
3.9.1.3 Derivational suffix tone change — **90**
3.9.1.4 Idiosyncratic tone allomorphy — **91**
3.9.2 Prefix allomorphs and prefix-induced stem vowel deletion — **91**

3.9.2.1 Prefix vowels: *ke-* 'NMLZ', *pV-* 'CAUS', *che-* 'RR', *cho-* 'AUTO.BEN/MAL', *a-* 'POSS' — **92**
3.9.2.2 *cho-* 'auto-benefactive/malefactive' and *che-* 'reflexive/reciprocal': Differences in phonological strength — **95**

4 The major word classes: Nouns and verbs and 'adjectival verbs' — 97
4.1 Nouns and verbs — **97**
4.1.1 Morphosyntactic criteria for nouns and verbs at the root level — **97**
4.1.2 Lack of word class distinctions with suffixal aspect, modality, and negation — **99**
4.1.3 Summary: Nouns and verbs in Karbi — **102**
4.2 Verbs and 'adjectival verbs': The status of property concept terms — **102**
4.2.1 Basic property concept terms (PCTs) — **104**
4.2.2 Possible property concept term (PCT) criteria: Gradability, intensification, adverb derivation — **105**
4.2.3 Possible verbhood criteria: Aspectual and modal suffixes, directives — **108**
4.2.4 Noun modification — **110**
4.2.5 Summary: Basic property-concept terms (PCTs) — **111**
4.2.6 Marginal types of property concept terms (PCTs) — **111**
4.3 Summary: The clausal functions of nouns, verbs, and property concept terms (PCTs) — **115**

5 Subclasses of nouns — 117
5.1 Classifiers — **117**
5.1.1 Sortal classifiers — **119**
5.1.1.1 Typical classifiers — **119**
5.1.1.2 Not fully grammaticalized classifier *bàng* 'CLF:HUM:PL' — **121**
5.1.2 Mensural classifiers — **122**
5.1.2.1 Typical classifiers — **122**
5.1.2.2 Not fully grammaticalized classifiers — **123**
5.1.3 Self-referential classifiers — **124**
5.1.4 Human and animal classifiers and personification — **125**
5.2 Nouns that are counted with numerals only (without classifiers) — **127**
5.3 Human/personified vs. non-human/non-personified — **129**
5.4 Relator nouns — **129**
5.4.1 General relator noun *-lòng* 'locative' — **133**
5.4.2 Locational relator nouns — **135**
5.4.3 Locational/temporal relator nouns — **139**
5.4.4 Temporal relator noun — **141**
5.4.5 Causal relator nouns — **141**

5.4.6 Relator nouns with other functions — **142**
5.4.7 Further grammaticalization of relator nouns — **144**
5.4.7.1 Subordinators — **144**
5.4.7.2 Adverbs — **145**
5.5 Bound noun roots — **146**
5.5.1 Body part terms — **146**
5.5.2 Kinship terms — **147**
5.5.3 Other inalienably possessed items — **147**
5.6 Other possible noun classes with frozen prefixes (Grüßner 1978: 44–6) — **148**
5.7 Collective noun roots — **148**

6 Other word classes — 149
6.1 Pro-forms — **149**
6.1.1 Personal pronouns and personal possessive prefixes — **149**
6.1.2 Reflexive/reciprocal pronouns — **151**
6.1.3 Demonstratives — **154**
6.1.4 Interrogative pronouns and pro-adverbs, and positive polarity indefinite construction — **156**
6.1.4.1 Positive indefinite construction with *-nē* 'indefinite' — **159**
6.1.4.2 Interrogative pronouns in indirect questions — **159**
6.1.4.3 Co-relative construction — **160**
6.1.5 Parallelism between demonstrative and interrogative adverbs — **160**
6.1.6 Pronouns and pro-adverbs of universal quantification — **161**
6.2 Verb subclasses — **163**
6.2.1 Property concept terms — **163**
6.2.2 Copulas — **163**
6.2.2.1 Existential and locative copulas — **163**
6.2.2.1.1 Positive *dō* and negative *avē* — **163**
6.2.2.1.2 Quantifying copula *óng* 'exist much' — **167**
6.2.2.2 Negative equational copula *kalī* — **167**
6.3 Frozen prefixes in disyllabic verb and noun roots — **168**
6.3.1 Frozen prefix *ing-* — **168**
6.3.2 Frozen prefix *ar-* — **170**
6.3.3 Other possible frozen prefixes in disyllabic noun and verb roots — **170**
6.4 Minor word classes — **170**
6.4.1 Adverbs — **170**
6.4.1.1 Temporal adverbs — **171**
6.4.1.1.1 Day and year ordinals — **171**
6.4.1.1.2 Other underived temporal adverbs — **173**
6.4.1.1.3 Temporal adverbs with *-váng* 'every' — **173**
6.4.1.1.4 Intensifier reduplication construction — **173**

6.4.1.2 Locative adverbs — **173**
6.4.1.3 Other adverbs — **174**
6.4.2 Numerals — **174**
6.4.3 Subordinators and coordinators — **176**

7 Nominal morphology — 179
7.1 Overview: Noun stems and noun words — **179**
7.2 Compounding — **180**
7.2.1 Simple root compounding — **180**
7.2.2 Higher level elaborate expression compounding — **181**
7.3 Nominal prefixes — **181**
7.3.1 General possessive or 'attributive' prefix *a-* 'POSS' — **182**
7.3.2 Personal possessive prefixes — **184**
7.3.3. Verbalizer *pe- ~ pa-* — **185**
7.4 Nominal suffixes — **186**
7.4.1 Gender suffixes *-pī* and *-pō* — **187**
7.4.1.1 Female and augmentative *-pī* — **187**
7.4.1.2 Male (and non-productive modifier-deriving) *-pō* — **188**
7.4.1.3 Augmentative and modifier-deriving *-pī* and *-pō* — **188**
7.4.2 Diminutive *-sō* — **189**
7.4.3 Gender/augmentative and diminutive suffixes in names of animal (/plant) subspecies — **190**
7.4.4 Human plural *-mār* — **191**
7.4.5 Superlative *-si* — **191**
7.5 Reduplication — **192**
7.5.1 Dual — **192**
7.5.2 (Distributive) plural — **193**

8 Verbal morphology: Overview and pre-root slots — 196
8.1 Compounding — **196**
8.2 Overview: Position classes in the Karbi verb — **196**
8.2.1 Derivational and inflectional affixes, and the verb stem — **197**
8.2.2 Overview of pre-root slots — **198**
8.3 Proclitic slot: Non-subject speech act participant cross-referencing and cislocative marking — **200**
8.3.1 Non-subject speech act participant marking — **200**
8.3.1.1 Introduction — **201**
8.3.1.2 Cross-referencing SAP primary objects — **201**
8.3.1.3 Cross-referencing SAP in non-core roles — **204**
8.3.1.4 Summary — **205**
8.3.2 Cislocative marking — **206**
8.3.2.1 Directional — **206**

8.3.2.2 Associated motion ('come and V') — **208**
8.3.2.3 Semantic extensions — **210**
8.3.2.4 Summary: Cislocative marking — **211**
8.3.3 Summary — **212**
8.4 Prefixal derivational morphology — **213**
8.4.1 Nominalizer *ke-* — **213**
8.4.2 Causative *pe- ~ pa-* — **214**
8.4.3 Reflexive/reciprocal *che-* — **215**
8.4.4 Auto-benefactive/malefactive *cho-* — **219**
8.4.5 On the functional overlap between reflexive/reciprocal *che-* and auto-benefactive/malefactive *cho-* — **220**

9 Verbal morphology: Post-root slots — 222
9.1 Overview of post-root slots — **222**
9.1.1 Cooccurrence restriction between negative *-Cē* and irrealis *-jí* and *-pò* (slots 8 and 10) — **223**
9.1.2 Slot 9 *-pín* 'experiential' — **223**
9.1.3 Slot 10 *-làng* 'still' and *-jí-lāng* 'IRR2-still' — **223**
9.2 Suffixal predicate derivations — **224**
9.2.1 Overview — **225**
9.2.1.1 Structural properties — **225**
9.2.1.1.1 Scope of negation and reduplication — **225**
9.2.1.1.2 Discontinuous predicate derivations — **227**
9.2.1.1.3 Productivity — **228**
9.2.1.2 Origins of predicate derivations — **229**
9.2.1.3 Functional categories of predicate derivations — **230**
9.2.2 Manner — **230**
9.2.2.1 Non-ideophonic manner — **230**
9.2.2.2 Ideophonic manner — **231**
9.2.2.3 Degree or extent — **231**
9.2.2.3.1 Quantification derivations — **232**
9.2.2.3.2 Comparative *-mū~-mūchòt* and superlative *-néi~-nái* — **233**
9.2.3 Result — **233**
9.2.4 Direction, (associated) motion, path — **234**
9.2.5 Argument and argument structure related functions — **238**
9.2.5.1 Argument quantification — **238**
9.2.5.2 Argument structure highlighting — **239**
9.2.5.2.1 Benefactive/malefactive *-pī* — **239**
9.2.5.2.2 Instrumental, comitative *-ī* — **240**
9.2.5.3 Argument classification — **241**
9.2.5.4 Argument structure changing — **242**
9.2.6 Aspect/aktionsart and time — **242**

9.2.6.1 Overview — **242**
9.2.6.2 Perfect *-ét* — **243**
9.2.6.3 Continuative *-bòm~-bōm* — **243**
9.2.6.4 Perfective2 *-tāng* — **244**
9.2.6.5 Completive *-ròk* — **244**
9.2.6.6 Perfective3 *-lèt* — **245**
9.2.6.7 Durative *-klùng* — **246**
9.2.7 Other functions — **246**
9.2.7.1 Involvement *-dùn~-dūn* 'JOIN' — **246**
9.2.7.2 Formal *-īk* — **248**
9.3 Reduplication — **249**
9.3.1 Habitual — **249**
9.3.2 Argument plurality and iterative — **250**
9.3.3 Intensification — **251**
9.4 Negative *-Cē* (onset reduplication) — **251**
9.5 Post-stem aspect: Exhaustive perfective *-dèt* and experiential *-pín* — **252**
9.5.1 Exhaustive perfective *-dèt* — **252**
9.5.2 Experiential *-pín* — **253**
9.6 Mood: Realis and irrealis — **254**
9.6.1 Realis *-lò* — **254**
9.6.1.1 Overview of verbal functions — **254**
9.6.1.2 Action verbs: The argument against 'past tense' — **255**
9.6.1.3 Copular and property concept term (PCT) verbs: Change of state — **257**
9.6.1.4 Indicating a logical relationship: Cause and result — **258**
9.6.1.5 Correcting a wrong assumption — **260**
9.6.1.6 Summary — **262**
9.6.1.7 Realis *-lò* on nominal predicates and focus *=lo* — **262**
9.6.2 Irrealis marking: *-pò* and *-jí* — **265**
9.6.2.1 Functions common to *-pò* and *-ji* — **266**
9.6.2.1.1 Future marking — **266**
9.6.2.1.2 Habitual marking in procedural texts — **266**
9.6.2.1.3 Hypothetical and counterfactuals — **267**
9.6.2.1.4 Epistemic reading — **269**
9.6.2.1.5 Expressing necessity/obligation — **270**
9.6.2.1.6 Expressing desiderative — **270**
9.6.2.1.7 Subordinate purpose clause marking — **271**
9.6.2.2 Past habitual marking via *-pò* — **271**
9.6.2.3 Summary — **271**
9.7 Subordinating verbal suffixes — **272**
9.7.1 Non-final marker: Realis *-si* and irrealis *-ra* (clause-chaining) — **272**
9.7.2 Non-final *-pen* — **273**

9.7.3 Conditional *-te* — **274**
9.7.4 Marker of complement clauses functioning as indirect questions: *-nē* 'indefinite' — **275**
9.8 Aspect II: *-làng* 'still' — **275**
9.9 Non-declarative speech act suffixes — **276**

10 The noun phrase — 277
10.1 Elements of the Karbi noun phrase — **277**
10.1.1 Overview: Karbi noun phrase structure — **277**
10.1.2 Occurrence of other elements in head noun slot — **278**
10.1.2.1 Conjunctive coordination constructions — **278**
10.1.2.1.1 Juxtaposition — **278**
10.1.2.1.2 With coordinator *=pen* or *lapèn* — **279**
10.1.2.2 Quotative *pu* — **280**
10.1.3 The noun phrase delimiter *abàng* — **281**
10.2 Evidence for Karbi noun phrase structure — **281**
10.3 Diachronic significance of the possessive construction — **287**
10.4 Possessive *a-* marking of head nouns — **288**
10.5 Preposed modifiers — **291**
10.5.1 Demonstratives — **292**
10.5.2 Possessives — **292**
10.5.3 Other preposed modifiers — **294**
10.5.3.1 Interrogative pronouns and derived indefinites — **294**
10.5.3.2 Adverbials — **296**
10.5.3.3 Clausal modifiers — **297**
10.6 Postposed modifier: Plural marking noun *-tūm* — **297**
10.7 Modifiers that occur preposed or postposed — **298**
10.7.1 Deverbal modifiers: PCT-based modifiers and relative clauses — **298**
10.7.1.1 Post-head PCT-based modifiers and pre-head relative clauses — **299**
10.7.1.2 Pre-head PCT-based modifiers (and post-head relative clauses) — **300**
10.7.2 Enumeration constructions — **303**
10.7.2.1 Four basic enumeration constructions — **304**
10.7.2.1.1 Typical classifier construction — **304**
10.7.2.1.2 Self-referential classifier construction — **305**
10.7.2.1.3 Not fully grammaticalized classifier construction — **305**
10.7.2.1.4 Non-classifier (direct enumeration) construction — **306**
10.7.2.2 Pre- vs. post-head order — **307**
10.7.2.3 Anaphoric use of classifiers — **308**
10.7.2.4 Juxtaposition of two numerals or classifier-numeral words to indicate indefiniteness or vagueness — **308**
10.7.2.5 'Another' additive construction — **310**
10.7.2.6 Constructions based on 'one' enumeration — **310**

10.7.2.6.1 Preposed ‘one’ enumeration as an indefinite article construction —— **310**
10.7.2.6.2 Indefinite pronoun construction —— **311**
10.7.2.6.3 Postposed ‘one’ enumeration expressing ‘whole’ —— **311**
10.7.2.6.4 Postposed ‘one’ enumeration expressing ‘same’ —— **311**
10.8 Noun phrase clitics —— **312**
10.8.1 Comitative, instrumental, ablative *=pen* —— **313**
10.8.2 Nominal quantifier constructions based on *=án* ‘this much; all’ —— **315**
10.8.3 Additive, topic, and focus clitics —— **317**
10.8.3.1 Additive *=tā* —— **317**
10.8.3.1.1 Overview —— **317**
10.8.3.1.2 Simple additive ‘also’ —— **317**
10.8.3.1.3 Bisyndetic coordination —— **318**
10.8.3.1.4 Scalar additive ‘even’ —— **318**
10.8.3.1.5 Universal quantification —— **319**
10.8.3.1.6 Intensifier verb construction —— **320**
10.8.3.1.7 Discourse (information structure) function —— **320**
10.8.3.2 Information and discourse structure clitics —— **320**

11 Monoclausal predicate constructions —— 321
11.1 Overview —— **321**
11.1.1 Non-verbal predicates —— **321**
11.1.2 Verbal and nominal predicate negation —— **322**
11.2 Modal and other markers at the monoclausal end of the complementation scale —— **324**
11.2.1 Overview —— **324**
11.2.2 Remarks on the complementation scale in Karbi —— **324**
11.2.3 Morphosyntactic tests for structural properties of modals —— **326**
11.2.3.1 Under scope of nominalization along with main verb root? —— **326**
11.2.3.2 Following adverbial construction $[V]_{\text{main verb}}$ $[pa\text{-}V]_{\text{adv}}$? —— **327**
11.2.3.3 Modified itself by a predicate derivation? —— **328**
11.2.4 Morphophonological evidence —— **329**
11.2.5 The modals —— **329**
11.2.5.1 Deontic *(-)náng* ‘need, must’ —— **329**
11.2.5.2 Non-control *(-)lōng* ‘GET’ —— **331**
11.2.5.2.1 Function —— **331**
11.2.5.2.2 Structure and distribution —— **333**
11.2.5.3 Skillful ability *(-)thèk~(-)thēk* ‘know how’ —— **334**
11.2.5.4 Physical ability *(-)ùn~(-)ūn* ‘be able’ —— **335**
11.2.5.5 Other markers —— **336**
11.3 Adverbial constructions —— **338**
11.3.1 Overview —— **338**
11.3.2 Causative adverbial construction $[V]_{\text{main verb}}$ $[pa\text{-}V]_{\text{adverbial}}$ —— **339**

11.3.3 Nominalization adverbial construction [*ke*-V]$_{\text{main verb}}$ [V]$_{\text{adverbial}}$ — **340**
11.3.4 Non-final preposed adverbial constructions — **341**
11.3.4.1 Non-final *-pen* construction [*ke*-V-*pen*]$_{\text{adverbial}}$ [V]$_{\text{main verb}}$ — **341**
11.3.4.2 Non-PCT root construction [*ke*-V-*si*]$_{\text{adverbial}}$ [V]$_{\text{main verb}}$ — **342**
11.4 Periphrastic constructions based on copulas — **344**
11.4.1 Progressive construction with non-final suffix *-si* plus copula *dō* — **344**
11.4.2 Copula argument quantification construction — **345**
11.5 Complex motion construction — **346**
11.6 'Noun plus verb' predicate constructions — **347**
11.6.1 Non-possessed noun incorporation — **347**
11.6.2 Psycho-collocations and possessed noun incorporation — **349**
11.6.3 Light verb construction — **352**
11.6.4 Cognate object construction — **353**
11.6.5 Hybrid construction — **355**
11.7 Other complex predicate constructions discussed elsewhere — **355**

12 Nominalization — 356
12.1 Derivational nominalization — **356**
12.2 Property concept term (PCT)-based noun modification — **358**
12.3 Relativization — **360**
12.3.1 Standard (externally-headed, pre-head) relativization — **360**
12.3.1.1 Relativization on different clause participants — **360**
12.3.1.2 Irrealis-marked relative clauses — **364**
12.3.1.3 Head noun occurring with personal possessive prefix — **365**
12.3.2 Internally-headed (or post-head) relativization — **365**
12.4 Complementation — **368**
12.4.1 Standard complementation — **368**
12.4.2 Irrealis-marked complement clauses — **371**
12.4.2.1 Irrealis-marked complement clauses with purpose marker *aphān* — **371**
12.4.2.2 Irrealis-marked complement clause with noun phrase delimiter *abàng* — **372**
12.4.3 Functional types of complement-taking verbs — **373**
12.5 Adverbial subordination — **373**
12.5.1 Nominalized adverbial subordination: Subordinators from relator nouns — **373**
12.5.2 Semantic types — **374**
12.5.3 Adverbial subordination constructions with additional marking — **381**
12.6 Irrealis-marked nominalized subordinate clauses — **381**
12.7 Main clause constructions — **384**
12.7.1 Nominalization plus existential copula construction — **384**
12.7.2 Adverbial constructions — **388**

12.7.3 Diachronic nominalization constructions in main clause grammar — **388**
12.7.3.1 Focus constructions — **388**
12.7.3.1.1 Inconsistent occurrence of *ke-* — **389**
12.7.3.1.2 General argument focus construction — **389**
12.7.3.1.3 Content question focus construction — **390**
12.7.3.1.4 Co-relative focus construction — **391**
12.7.3.1.5 Historical development — **392**
12.7.3.2 Imperfective construction — **393**
12.7.3.3 Ambiguity between focus and imperfective interpretation — **396**
12.8 Inconsistent occurrence of *ke-* 'nominalizer' in nominalization constructions — **397**

13 Clause participants: Overview, participants, noun phrase delimiter — 401
13.1 Preliminaries — **401**
13.1.1 Terminology — **401**
13.1.2 The argument-oblique continuum and the syntax, semantics, and pragmatics in role marking — **402**
13.2 The predicate: Defining argument roles — **403**
13.2.1 Argument roles in typical declarative clauses — **403**
13.2.1.1 Monovalent predicate: Unmarked S argument — **404**
13.2.1.2 Bivalent predicates: Marking of A and O arguments — **405**
13.2.1.3 Trivalent predicates: Marking of A, R, and T arguments — **406**
13.2.1.3.1 R-marked trivalent construction — **406**
13.2.1.3.2 T-marked trivalent constructions — **408**
13.2.1.3.3 Unmarked trivalent construction — **411**
13.2.1.3.4 Summary: Trivalent constructions — **412**
13.2.1.4 Alignment in typical declarative clauses — **413**
13.2.2 Declarative clause constructions with non-typical role marking of arguments — **414**
13.2.2.1 Motion constructions with unmarked and *-lòng* marked goals — **414**
13.2.2.2 'Need' construction — **416**
13.2.2.3 Existential copula constructions: Simple locative and possessive constructions — **417**
13.2.2.3.1 Simple locative construction — **417**
13.2.2.3.2 Predicational possession construction with unmarked A/possessor and possessive-marked O/possessed — **418**
13.2.2.3.3 Predicational possession construction with locative-marked A and unmarked O argument — **420**
13.2.2.4 Topical possessor construction — **420**
13.2.2.5 Comparative constructions — **422**
13.2.3 Predicates with derivationally changed argument structure — **423**

13.2.3.1 Causative *pe-~pa-* — **423**
13.2.3.2 Benefactive/malefactive (affective) *-pī* — **425**
13.2.4 Non-subject speech act participant indexing on the verb — **427**
13.2.5 Other grammatical relations constructions — **428**
13.3 Overview: Clause participant marking — **429**
13.4 Lexical noun phrase versus pronoun versus zero anaphora — **429**
13.4.1 Lexical noun phrase — **430**
13.4.2 Pronoun — **431**
13.4.3 Zero anaphora — **432**
13.5 Noun phrase delimiter *abàng* — **433**
13.5.1 Distribution of noun phrase delimiter *abàng* — **435**
13.5.2 Diachronic source construction: Human classifier as syntactic head noun — **437**
13.5.3 Coocurrence of *abàng* with role markers — **438**
13.5.4 Co-occurrence of *abàng* with core information/discourse structure clitics — **440**
13.5.5 Differences between speakers, texts, genres — **441**

14 Role marking and core information/discourse structure marking — 444
14.1 Role marking — **444**
14.1.1 Unmarked noun phrases — **445**
14.1.1.1 S argument — **445**
14.1.1.2 A argument — **446**
14.1.1.3 O-low argument — **446**
14.1.1.4 T argument — **447**
14.1.1.5 R argument (T-marked trivalent construction) — **447**
14.1.1.6 Goal/locative argument/participant of motion verbs — **448**
14.1.1.7 Other types of participants — **448**
14.1.2 Functions of 'non-subject' *-phān* — **450**
14.1.2.1 O-high argument — **450**
14.1.2.2 R argument (R-marked trivalent construction) — **451**
14.1.2.3 T argument (T-marked trivalent construction) — **452**
14.1.2.4 Semantic marking with *náng* 'need' — **452**
14.1.2.5 Marking the standard of comparison — **452**
14.1.2.6 Other types of participants — **453**
14.1.3 Functions of 'locative' *-lòng* — **455**
14.1.3.1 Locative/goal R argument (T-marked/unmarked trivalent constructions) — **455**
14.1.3.2 Locative O-like argument — **456**
14.1.3.3 Motion verb locative/goal — **457**
14.1.3.4 Semantic marking with possessor construction — **458**
14.1.4 Semantically marked participants — **458**

14.1.4.1 Semantically specific marking of functionally core roles — **458**
14.1.4.1.1 Comitative, instrumental, ablative *=pen* — **459**
14.1.4.1.2 Goal arguments marked with semantically specific relator nouns — **460**
14.1.4.2 Semantically specific relator noun marking of obliques — **461**
14.1.5 Differential marking — **461**
14.1.6 Marking variation — **462**
14.1.6.1 *che-tòng* 'RR-meet' — **462**
14.1.6.2 *arjū-dām* 'ask-GO' — **463**
14.2 Core information/discourse structure marking — **464**
14.2.1 Topic *=ke* — **465**
14.2.2 Additive *=tā* — **468**
14.2.3 Realis focus *=si* — **471**
14.2.4 Irrealis focus *=le* — **474**
14.2.5 Relationships between core information/discourse structure clitics — **476**
14.2.6 Other information/discourse status constructions — **479**
14.2.6.1 Constituent order — **479**
14.2.6.2 New participant marking — **482**
14.2.6.3 Restrictive focus markers — **482**

15 Clause types and clause combining — 486
15.1 Non-declarative main clause types — **486**
15.1.1 Interrogatives — **486**
15.1.1.1 Question particle *=ma* — **487**
15.1.1.2 Content questions with interrogative pronouns and adverbs — **487**
15.1.1.2.1 Verb occurring without *=ma* — **487**
15.1.1.2.2 Verb occurring with *=ma* — **488**
15.1.1.3 Polar interrogatives and disjunctive interrogatives — **489**
15.1.1.4 Polar interrogatives — **489**
15.1.1.4.1 Polar interrogatives with *=ma* — **489**
15.1.1.4.2 Polar interrogatives with prosody only — **491**
15.1.1.5 Disjunctive interrogatives — **493**
15.1.1.5.1 Type 1: A=*ma* A-NEG — **493**
15.1.1.5.2 Type 2: A=*ma* B — **493**
15.1.1.5.3 Type 3: A=*ma* B=*ma* — **494**
15.1.1.5.4 Type 4: A=*ma ma* B=*ma* — **494**
15.1.1.6 Interrogative assumption *=bo* — **495**
15.1.1.7 Tag question *dī* — **496**
15.1.1.8 Feedback request with *déi* — **496**
15.1.2 Imperatives and prohibitives — **497**
15.1.2.1 Bare stem imperative — **497**

15.1.2.2 Informal conditioned imperative *-nōi* — **498**
15.1.2.3 Conditioned imperative *-nōn* — **499**
15.1.2.4 Conative imperative *-thā* — **499**
15.1.2.5 Unconditioned imperative *-tū* — **500**
15.1.2.6 Imperatives *-nōn*, *-thā*, and *-tū* and Grüßner's (1978) account of politeness differences — **501**
15.1.2.7 Prohibitive *-rī* — **501**
15.1.2.8 Prohibitive construction via combination of prohibitive and imperative suffix — **501**
15.1.3 Hortatives — **502**
15.1.3.1 General hortative *-nāng* — **502**
15.1.3.2 Jussive construction with causative *pa-* and hortative *-nāng* — **504**
15.1.3.3 Extended forms: Emphatic hortative *-lonāng* and conative hortative *-sināng* — **505**
15.2 Non-nominalized subordinate clause types — **507**
15.2.1 Non-final clauses in clause chains — **507**
15.2.1.1 Morphologically marked clauses: *-si* 'non-final:realis', *-ra* 'non-final:irrealis', *-pen* 'non-final:with' — **507**
15.2.1.2 Prosodically marked clauses — **510**
15.2.2 Complement clauses — **512**
15.2.2.1 Verb juxtaposition — **512**
15.2.2.2 Indirect questions — **513**
15.2.2.3 Topic *=ke* marked complement clauses — **514**
15.2.2.4 Quotative *pu* and *pusi* complementizers — **514**
15.2.3 Adverbial clauses — **516**
15.2.3.1 Conditional *-te* — **516**
15.2.3.2 Purpose clauses with quotative complementizers — **517**
15.2.3.3 Concessive *sitā~setā* — **517**
15.3 Irrealis clause types: Irrealis-sensitivity in non-final and focus markers — **518**
15.3.1 Overview — **518**
15.3.2 Non-declarative speech acts — **519**
15.3.3 Negation — **521**
15.3.4 Deontic clauses (expressing necessity/obligation; with *náng* 'need, must') — **522**
15.3.5 Conditional subordinate clauses — **523**
15.4 Non-nominalized insubordination (formally non-finite declarative main clause types) — **523**
15.4.1 Main clauses marked with *=ke* 'topic': Background information construction — **523**
15.4.2 Main clauses marked with *pu* 'quotative': Desiderative construction — **524**

15.4.3 Stand-alone indirect questions — **526**
15.5 Clause coordination — **527**
15.5.1 Conjunctive coordination — **527**
15.5.1.1 Conjunctive coordinator *lapèn* 'and' — **527**
15.5.1.2 Additive particle clause/VP coordination constructions — **527**
15.5.1.3 Clausal NP coordination — **528**
15.5.2 Disjunctive coordination — **529**
15.5.3 Adversative coordination — **530**
15.6 Lack of a syntactic pivot — **531**

16 Discourse constructions — 532
16.1 Discourse structuring constructions — **532**
16.1.1 Clause parallelism — **532**
16.1.2 Tail-head linkage — **534**
16.1.3 Discourse connectors — **536**
16.1.4 Discourse structuring markers e and *'mh* — **537**
16.2 Rhetorical constructions — **539**
16.2.1 General extender constructions — **539**
16.2.2 Elaborate expression constructions — **543**
16.2.2.1 Forms of elaborate expressions — **543**
16.2.2.2 Embedding into parallelism or compound construction — **547**
16.2.2.3 Functions of elaborate expressions — **549**
16.2.3 Copy verb constructions — **550**
16.2.3.1 Assertive (with *=ke* 'topic') — **551**
16.2.3.2 Intensifier declarative (with *=tā* 'additive') — **553**
16.2.3.3 Intensifier non-declarative (with *=le* 'focus:irrealis) — **554**
16.2.3.4 Perseverance construction (with *=ma* 'question particle') — **555**
16.2.4 Constructions with negative equational copula *kalī* — **556**
16.2.4.1 Disagreement construction based on quasi-reduplication and negative equational copula — **556**
16.2.4.2 Nominalization-based intensifier construction *ke*-V-*sō kalī* — **557**
16.2.5 Prosodic emphasis — **558**
16.3 Particles — **558**
16.3.1 Quotative *pu* — **560**
16.3.1.1 Reportative function of *pu* — **560**
16.3.1.2 Desiderative function of *pu* — **561**
16.3.2 Reportative *tànghò* — **562**
16.3.3 Dubitatives *bón, tahái, menē* — **564**
16.3.4 Always *titī* — **565**
16.3.5 Emphatic *ti* — **566**
16.3.6 Interactive emphatic *ho* — **569**

16.3.7 Vocative ó — **570**
16.3.8 Exclamative function of irrealis focus *=le* — **571**
16.3.9 Afterthought *=he* — **572**
16.3.10 Background information: *=ke* 'topic' marked main clauses — **574**
16.3.11 Common ground marker *=mati* — **577**
16.3.12 Narrative style *hedī* — **578**
16.4 Honorific and formality marking — **579**
16.4.1 Honorific *-lī* on pronouns and addressing words — **580**
16.4.2 Formal *-īk* on predicates — **580**
16.4.3 Honorific *-héi ~ -hái* on kinship terms — **582**
16.5 Interjections — **582**
16.6 Hesitation and correction words — **584**
16.6.1 Hesitation words *kenē* and *mane* (<Assamese) — **584**
16.6.2 Correction words *chē* and *bá* (<Assamese) — **586**

Appendix A
Abbreviations — **589**

Appendix B
Folk story: Chonghokaloso lapen Misorongpo (RBT, ChM) — **591**

Appendix C
Stimuli-based narrative: Pear Story (SiT, PS) — **610**

Appendix D
Metadata — **622**

Appendix E
Glossary — **626**

Bibliography — 647

Index Karbi grammar — 657

Author Index — 665

List of Figures

Figure 1 Northeast India (taken from openstreetmap.org, “Myanmar” and “Karbi Anglong” labels added, accessed February 28, 2020) —— **3**

Figure 2 Approximate outline of the Karbi-speaking area (based on a map taken from openstreetmap.org, accessed on February 28, 2020; outline and names for countries as well as Indian states added) —— **5**

Figure 3 Priest performing a duck sacrifice —— **6**

Figure 4 Jambili athon —— **6**

Figure 5 Classification of Tibeto-Burman according to Bradley (2002) —— **10**

Figure 6 Burling’s (2003: 184) ‘Relationships among the languages of the Eastern Border’ —— **10**

Figure 7 Pnar, Khasi, War, and Lyngam speaking areas (Daladier 2014) —— **15**

Figure 8 Standard Karbi vowel monophthongs —— **55**

Figure 9 Amri Karbi vowels —— **57**

Figure 10 Semantic map of basic clausal functions of Karbi parts of speech (following Croft’s (2001: 99) model) —— **116**

Figure 11 Position classes in the Karbi verb —— **197**

Figure 12 Overview of cislocative functions of *nang=* —— **211**

Figure 13 Post-root slots of the Karbi verb —— **223**

Figure 14 Karbi noun phrase structure —— **277**

Figure 15 Noun phrase clitic slots —— **312**

Figure 16 Grammaticalization scenario for *=si* ‘focus’ (starting as copula) —— **392**

Figure 17 Possible grammaticalization pathway for the imperfective construction —— **396**

Figure 18 Schematic overview of argument expression in Karbi —— **430**

Figure 19 Role marking possibilities —— **444**

Figure 20 Waveform and spectrogram of interrogative and subsequent declarative *enutvetlo* ‘he was alone’ (HK, TR 005-6) —— **492**

Figure 21 F_0 contour of interrogative and subsequent declarative *enutvetlo* ‘he was alone’ (HK, TR 005-6) —— **492**

Figure 22 Prosodic emphasis in full predicate reduplication of *baithekthe* (KK, CC 012) —— **559**

List of Tables

Table 1 Locations of Karbi-speaking villages outside Karbi Anglong —— **4**
Table 2 Vowel alternations —— **27**
Table 3 Subcorpora of Karbi data —— **38**
Table 4 Text genres —— **42**
Table 5 Syllable-initial consonants —— **49**
Table 6 Minimal set for bilabial stop onsets —— **51**
Table 7 Minimal set for alveolar stop onsets —— **51**
Table 8 Minimal set for palatal stop onsets —— **51**
Table 9 Minimal set for velar stop onsets —— **52**
Table 10 Minimal set for fricative onsets —— **52**
Table 11 Minimal set for nasal onsets —— **52**
Table 12 Minimal set for liquid and glide onsets —— **52**
Table 13 Syllable-final consonants —— **54**
Table 14 Minimal sets for stop codas —— **54**
Table 15 Minimal sets for sonorant codas —— **54**
Table 16 Minimal sets for vowels in open syllables —— **56**
Table 17 Minimal sets /ei/, /ai/, /oi/, /ui/ —— **56**
Table 18 Set of items with /ɪ/ in Amri Karbi —— **57**
Table 19 Marginal vowels —— **58**
Table 20 Syllable types —— **59**
Table 21 Onset cluster types —— **60**
Table 22 Minimal sets for tones after voiceless onsets —— **63**
Table 23 Minimal sets for tones after voiced onsets —— **63**
Table 24 Additional stopped syllable tone minimal triplet —— **64**
Table 25 Stopped syllable tone minimal pairs for M-H, L-H, and L-M —— **64**
Table 26 Minimal triplets of disyllabic roots —— **65**
Table 27 Minimal pairs of disyllabic roots —— **66**
Table 28 Tone sandhi effects on low tone suffix after mid tone stem—— **72**
Table 29 Internal sandhi effects on stem induced by prefix *cho-* 'AUTO.BEN/MAL' —— **73**
Table 30 Grüßner's (1978) account of tone assignment to toneless syllables —— **74**
Table 31 Compound tones —— **75**
Table 32 Tone representation on unstressed syllables in disyllabic morphemes (default and non-default) —— **78**
Table 33 Verb-Noun tone minimal pairs (G = Grüßner 1978) —— **79**
Table 34 Tone minimal pairs in semantically related verbs according to Grüßner (1978: 91) —— **79**
Table 35 Karbi prefixes —— **83**
Table 36 Karbi clitics —— **84**
Table 37 Full reduplication of roots or suffixes without vowel change —— **85**
Table 38 Vowel change patterns in quasi-reduplicative constructions —— **86**
Table 39 Forms of quasi-reduplicative *-Cē* 'NEG' with different stems —— **87**
Table 40 Monosyllabic stem tone changes after *che-* 'RR' and *cho-* 'AUTO.BEN/MAL' —— **88**
Table 41 Disyllabic stem tone changes following possessive prefixation —— **89**
Table 42 Tone change patterns for derivational suffixes —— **90**
Table 43 List of derivational suffixes that participate in tone change —— **90**
Table 44 Idiosyncratic tone allomorphy of *-vèt* ~ *-vét* 'only' —— **92**
Table 45 Forms of prefix *ke-* 'NMLZ' —— **93**

Table 46 Forms of prefix *pV-* 'causative' —— **93**
Table 47 Forms of prefix *che-* 'reflexive/reciprocal' —— **94**
Table 48 Initial vowel deletion in *ing-* stems after prefixes *ke-*, *pV-*, *cho-* —— **94**
Table 49 Vowel deletion between *ar-* stems and prefixes *ke-*, *pV-*, *che-*, *a-* —— **95**
Table 50 Morphosyntactic criteria for the attempt to identify a class of 'adjectives' —— **103**
Table 51 Basic property concept terms: Karbi sample roots —— **104**
Table 52 Sample complex stems expressing concepts of HUMAN PROPENSITY —— **105**
Table 53 Possible morphosyntactic criteria for 'adjectives' —— **111**
Table 54 Marginal types of property concept term —— **112**
Table 55 Classifier types —— **118**
Table 56 List of sortal classifiers that occur in typical classifier construction (G = Grüßner (1978: 68–70)) —— **119**
Table 57 Not fully grammaticalized classifier *bàng* 'CLF:HUM:PL' —— **121**
Table 58 List of mensural classifiers (G = Grüßner (1978: 68–70)) —— **123**
Table 59 Not fully grammaticalized mensural classifiers —— **124**
Table 60 Self-referential classifiers —— **124**
Table 61 Nouns counted with numeral only (without classifier) —— **128**
Table 62 Relator nouns —— **130**
Table 63 Overview of locational relator nouns —— **136**
Table 64 Locational/temporal relator nouns —— **139**
Table 65 Relator nouns with other functions —— **142**
Table 66 Personal pronouns and personal possessive prefixes —— **150**
Table 67 Morphological structure of independent pronouns and pronominal possessive prefixes —— **151**
Table 68 Demonstratives —— **154**
Table 69 Words with demonstrative roots —— **156**
Table 70 Interrogative pronouns and adverbs —— **157**
Table 71 Corresponding demonstratives and interrogatives —— **161**
Table 72 Pronouns of universal quantification —— **161**
Table 73 Day ordinals —— **171**
Table 74 Year and day ordinals —— **171**
Table 75 Numerals from 'one' to 'ten' —— **174**
Table 76 Numerals over 'ten' —— **175**
Table 77 Subordinators grammaticalized from relator nouns —— **176**
Table 78 Coordinators —— **177**
Table 79 Noun-noun and noun-verb compounds —— **180**
Table 80 Personal possessive prefixes —— **185**
Table 81 Proper nouns with *-pī*, *-pō*, or *-sō* suffix —— **190**
Table 82 Cross-referencing 'paradigm' of *nang=* —— **201**
Table 83 Proclitic cross-referencing 'paradigm' —— **206**
Table 84 Discontinuous predicate derivations —— **227**
Table 85 Sample highly productive PDs —— **228**
Table 86 Sample PDs that mean 'quite' (productive and non-productive PDs) —— **228**
Table 87 Non-productive intensifier PDs —— **229**
Table 88 Some predicate derivations and related lexical items —— **229**
Table 89 Sample non-ideophonic manner predicate derivations —— **230**
Table 90 Sample ideophonic manner derivations —— **231**
Table 91 Sample PDs that indicate a considerable degree ('quite') (productive and non-productive PDs) —— **231**

Table 92 Non-productive intensifier PDs — **232**
Table 93 PDs indicating argument quantification and degree or extent — **232**
Table 94 Sample result derivations — **233**
Table 95 Direction, motion, and path derivations — **234**
Table 96 Argument quantification derivations — **238**
Table 97 Sample predicate derivations that function as argument classifiers — **241**
Table 98 Aspect/aktionsart and time derivations — **243**
Table 99 Non-declarative speech act suffixes — **276**
Table 100 NP structure — **282**
Table 101 Enumeration constructions — **304**
Table 102 Examples of enumeration constructions — **304**
Table 103 Proper adverbial constructions (AdvCs) — **338**
Table 104 Non-final adverbial constructions (NF-AdvCs) — **338**
Table 105 Noun incorporation — **348**
Table 106 Sample psycho-collocations — **350**
Table 107 Possessed noun incorporation expressions (non-psycho-collocations) — **351**
Table 108 Semantic types of nominalized adverbial subordinate clauses — **374**
Table 109 Object marking in trivalent constructions — **406**
Table 110 Co-occurrence of *abàng* with core information/discourse structure clitics — **440**
Table 111 Occurrence of *abàng* across texts — **441**
Table 112 Frequencies of occurrence of core information/discourse structure clitics — **465**
Table 113 Distribution of restrictive focus markers — **482**
Table 114 Functions of restrictive focus markers — **483**
Table 115 Disjunctive interrogative types — **493**
Table 116 Irrealis contexts for *-ra* 'NF:IRR' and *=le* 'FOC:IRR' — **519**
Table 117 Discourse connectors — **536**
Table 118 Formal patterns in elaborate expressions — **545**
Table 119 Interjections — **582**
Table 120 Metadata of texts representing the main corpus for this grammar — **622**
Table 121 Brief descriptions of texts that have formed the main corpus for this grammar — **623**
Table 122 Metadata of speakers of texts that have formed the main corpus for this grammar — **625**
Table 123 Word classes and grammatical morpheme types indicated in the glossary — **626**

1 Introduction

This is a grammar of Karbi as spoken in the hills of the Karbi Anglong district in Assam, Northeast India. It expands on research findings reported in Grüßner's (1978) grammar of the phonology and morphology of the language, but also offers a more comprehensive treatment of issues in Karbi syntax, semantics, and pragmatics.

This chapter is organized as follows. In §1.1, the Karbi people and their language and culture are introduced. Next, an overview of the linguistic context of Karbi is offered, i.e., the relationships between Karbi and surrounding languages, as that context helps understand why modern Karbi grammar is the way it is. On the one hand, that involves the relationship to other Tibeto-Burman languages, as discussed in §1.2. On the other hand, and importantly, it also involves the contact relationship specifically with the Austroasiatic Khasi languages, and possibly other Tibeto-Burman languages, as discussed in §1.3.

In §1.4, an overview of what is known on different varieties of Karbi is provided, including the major dialectal divide between Hills and Plains (or, Amri) Karbi, and some notes on variation within Hills Karbi, the major dialect group that this grammar is based on. A sociolinguistic profile of language endangerment (mostly based on Hills Karbi) is included in §1.5. In §1.6, an overview of the linguistic literature and linguistic resources on Karbi is provided.

Finally, §1.7 discusses Karbi orthography and ongoing issues in standardization efforts as well as an outline of conventions followed in this grammar. The organization of this grammar is outlined in §1.8.

1.1 Karbi people, language, and culture

1.1.1 Names and ISO codes for the Karbi language

In the last few decades, there has been a movement among the Karbis to push for the autonym *Karbi* or the elaborate form *Karbi Karbak* (see §16.2.2 on elaborate expressions). While this name has long been in use, it is a recent development that *Karbi* is favored over the logonym *Arleng* (i.e., *arlēng* 'man, person').[1] This might be due to the existence of *arlēng* as a simple noun root for 'man, person'.[2]

1 In one of the recorded texts collected for the corpus of this grammar, the storyteller finds himself saying *Arlengpi* for 'Karbi woman' (using the female *-pī* suffix) and corrects himself and says *Karbipi*.
2 Note, however, that there also is another general noun *monít* 'person, man', which is a borrowing from Asamese.

Mikir is a formerly commonly used exonym, which has become pejorative within the last few decades in particular. Now most Karbis have strong objections against it, which has to do with a number of offensive hypotheses for the etymology of this name. *Bhoi Mynri* is mentioned by Grüßner (1978: 6) as an exonym used by the neighboring Khasis of Meghalaya, to the immediate west of Karbi Anglong. According to my language consultants, this term *Bhoi Mynri* may also specifically refer to the variety of Plains Karbi spoken across the western border of Assam in Meghalaya (§ 1.4.1).

The Hills Karbi variety has the ISO 639-3 code 'mjw', whereas the Plains Karbi variety has the ISO code 'ajz' (for dialect differences, see § 1.4.1).

1.1.2 Number of speakers and geographical spread of Karbi

The Census of India from 2001 reports a total of 419,534 native speakers of Karbi, which is also the figure cited in the Ethnologue (Simons and Fennig 2017). The *Karbi Lammet Amei* (§ 1.1.4) estimates a higher number of speakers, at over half a million. Karbi is the third-largest minority language in the state of Assam in terms of number of speakers, following Boro and Mising.

Karbi is spoken in Assam and adjacent areas in neighboring states in Northeast India. For a map of the Eastern Himalayan region, including Northeast India and surrounding countries, see Figure 1.

Today most Karbi speakers live in the Karbi Anglong and West Karbi Anglong districts of Assam, which until 2015 formed a single Karbi Anglong district (see § 1.1.5). However, the geographic spread of Karbi speaking villages is much larger. Table 1 provides the locations of Karbi speaking villages outside the Karbi Anglong district.

Based on the locations given in Table 1, we can plot the approximate outline of the Karbi speaking area as done in Figure 2.

As the topographical map in Figure 2 shows, the Karbi speaking area extends from the valley of the river banks of the Brahmaputra southwards across plains and low to moderate hills into the Barak Valley around Silchar.

The present-day core region of the Karbi speaking area are the Karbi Anglong and West Karbi Anglong districts. These districts are located in the lower hills that mark the transition between the Brahmaputra valley area and the hill range that extends to the south and the southeast.[3]

3 This is part of the hill range that extends all the way into Southeast Asia and represents something of a cultural area, with similar histories of the people inhabiting them, see Scott (2009).

Figure 1: Northeast India (taken from openstreetmap.org, "Myanmar" and "Karbi Anglong" labels added, accessed February 28, 2020).

1.1.3 Aspects of Karbi culture

1.1.3.1 Traditional culture and social organization

Changes in the lifestyle of the Karbis are occurring at an exponentially increasing pace in recent years. Due to urbanization and increased physical and virtual infrastructure, elements of the traditional culture are becoming both more endangered as well as newly embraced and cherished. Traditional village life involving *jhum* cultivation and collecting wild vegetables and fruit in the jungles and forests is becoming more and more confined to remote places that are not connected with physical infrastructure.

While a substantial number of Karbis have nowadays converted to Hinduism or Christianity, the traditional religion of the Karbis is still practiced by a considerable portion of the population. It involves different gods and goddesses, but also has a strong animist element. In cases of major life events such as weddings or deaths, as

Table 1: Locations of Karbi-speaking villages outside Karbi Anglong.

State	District	# of villages
Assam	Golaghat	?
	Marigaon	?
	Biswanath	19
	Lakhimpur	5
	Nagaon	18
	Hojai	18
	Kamrup (mostly Dispur LAC, but also Guwahati East LAC and Guwahati West LAC)	180
	NC Hills	64
	Kachar Plains	?
Meghalaya	Ri-Bhoi and West Jaintia Hills	58
Arunachal Pradesh	Papum Pare	5

For the information in this table, I am indebted to Ajit Kathar, who provided a list of villages in Kamrup; to Manik Rongpi, a student at Tezpur University, who contributed lists of Karbi speaking villages in the Biswanath and Lakhimpur districts of Assam as well as in Arunachal Pradesh; to Keson Klein of Marmein in Meghalaya for a list of villages in Meghalaya; and to Joysing Ronghi of Umrongso, Dima Hasao, for a list of villages in the NC Hills. And, most importantly, many thanks to Sikari Tisso for contacting them and others and collecting all this information.

well as other crucial times such as sickness or before going on a long trip, priests perform rituals that typically involve sacrifice of animals (such as chickens or ducks) in conjunction with chants that are orally transmitted from generation to generation, typically using the Karbi song language (§ 1.1.3.3), see Figure 3.

An important cultural symbol is the Jambili athon (Figure 4). The bird on top represents values such as wisdom, intellectuality, and leadership. The lower four birds in the four directions are the followers.

There are five major clans in Karbi society: Terang, Teron, Inghi (also spelled Enghi or Enghee), Ingti (also spelled Engti), and Timung.[4] These five major clans are further divided into subclans. This division into clans and subclans has important societal consequences such as marriage restrictions.

An excellent resource on Karbi cultural studies are the two volumes 'Karbi Studies'. The first volume is edited by Dharamsing Teron, with contributions both from Karbi

4 These are the clan names in the Hills Karbi variety; in Plains Karbi, some names are slightly different, e.g., *Timung* is *Tumung* (see § 1.4.1 on dialect differences between what I refer to as Plains and Hills Karbi).

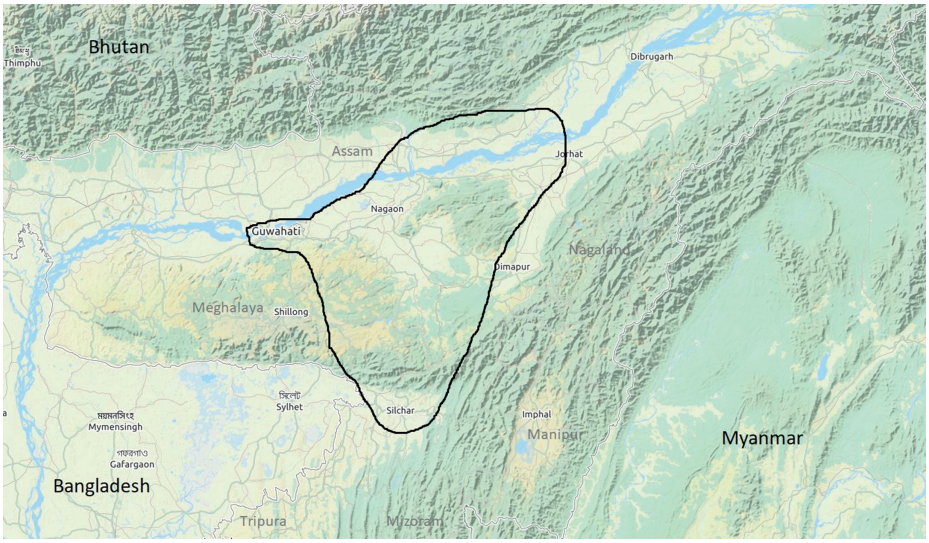

Figure 2: Approximate outline of the Karbi-speaking area (based on a map taken from openstreetmap.org, accessed on February 28, 2020; outline and names for countries as well as Indian states added).

and international scholars (Teron 2012), while the second volume is entirely authored by Teron (Teron 2011). Further information can also be obtained from a blog originally created by Morningkeey Phangcho although no new posts have appeared since 2009 (http://karbi.wordpress.com/).

1.1.3.2 Oral literature

The orally transmitted traditional literature of the Karbis is a fundamental part of Karbi culture. As part of data collection for this grammar, a number of folk stories were recorded that tap into this rich treasure of Karbi oral literature. While the stories are always narrated in the ordinary language, there are a lot of songs (mostly ballads that tell a particular story) as well as (religious) chants which are sung using the song language (see § 1.1.3.3 below). However, also the ordinary language used to tell folk stories has elements specific to the genre, see § 16.1.4 on the discourse structuring markers *e* and *'mh*, as well as § 16.3.12 on the 'narrative style marker' *hedī*, in particular.

Karbi oral literature shares many types of folk stories with other ethnic groups in Northeast India. A typical genre is folk stories about the origins of subclans, such as the story about the three Bey brothers (Konnerth and Tisso 2018: 334–48). Typically, these stories offer a (mythological) explanation of how the division into subclans among members of a particular clan or subclan came about, and they often also contain societal rules such as restrictions on (everyday life) interactions between members of particular subclans.

Figure 3: Priest performing a duck sacrifice.

Figure 4: Jambili athon.

Second, an apparently common story in the context of Northeast India that exists in Karbi oral literature as well is the story *Miso-rongpo lapen Chongho-kaloso* (see Appendix B). This folk story starts with a fight between a frog and an ant (although it might involve other animals in the traditions of other language communities), resulting in a chain reaction of events, in which one animal suffers from being disturbed or hurt by another animal, and as a consequence accidentally disturbs or hurts another animal, and so on. Examples from what appears to be the same basic story in Khumi (South-Central or 'Kuki-Chin'[5], spoken in the Bangladesh/Burma border area) are used in a discussion of elaborate expressions in Khumi by Peterson (2010: 96–7), and I have come across stories with the same basic structure in other South-Central languages as well, such as Monsang, Lamkang, and Thadou.

Another narrative that is characteristic of the region (specifically the hill region stretching from Northeast India across Southeast Asia) concerns the loss of an allegedly previously existing script. This narrative is analyzed by Scott (2009) as a literary-mythological account of an intentional decision by these peoples for an oral literary tradition and against a written tradition. He makes this argument in the context of his larger hypothesis that the hills peoples of Southeast Asia[6] have a history of intentionally fleeing the developing civilizations in the valleys (which were heavily built on slavery in their early beginnings) in order to maintain (cultural and political) independence and societal equality. In Karbi, similar to other languages of the region, the lost script narrative tells that the only record of the script was on a deer hide, which in a time of starvation had to be eaten in order to survive, and was therefore lost.

1.1.3.3 Song language

The Karbi song language is used for oral literature that is sung or chanted (hence the name) rather than narrated. It is also referred to as the poetical language. According to my language consultants, between the two major dialects of Hills and Plains Karbi, there is an interesting relationship between song language and ordinary language such that Hills Karbi song language words are ordinary language words in Plains Karbi and vice versa.[7]

---

5 'Kuki-Chin' is referred to here in single quotes since this label is offensive to speakers of some of the languages included within it, such as Monsang, Moyon, Lamkang, Anal from the state of Manipur in Northeast India. I am also using the name "South-Central" here, following the considerations mentioned in Konnerth (2018: 19).

6 That is, the peoples inhabiting the hill range that stretches across Southeast Asia, which Scott (2009) refers to as *Zomia*, with the claim that that is not only a geographic label but also needs to be understood as an area of a shared cultural-political history.

7 Note that it is not common for members of the Karbi language community to understand all the song language words. They typically know a few individual words, but no more than that.

Many lexical items that occur in the Hills Karbi song language represent borrowings from Khasi languages (with which there exists a history of contact, §1.3). For example, the word *um* is used for 'water' in the song language, which is a common component of toponyms in West Karbi Anglong in names such as Umswai, Umlapher, Umkachi (or Amkachi), etc. The song language is thus an important object for further study in order to trace Khasi borrowings in Karbi. A first move in this direction is the book *Karbi lamlir achili* (lit., 'the seeds of the Karbi poetical language'), a collection of Hills Karbi song language words (some of which with context in songs and chants in which they are used) edited by eminent Karbi language and literature scholar Longkam Teron (Teron 2008). It is furthermore a topic for future study to investigate the grammatical structure of song language texts.

Note that an interesting aspect in the transition from traditional to modern culture is that the song language is also used in modern (Indic, Bollywood-style) Karbi pop songs. However, Christian songs do not make use of the song language but of the ordinary language.[8]

1.1.4 The Karbi Lammet Amei (KLA)

This grammar is the result of close collaboration with members of the Karbi Lammet Amei, who in fact initiated the project in 2007: most notably Mr. Sikari Tisso, as well as Mr. Khor Sing Teron. The Karbi Lammet Amei (KLA; from *Karbì lám-mét a-méi* 'Karbi word-artful POSS-assembly') is a language and literature organization based in the Karbi Anglong district capital Diphu, but with branches in larger villages and towns. The KLA was founded on March 27, 1966, with the goal of preserving and promoting the Karbi language so it could be taught in schools and other institutions of higher education, while also engaging in the promotion of Karbi literature (Khor Sing Teron, p.c.).

1.1.5 The Karbi Anglong and West Karbi Anglong districts

The Karbis have had their own autonomous Karbi Anglong district (i.e., *Karbi a-inglóng* 'Karbi POSS-hill' > 'Karbi hills') for a number of decades. The district was first formed in 1951, although at that time, the North Cachar Hills region to the south still belonged to the district (then called 'United Mikir and North Cachar Hills District'). In 1970, the two parts were separated, and the 'Mikir Hills' district was renamed as Karbi

8 There might be several reasons for this. For once, it might be because most Karbi speakers do not understand song language words. Another reason could be that the song language is closely linked to the traditional religious belief and rituals.

Anglong in 1976, with Diphu as the capital.[9] Until 2015, the Karbi Anglong district comprised of an eastern and a western part, with the capital Diphu in the eastern part. In 2015, the western part split off and is now a separate district of Assam, called West Karbi Anglong. The term Karbi Anglong now refers only to the eastern part.

The Karbi Anglong and West Karbi Anglong districts are located at the southern edge of the Brahmaputra Valley (the Brahmaputra being the river to the north) and most of it in the lower hills that mark the transition between the river valley area and the hill range that extends to the south and the southeast as hinted at in Figure 1 above.[10]

The district capital of Karbi Anglong is Diphu and is located in the southern portion of the district. West Karbi Anglong is generally considered to be home to the traditional-cultural center of the Karbis. Specifically, the village of Ronghang Rongbong near Hamren, the district capital of West Karbi Anglong is considered to be the major traditional-cultural center of the Karbis as it is home to the Karbi kings, i.e., *lindókpō*.

1.2 Karbi in Tibeto-Burman

While there has never been any doubt that Karbi is a Tibeto-Burman language, the exact phylogenetic status inside Tibeto-Burman has not been possible to determine. This is despite the fact that information on Karbi has been available early on (§ 1.6.1), and that it has been considered in the early large-scale Tibeto-Burman classification proposals as well as in the modern, detailed comparative work. The difficulties of working out the exact phylogenetic status of Karbi are likely in large part due to a history of language contact and grammatical reorganization as a result of it (§ 1.3).

1.2.1 Tibeto-Burman languages of Northeast India

Northeast India is home to the greatest diversity of Tibeto-Burman languages, including languages from several different branches, such as Bradley's (2002) Western, Sal, and Central branches, see Figure 5.

The Sal branch in Bradley's proposal is a more inclusive version of this branch whose name was coined by Burling (1983), but renamed later more transparently as the 'Bodo-Konyak-Jinghpaw' branch (Burling 2003). Besides this Bodo-Konyak-Jinghpaw subbranch, Burling's (2003) proposal to classify the Tibeto-Burman languages

9 This information comes from http://www.karbianglong.nic.in/, which is the official website of the Karbi Anglong District Administration, accessed on February 3, 2014.

10 This is part of the hill range that extends all the way into Southeast Asia and represents something of a cultural area, with similar histories of the people inhabiting them, see Scott (2009).

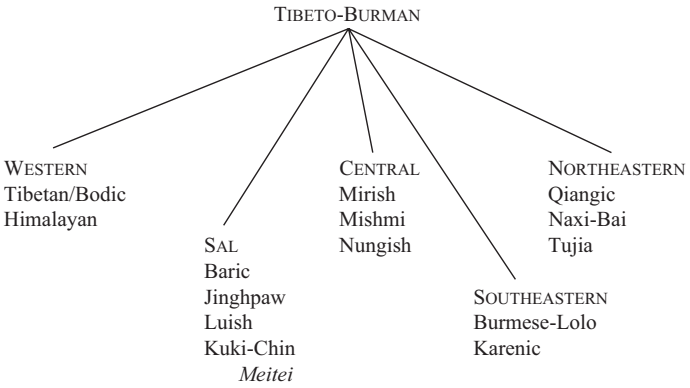

Figure 5: Classification of Tibeto-Burman according to Bradley (2002).

of Northeast India includes a substantial number of other low-level branches, whose higher-level groupings remain far from clear. As seen in the classification of the languages of the 'Eastern Border' in Figure 6, Karbi has in this context always been one of two languages (the other being Meitei, the state language of Manipur) that have been particularly difficult to associate with one of the other low-level branches.

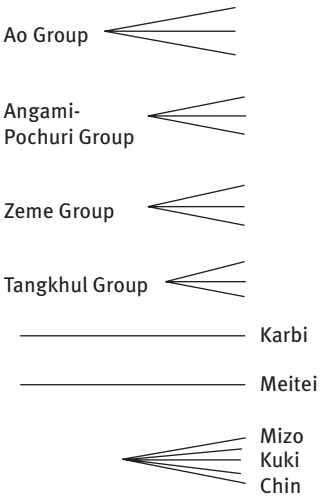

Figure 6: Burling's (2003: 184) 'Relationships among the languages of the Eastern Border'.

1.2.2 Karbi in Tibeto-Burman classification proposals

Karbi (then referred to as 'Mikir') was included in the Linguistic Survey of India (LSI) by Grierson and Konow in the early 20th century (Grierson 1903), which represents the first attempt at classifying Tibeto-Burman languages. Already at that time, there

was a fair amount of information available on the language. In the LSI, it is noted that Karbi "has received some attention from the missionaries who work among them", and "we have a vocabulary and some short pamphlets written in it and an admirable grammar with selected texts from the pen of the late Sir Charles Lyall" (Grierson 1903: 69). Still it was unclear where in the classification Karbi belongs:

> In Volume III, Part ii of the Survey I have classed Mikir as falling within the Nāgā-Bodo Sub-Group. The language has affinities with Bodo, but subsequent investigation has shown that it is much more closely connected with Kuki, and that it should be classed [...] as belonging to the Nāgā-Kuki Sub-Group, in which it occupies a somewhat independent position. (Grierson 1903: 69)

While the absence of a closer link between Karbi and Bodo-Garo[11] has not been controversial since, there are three groups in particular that have been linked to Karbi in proposals in the literature: Meitei, 'Naga', and 'Kuki-Chin' (the latter two of which were put into one group, going back to the LSI, see above).[12]

However, the evidence that underlies the grouping of Karbi with Meitei might better be analyzed as borrowings (§ 1.3). The putative grouping with 'Naga' is complicated due to the fact that it is not currently clear at all what 'Naga' actually is, as there is a long-standing confusion of ethnic and linguistic labels surrounding the term 'Naga' (Burling 2003) (i.e., using ethnic labels as linguistic labels, see also § 1.4.2 for a similar problem within Karbi 'dialects'). A possible link to South-Central or 'Kuki-Chin' currently appears promising. As pointed out in various places throughout this grammar, particular links to 'Kuki-Chin' exist, for example with respect to: the negative equational copula (§ 6.2.2.2); the cislocative as well as speech act participant non-subject marking (§ 8.3.1.4); the reflexive/reciprocal prefix (§ 8.4.3); and the realis focus marker *=si* (§ 12.7.3.1.5), among other constructions. However, working out the exact details as well as implications of these similarities and apparent cognates is a matter of future research. This will likely include the difficult work of carefully disentangling areally from genetically shared features.

In sum, it has remained difficult to come up with a classification proposal that places Karbi in a closer relationship with one of the (geographically) neighboring branches. This is despite early availability of information on Karbi grammar and lexicon (§ 1.6.1), but certainly has to be seen in the context of the remaining lack of information on some of the 'Kuki-Chin' and so-called 'Naga' languages. What appears quite obvious, however, is that a major factor in obscuring the relationships between Karbi and other Tibeto-Burman languages has been language contact and contact-

11 Bodo-Garo languages form a "compact, low-level branch of Tibeto-Burman" (DeLancey 2012). The few similarities that exist between Karbi and Bodo-Garo, such as the Karbi *ke-* nominalizer that is a cognate of a Bodo-Garo adjectival prefix (Konnerth 2009, 2012), stem from a very high node, possibly Proto-Tibeto-Burman.

12 There also was a proposal by Bauman (1976) to consider Karbi the missing link in a connection between 'Kuki-Chin' and Lepcha.

induced changes in Karbi grammar and lexicon. In particular, it has been known since the Linguistic Survey of India that Karbi has been in close contact (and, in fact, the closest contact of all TB languages) with the Austroasiatic Khasi languages to the west in Meghalaya.

1.3 On the role of contact in the development of Karbi

1.3.1 Linguistic evidence for contact with Austroasiatic, Indo-Aryan, and other Tibeto-Burman languages

A proper investigation of lexical and grammatical/constructional borrowings in Karbi has not yet been conducted. In what follows, I will describe the evidence that I have encountered in the course of preparing this grammar.

As far as lexical borrowings are concerned, a very preliminary picture can be obtained from the glossary appended to this grammar. Of a total of about 1,600 entries, there are 88 lexical items identified as apparently coming from Assamese or Indo-Aryan more broadly (e.g., Bengali); 11 as Khasi; 5 as English; and 2 as Pnar. A few caveats limit the informative value of these numbers, however. First, these items were only identified in an opportunistic way, rather than systematically. Therefore, given closer examination, we would expect to find more borrowings. This is particularly true for borrowings from Khasi and Pnar, as the language consultants I have mostly worked with do not know these languages well. Further, although I have tried to eliminate from the glossary all those words in particular from Assamese and English that seemed to be instances of spontaneous code-switching rather than being widely used borrowings, there may still be some code-switching items from Assamese listed in the glossary that I did not identify as such. Also, it needs to be kept in mind that some borrowings might have passed through another language before entering Karbi (see also the discussion in Joseph and Konnerth 2015). Thus, the numbers from the glossary represent merely a rough picture of the amount of borrowing in the Karbi lexicon. Nonetheless, it is interesting to note that the Karbi lexicon apparently only contains a rather small amount of borrowed vocabulary.

Another set of lexical borrowings was identified by Grüßner (1978): a number of lexical items referring to social organization as well as the kingdom system are borrowings from Khasi or Pnar, obviously suggesting that the concept was borrowed along with the word. Examples include the Karbi *lindók-pō* 'king' (with the male suffix *-pō*) from Khasi *lyngdoh* 'priest' (or Pnar *lŋdɔʔ* (see Ring 2015)), and the word *kúr* for 'clan'.

With respect to closed class borrowings, we find a small number of borrowings from Assamese, such as the reflexive pronoun *anijé* 'REFL' or *-lokòt* 'along with', which is used as a relator noun in Karbi. Also borrowed from Assamese, we find interjections, such as the exclamation *ekdóm*, the correction word *bá*, and the hesitation

word *mane* (§16.5, §16.6). As far as closed class borrowings from other languages are concerned, one example is the Karbi singular human classifier *-nūt*, which Joseph (2009) argues to derive either from Standard Khasi (there reconstructed as **shi-ngut*) or Pnar (reconstructed form **chi-ngut*).

No clear cases of borrowings have been identified in the domain of grammatical constructions.[13] However, the causative prefix *pe-~pa-* may represent a borrowing from Austroasiatic (cf. §8.4.2).

Otherwise, one case of possible contact with the Tibeto-Burman language Meitei are the peculiar, corresponding numeral systems in both languages. As discussed in §6.4.2, the numerals 'eight' and 'nine' are morphologically complex forms that translate as 'ten minus two' and 'ten minus one' in both languages. This subtractive construction for 'eight' and 'nine' is not attested so far in any other language in the region. The corresponding constructions in Karbi and Meitei look calqued, since the individual morphemes do not correspond.

Another form that connects Karbi to Meitei, although also to 'Kuki-Chin', is the suffix *-pī* 'female; augmentative' (§7.4.1.1). The Meitei suffix has the segmentally identical form *-pi* 'female'. Likewise, in 'Kuki-Chin', we find an augmentative form reconstructed by VanBik (2009: 84) as **puy*, which in most of the languages takes the form *-pi*. This is noteworthy because it is not a typical Tibeto-Burman form and to my knowledge not attested anywhere else in the family. It is not clear yet what to make of this apparent correspondence.

In addition to the evidence from numerals and the female suffix, there are several other correspondences that could potentially contribute to a model of Karbi-Meitei contact, but without being strong evidence. For example, the Karbi word *ōk* 'meat' is peculiar because the more common Tibeto-Burman root for a word 'flesh' is something like Matisoff's (2003) reconstructed **sya*. The Meitei word for 'pig' is *ók* (Chelliah 1997), and considering that pork is the major and favorite type of meat eaten by the Karbis, *ōk* might be a borrowing from this Meitei word for 'pig'. At the same time, the Meitei *ók* is similar to roots for the word 'pig' in other Tibeto-Burman languages. Mat-

13 Note that one intruigingly parallel construction between Karbi and 'Kuki-Chin' is the marking of non-subject, or object, speech act participants on the verb. In Karbi, the marker is *nang=*, derived from a second person form (§8.3.1). The exact same construction, with a slightly different second person possessive form *nə-* is also found in Purum (Sharma and Singh 2008; see Konnerth 2015) and a vowel-harmonic form *nV-* in Chiru (Chiru 2019), both languages from the Northwestern subbranch of 'Kuki-Chin'. In addition, Aimol (Northwestern 'Kuki-Chin') also has a marker *na-* for indexing first person O arguments (personal fieldwork). Other 'Kuki-Chin' languages have innovated first person or first/second person O argument marking from different source constructions, but the prevalence of this kind of construction is very typical for 'Kuki-Chin'. But despite the near-identity of the construction in Karbi on the one hand and Purum and Chiru on the other hand in terms of synchronic function and diachronic source, it is unclear whether to propose that language contact plays a role here. There are other similarities between Karbi and 'Kuki-Chin', but they may be due to a phylogenetic link rather than contact.

isoff (2003) here reconstructs **p^{w}ak* (the Karbi word is *phāk*). Therefore, this does not represent strong evidence for a borrowing from Meitei into Karbi.[14]

Finally, suggestive evidence for contact between Karbi and Austroasiatic populations comes from the set of common female Karbi names that start with *ka-*, which is the female article in Khasi and Pnar, for example, *Kare*, *Kasang*, *Kahan*, *Kache*, as insightfully observed by U.V. Joseph (personal communication).

1.3.2 Non-linguistic evidence for contact with populations speaking Austroasiatic and Indo-Aryan languages

The current state of research presented in the previous section, while being only a preliminary survey, suggests that there is not much linguistic evidence for contact between Karbi and other languages. However, there is also some non-linguistic evidence to complement the linguistic evidence. This non-linguistic evidence so far includes just several pieces of the much larger puzzle regarding the history of language contact between Karbi and Austroasiatic languages on the one hand and Indo-Aryan languages on the other hand. The limited facts presented here do not allow yet to model the contact scenarios. But they do bear witness to the existence of considerable contact between Karbi and languages of these two language families. Note that I remain purposefully vague about the exact languages involved. Among the Austroasiatic languages, the two most likely candidates for contact with Karbi appear to be Pnar and Khasi; among Indo-Aryan languages, it would be Assamese and Bengali. All four languages (and perhaps others) have probably been in contact with Karbi at some time, and it is difficult and in many cases probably impossible to tell a Pnar origin from a Khasi origin, and an Assamese origin from a Bengali origin.

As a first piece of evidence, consider the present-day geographic distribution of languages. Figure 7 provides the locations of Pnar, Khasi, War, and Lyngam speaking populations. The West Karbi Anglong district is located at the northeastern border of these areas. This map shows that the geographic spread of populations speaking the Austroasiatic Pnar and Khasi languages overlaps with the Karbi speaking area. It is in particular Pnar that is in close contact with Karbi as it is spoken in pockets inside the West Karbi Anglong district and borders the Karbi speaking area to the south. Khasi is spoken along the western border of the Karbi speaking area but not inside it.

A history of close contact with Pnar and Khasi in the western part of the Karbi speaking area is also evidenced by the names of Karbi villages whose etymologies go back to Pnar or Khasi. For example, a number of village names in West Karbi Anglong

14 In addition, there is also a demonstrative *si* in Meitei (Chelliah 1997) that represents an alternative (or possibly ultimately the same) cognate for the focus marker *=si* in Karbi (which is suggested to be connected to an equational copula *si(i)* in Central 'Kuki-Chin' in § 12.7.3.1.5).

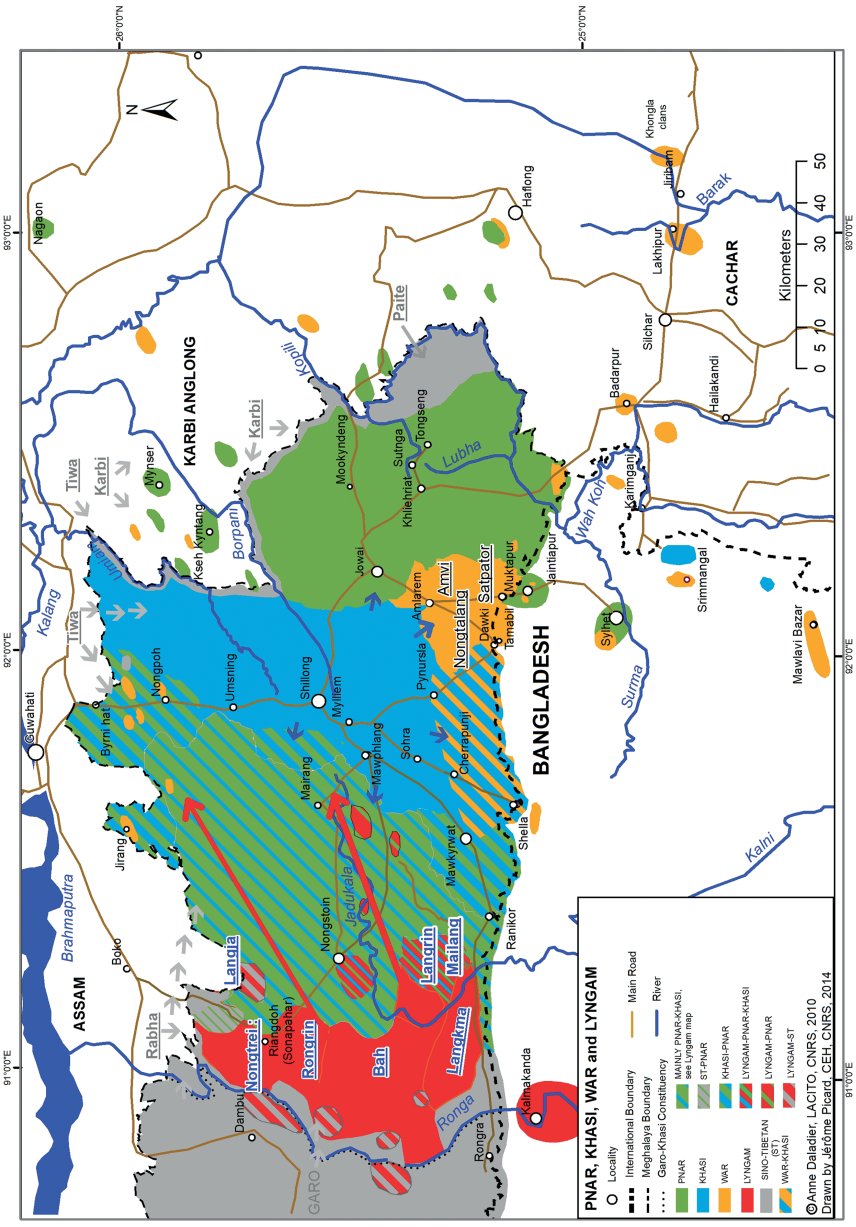

Figure 7: Pnar, Khasi, War, and Lyngam speaking areas (Daladier 2014).

contain a syllable *um*, which is the word for 'water' in both Pnar and Khasi. Examples are *Umswai*, *Umpanai*, *Umlaper*, as well as probably village names with *am*, such as *Amkachi* or *Amtereng*. This evidence suggests that these villages used to be Khasi or Pnar territory and at some point were resettled by Karbis.

Another piece of non-linguistic evidence for contact between Karbi and Austroasiatic and Indo-Aryan comes from the ethnographic-historical work by Stack and Lyall (1908). Without citing a source, they claim that, "during the Burmese wars in the early part of the last century [...] many Assamese are reported to have taken refuge with [the Karbis], and to have become [Karbis]" (Stack and Lyall 1908: 22). They even provide further detail with respect to how people from other communities were able to become Karbis: "[...] in North Cachar outsiders are admitted into the tribe and are enrolled as members of one of the *kurs*, after purification by one of the Bē-kuru *kur*" (Stack and Lyall 1908: 23; italics original). They also provide a remarkable photograph with the caption "Group of Mikirs [=Karbis] (North Cachar)", where five men of widely varying physical appearance are seen. The authors explain that "[i]n the group [...] the short man is evidently a Khasi, while the man to his left appears to be an Assamese" (Stack and Lyall 1908: 23).

This ethnographic evidence thus provides specific information on practices of accepting non-Karbis, and therefore importantly, non-Karbi speakers, into Karbi society.

1.3.3 A creolization account for grammatical characteristics of Karbi

Many aspects of Karbi grammar can be understood in terms of a more general type of grammatical profile that DeLancey (2010, 2011b, 2012, 2013) has called "creoloid". This section is dedicated to a discussion of these grammatical characteristics as a way of connecting facts about Karbi grammar. It is argued that we find the "creoloid" signature across many grammatical domains.

In a number of recent articles, DeLancey (2010, 2011b, 2012, 2013) has offered an explanatory account for the divergent grammatical profiles that are found in the Tibeto-Burman family: Some have a substantial amount of fusional morphology; have verbs with rigid syntagmatic slots and person indexation systems of, in some cases, extremely high synchronic and diachronic complexity. Others may have the opposite: agglutinating, transparent morphology; verbal systems with grammatical markers in flexible position and hence semanto-pragmatic rather than morphosyntactic organization, and no verbal person indexation.

DeLancey argues that the latter type of grammatical profile is innovative. It is correlated with what we may refer to as a particular 'ethnolinguistic profile', which includes several factors that are related to language contact. Particular case studies of this grammar type, characterized by transparent morphology and the absence of

verbal person indexation, among other features, include the ancestral languages of the Bodo-Garo branch (DeLancey 2012) and of Sinitic (DeLancey 2011b).

While Bodo-Garo is argued to have an origin in a lingua franca in DeLancey's proposal,[15] such an extreme case of contact influence does not need to be assumed in the case of Karbi. However, a considerable impact from contact has to be part of the history of Karbi as well. As I argue in this section, DeLancey's framework provides a useful approach to understanding Karbi grammar in a holistic way. It allows us to see parallels across different grammatical domains. The following sections will provide an overview of the areal ethnolinguistic profile of Karbi (§ 1.3.3.1); aspects of its overall grammatical profile (§ 1.3.3.2); as well as one brief case study that illustrates a more general characteristic of the grammatical make-up of the language. It shows what we may refer to as the non-obligatoriness of grammatical markers: they are found lacking in morphosyntactic contexts that should require their presence (§ 1.3.3.3).

1.3.3.1 Areal ethnolinguistic profile

DeLancey (2013) provides case studies in which languages with a "creoloid" grammatical profile (§ 1.3.3.2) are argued to have become that way due to a situation of intense language contact. The type of language contact that is required in order to reshape the grammatical make-up of a language involves one or several events of a substantial number of adult second language learners entering into the community. The idea is that these adult non-native speakers cannot learn the language perfectly and thus speak a 'decomplexified' version of it, which becomes the way the language is spoken in the community at large.

Within the local context, the (pre)historical events of adult non-native speaker influxes would have naturally been common in the valley areas of the Eastern Himalayan region where the expansions of kingdoms and new conquests happened on a regular basis. Certain hilly areas also underwent similar events, in particular those near the large valley areas where there has always been a lot of back and forth between valley and hills.

If the historical scenario for the reshaping of grammar towards a "creoloid" profile has to do with one or several periods of adult non-native speaker influxes, then we will also expect that we are dealing with a community of a substantial size to accept the incoming strangers into the society. This kind of account is thus more plausible if the community is large both in terms of number of speakers and in terms of the size of the area where the language is spoken.

To sum up, the ethnolinguistic profile that lends itself naturally to correlate with "creoloid" grammar involves variables such as (a) number of speakers; (b) size of area covered by the language; and (c) location of the area covered by the language in terms of topography.

---

15 This proposal goes back to Burling (2007).

As for (a), the estimate of the total number of speakers is somewhere around a half million (§ 1.1.2). This makes Karbi the third largest tribal language in terms of number of speakers in Assam. The language with the highest number of speakers is Boro, which is one of DeLancey's (2013) case studies for his argument.

In terms of (b), the size of area covered by the language, we find that Karbi speakers are spread across a very large area. The approximate outline of today's Karbi speaking area is shown in Figure 2 in § 1.1.2. It extends across all of the central districts of Assam as well as into eastern Meghalaya and southern Arunachal Pradesh.

Karbi has also historically been recognized as a particularly large community and language in the local context. In the preface to his Karbi dictionary, Walker (1925) says, "[...] the Mikirs [=Karbis] are among the more numerous of the Assam frontier races, and [...] they are scattered over a wide area, from Golaghat to Kamrup and the Khasi Hills beyond Gauhati, and from the Cachar plains near Silchar to the forests north of Bishnath in Darrang [...]." A similar remark about the wide geographic spread of the Karbi community stems from the Linguistic Survey of India, where it is noted that "it cannot be doubted that in former times the Mikirs occupied a comparatively large tract of country in the lower hills and adjoining lowlands of the central portion of the range stretching from the Garo Hills to the Patkoi" (Grierson 1903: 69). According to this account, the historical spread of the Karbi speaking area extended even further east than today as there do not appear to be villages found nowadays in the Patkoi hill range, i.e., in any of the far eastern districts of Assam or in Nagaland, Manipur, or Mizoram.

This last quote also mentions the topographic location "in the lower hills and adjoining lowlands", which is an accurate characterization for the modern extension of the Karbi speaking area as well. Therefore Karbi does not exactly align with either the "Valley" or the "Hills" type of languages and cultures that have repeatedly been found to exhibit very different characteristics (cf. Scott 2009; DeLancey 2013 and references therein). However, the immediate adjacency of the Karbi speaking lowlands area to the Brahmaputra Valley clearly represents another factor in favor of reconstructing the kind of high contact scenario that has shaped the evolution of Karbi grammar.

In sum, Karbi has an ethnolinguistic profile very similar to that of other languages argued to have "creoloid" grammar by DeLancey (2012; 2013): large community size, large area covered, and location in the lower hills adjacent to the Brahmaputra Valley.

1.3.3.2 Aspects of a "creoloid" grammatical profile

DeLancey's "creoloid" grammatical profile revolves around the notion of morphosyntactic transparency. As DeLancey puts it, "A characteristically creoloid morpheme has a unitary, coherent meaning, which is inherent to the morpheme itself, not dependent on paradigmatic or syntagmatic relations to other morphemes" (2013: 45). This notion of transparency can be broken down into a number of characteristics, which Karbi shares with languages of DeLancey's "creoloid" type.

First, Karbi has consistently agglutinating morphology. The amount of morphophonemics is minimal, and is restricted to a small number of tone changes and allomorphy in the limited prefixation processes (§ 3.9). There are no irregular forms (such as, for example, verb stem alternations in the 'Kuki-Chin' languages (King 2009)).

There is also a considerable amount of evidence that Karbi is not concerned with transitivity as a morphosyntactic notion. Although on the one hand, it is the case that several predicate derivation suffixes can be argued to change the argument structure, as shown in § 9.2.5.4. On the other hand, however, other, including more typical derivational categories do not always affect the argument structure in a consistent way. First, the causative prefix *pe-~pa-* is shown to produce different argument structures in § 13.2.3.1 without any additional marking. The examples show that the same transitive verb *chetòng* 'meet' can occur with two different argument structures when the causative prefix is added to derive the meaning 'make somebody meet somebody'. Also, the reflexive/reciprocal marker *che-* can be used on a transitive verb without resulting in an intransitive verb with a single argument. As § 8.4.3 shows, there are a number of examples where a *che-* marked, reflexive verb still has what we would analyze as an A and an O argument. In those cases, the reflexive meaning lies in the identity of the A with the possessor of the O argument. Furthermore, as discussed in § 9.2.5.2, there are two applicative-like verbal suffixes *-pī* 'benefactive/malefactive' and *-ī* 'instrumental, comitative', which cannot be considered true applicatives precisely because they do not lead to a consistent change in argument structure (see also § 13.2.3.2). Also more generally, Chapter 13 argues that there is no direct morphosyntactic assignment of clausal participants to either argument or oblique roles. The lack of a clear distinction further speaks to the fact that transitivity is not an important notion in Karbi (see specifically § 13.1.2).

There are also no person indexation paradigms in Karbi as found in the conservative Tibeto-Burman languages. There is the phenomenon of indexing non-subject speech-act participants, as well as the cislocative, via *nang=* as discussed in § 8.3, but this is strikingly different from the rich and archaic person marking systems found elsewhere in the family. More generally, Karbi has shed much of the archaic morphology that we can still find in other Tibeto-Burman languages. While there are reflexes of the Proto-TB nominal prefixes **a-* '3SG' (§ 7.3.1), **m-* 'intransitive, durative, reflexive' (§ 6.3.1), and **r-* with an unclear function (§ 6.3.2), as well as the verbal prefix **gV-* 'nominalizer' (Chapter 12), the language is predominantly suffixing. A particularly rich suffixal category are predicate derivations, among which a number of recent grammaticalizations can be found (§ 9.2).

1.3.3.3 Non-obligatoriness of grammatical markers

Another characteristic of Karbi grammar that is arguably a direct outcome of its "creoloid" nature surfaces in many different grammatical domains. It can be described as a general non-obligatoriness of grammatical markers. It is another way in which Karbi has apparently replaced rigid morphosyntax with pragmatics.

As discussed in Chapter 12, the nominalizer *ke-* not only has synchronically productive functions but also has grammaticalized, within particular constructions, to verbal markers of imperfective aspect and focus. However, the focus construction is difficult to analyze because the occurrence of *ke-* is not consistent. Possibly, this has to do with the grammaticalization of the construction, which has resulted in the reanalysis of one component of the construction as a dedicated focus particle, as argued in §12.7.3.1.1, which would have left the *ke-* without a synchronic function. Perhaps it is a result of not serving a synchronic function that *ke-* is often left out in this construction. However, even in those cases where *ke-* does serve a synchronically nominalizing function, it turns out that its occurrence is not consistent in all cases where its presence is expected to be required. These cases are discussed in §12.8. One example is (616), which is repeated below as (1).

(1) Lack of *ke-* 'NMLZ' on relative clause verb
[...] "he matsi", hala apiso abang pulo,
he komāt=si hála a-pisò abàng pù-lò
hey! who=FOC:RL that POSS-wife NPDL say-RL

"he therak thekthe apinso"
he **[[therā**k **thèk-Cē**] a-pinsò]
hey! **be.ashamed** **know.how-NEG** POSS-married.man
'[...] "Hey, who is that!", the wife said, "hey, (you are) a man who doesn't feel any shame".' [SeT, MTN 034]

Another highly frequent grammatical marker of Karbi is *a-* 'possessive'. It attaches to a head noun that is modified by a preceding modifier of any kind, be it a relative clause, a derived property concept term modifier, or a demonstrative (§10.4). However, also *a-* is occasionally found lacking, for example in (381), repeated as (2).

(2) Preposed demonstrative *lasō* 'this' without *a-* on head noun
[...] amat laso sarpita ajo mek janglo [...]
[amāt **[lasō** **sarpī=tā**] a-jó mēk jáng-lò]
and.then **this** **old.woman=ADD** POSS-night eye fall-RL
'[...] [A]nd then also that old woman slept at night. [...]' [KK, BMS 118]

In addition to what we have just seen with *ke-* and *a-*, we also find "non-obligatoriness" of marking in the clausal domain, more specifically with respect to noun phrases that express additional, "non-core" roles.[16] Typically, noun phrases that express these additional participants are found inside a relator noun construction or with the

16 See, however, Chapter 13 for argumentation that Karbi does not strictly distinguish between core and oblique participants in its morphosyntax.

comitative/instrumental/ablative clitic *=pen*. However, even these oblique NP's may remain unmarked as discussed in § 14.1.1.7. An example is (708), which is repeated as (3). In this example, the second clause *nangtumke mandule cho* has an unmarked locative NP *mandule* 'in the *mandu* (field hut)'.

(3) Unmarked (non-salient) locative NP with *chō* 'eat', but marked salient locative NP (*angsóng* 'high up')
*[...] nangpole **hemtap angsong** chote, nangtumke mandule*
[[nang-pō=le **[hēmtā**p **a-ngsóng**] chō-tē] nang-tūm=ke
2-father=FOC:IRR **tree.house** **POSS-high.up** eat-if 2-PL=TOP
cho
[mandú=le] chō]
field.hut=FOC:IRR eat
'[...] "[I]f your father takes his meal in the *hemtap*, you eat in the *mandu*".' [CST, RO 017]

We also find "non-obligatoriness" in subordinate clause markers. The marker *-ī* 'with' on a relative clause verb, which regularly indicates that an instrumental clause participant is being relativized on, is not obligatory (§ 12.3.1.1). This is illustrated by (550), repeated as (4), where the relative clause verb *tòk* 'pound' is only nominalized by *ke-* rather than additionally being marked with *-ī* as would be expected.

(4) Instrument relativization without *-ī*
lasi la thap ketok alengpumta
lasì [là **[thàp** **ke-tòk]$_{RC}$** **a-lengpūm=tā]$_{HN}$**
therefore this **cake.for.rice.beer** **NMLZ-pound** **POSS-pestle=ADD**
otdunno, [...]
ót-dùn-nō
touch-JOIN-be.bad
'The pestle with which the rice beer cake is ground is bad to touch. [...]'
[WR, BCS 037]

Relative clauses with a future sense are also expected to be explicitly marked as such by carrying the irrealis2 marker *-jí* (§ 12.3.1.2). Nonetheless, *-jí* also turns out to occasionally be lacking where it should be required. An instance of this is discussed in (553), (partially) repeated as (5).

(5) Future relative clause without *-jí*
ta ne kethan atomo abangke [...]
tā [[nè **ke-thán**] a-tomó abàng=ke]
but 1EXCL **NMLZ-tell** POSS-story NPDL=TOP
'the story I'll be telling now, [...]' [KK, CC 008]

Finally, the proclitic *nang*=, which highlights the involvement of non-subject speech-act participants, may be left out. While *nang*= regularly occurs on the verb to cross-reference first or second person O arguments, this is only the case in one of two instances in (210), repeated as (6) (§ 8.3.1.2). There are two parallel sentences in this example, with parallel relative clauses in which A and O are the same and only the relative clause verb changes. Despite the parallel structure, or more likely because of it, the second instance of the same second person O argument is not cross-referenced with *nang*= while the first one was.

(6) Third person acting on second person (3→2)
*athema nangphan **nangkelang** inut donangji*
athēma **[[[nàn**g-phān **nang**=ke-làng] e-nūt] dō-náng-jí]
because **you**-NSUBJ **1/2:NSUBJ**=NMLZ-see one-CLF:HUM:SG exist-need-IRR2

kevan kepon inut donangji […]
[[[ke-vàn ke-pòn] e-nūt] dō-náng-jí]
NMLZ-bring NMLZ-take.away one-CLF:HUM:SG exist-need-IRR2
'Because there needs to be somebody to look after you, there needs to be somebody to bring you and to take you. […]' [SH, CSM 066]

This section has surveyed a number of different grammatical markers, across a variety of grammatical domains, which all share the property that they are not syntactically obligatory. What all these cases arguably have in common then is that rather than exhibiting the expected one-to-one mapping between form and function, these grammatical morphemes indicate a function that can also be left unmarked, letting the context disambiguate what the utterance is about. Bringing this back to the proposed "creoloid" profile of the language, it is argued that such a pragmatically oriented morphosyntax, which leaves a considerable amount of disambiguation to the (non-linguistic) context, is likely to have its origin in a high contact situation where the influx of adult second language learners "interrupts" the language transmission process to the new generation (McWhorter 2007).

1.3.4 Summary

The current state of research suggests that there is not much linguistic evidence for contact between Karbi and languages from other language families, specifically Austroasiatic or Indo-Aryan, that would help explain the creoloid aspects of Karbi grammatical structure as argued above. Nonetheless, there is non-linguistic evidence for contact that strongly suggests that there were second language learners of Karbi joining the community. The hypothesis at this point can thus be that there may have been native speakers of different languages joining Karbi society. If it was not a single

group that joined then it makes sense that we do not see large amounts of borrowings from a single language but rather the kind of systematic grammatical restructuring that has resulted in present day Karbi grammar.

As discussed in the next section, apart from a single major dialect divide, there is little dialectal variation in Karbi. This may reflect a rather shallow time depth of the modern Karbi language that has been emerging as a consequence of the assumed restructuring.

1.4 Varieties of the Karbi language

While the details of the Karbi dialect situation are outside the scope of this grammar, it appears that there is a high degree of homogeneity – perhaps surprisingly so, given the large geographic spread of the language. This was also noted by Walker (1925) as he writes in the preface to his dictionary that "in spite of the fact [...] that [the Karbis] are scattered over a wide area, [...], the language is practically one and the same throughout."

The simplified 'big picture' of the dialect situation, is that there is a major dividing line (political as much as linguistic in nature) between the Hills Karbis (Karbis from *Karbi Anglong* and *West Karbi Anglong* ((W)KA)) and the Plains Karbis (Karbis mostly living in the plains of Assam largely north of (W)KA), as discussed in § 1.4.1.

Within each of these major two varieties, there is relatively little dialectal variation. However, investigating the nature of dialectal variation is complicated due to the application of dialect labels by Karbi native speakers, which are grounded in historical ethnic/familial and/or geographical affiliation, as outlined in § 1.4.2.

Following this discussion, § 1.4.3 further discusses two of these dialect labels from the Hills Karbi variety: the Rongkhang or Ronghang dialect, which (with apparently wide-spread acceptance) is being used as the basis for standardization; and the Hills (not Plains) Amri Karbi dialect, which is spoken in West Karbi Anglong, where the traditional-cultural center of the Karbis lies.

Finally, § 1.4.4 offers a list of some lexemes that have been found to exhibit (mostly, vowel) alternations in the speech of different native speakers, without, however, actually appearing to represent dialect isoglosses.

Note that besides these geographical and historical/ethnic dialect groups, there appears to be some evidence for a Christian sociolect (possibly specifically in the Tika region), with some slight differences in lexicon and grammar from the non-Christian sociolect (see § 1.6.2 on Grüßner's work, which was based on a variety with some such features). All of these issues pertaining to linguistic varieties of Karbi require further research.

In the discussion of varieties of the Karbi language, in the following subsections as well as in the entire grammar, I want to emphasize that not a single word is written with a political motive behind it. My goal has always been to describe the linguistic

landscape in a scientific way and to be as neutral as possible when it comes to the politics that are, of course, tied to it in real life. I truly hope that no part of the discussion of the different varieties of the Karbi language is offensive to anybody.

1.4.1 Plains Karbi ("Amri Karbi") and Hills Karbi

The Plains Karbi variety spoken in the Kamrup and Marigaon districts of Assam as well as partly in the Ri-Bhoi district of Meghalaya is commonly referred to as 'Amri Karbi' in the linguistic literature and in the Ethnologue (Lewis, Simons, and Fennig 2013). However, it should be noted right away that Karbis identifying with the western subvariety of Hills Karbi use the same name for themselves, possibly because of an ultimately shared geographical origin and/or common ancestors (§ 1.4.3).

Members of the Karbi Lammet Amei (§ 1.1.4) have expressed their concern to me over the use of the name Amri for the Plains Karbi variety, while this name is embraced by many speakers of this variety. The term 'Dumra' or 'Dumrali' is also used to refer to this variety of the Karbi language and the people that speak it.

In the following discussion, I will use the geographically based terms 'Plains Karbi' and 'Hills Karbi' to refer to the two major linguistic varieties of the Karbi language, which, again, include further 'sub'-varieties based on linguistic features, which are, however, not as different from one another as are the two major varieties. Although these geographic terms are not ideal either since there are Karbis living in the plains who do not speak the 'Plains Karbi' variety, I follow Teron and Tumung (2007) in using 'Plains Karbi' and 'Hills Karbi' in these ways, as the terminological debate is currently still ongoing and there simply is no ideal set of terms to use at this point. Note that the name Amri refers to a historical administrative unit in the Karbi kingdom, and, as mentioned above, in addition to Plains Karbis, the group of Hills Karbis living in the West Karbi Anglong district also identify with this name, see § 1.4.2 and § 1.4.3 below.

There is a strong political movement on part of the Plains Karbis to consider their variety of Karbi a different language rather than just a different dialect from the variety of Karbi that is spoken in the Hills. This likely has to do with the unequal power relations between the two groups. While there are close to half a million native speakers of Hills Karbi that have autonomy in the Karbi Anglong and West Karbi Anglong districts, the Plains Karbi speaking population is scattered across a number of districts, and a 2003 figure reported by the Ethnologue estimates the number of speakers at a total of 125,000 (Simons and Fennig 2017).[17]

17 It is not clear what the basis is for this figure provided by the Ethnologue. The Census data of close to a half a million native speakers of Karbi do not specify whether a distinction was made between Hills and Plains Karbi (and hence it likely was not).

According to my Hills Karbi language consultants, there is a high degree of mutual intelligibility between the Hills and Plains Karbi.[18] This is especially true for Hills Karbi speakers that are fluent in Assamese (which most people living in the urban areas are), as the Plains Karbi variety has a large number of Assamese loans due to closer contact with Assamese in the plains. It is also noted in the Ethnologue that "some Amri Karbi villages shifted completely to Assamese due to intermarriage and the perception that Assamese is preferred for children to do well in school."

In addition to the larger number of Assamese loans, some of the more noticeable ways in which Plains Karbi is different from Hills Karbi are the following. First, there are differences between the song language (§ 1.1.3.3) and the ordinary language. Specifically, Plains Karbi uses lexemes in ordinary, colloquial speech, which are only used in the song language of Hills Karbi, and vice versa.

A phonological difference is that Plains Karbi has preserved coda /l/, which in Hills Karbi has changed to a glide codas (see § 3.4).

In the domain of morphology, there are two salient differences. First, the unusual onset-reduplicative negative suffix *-Cē* in Hills Karbi (§ 3.8.6.3) corresponds to just *-e* without the onset reduplication in Plains Karbi. Second, the Hills Karbi negative existential copula is *avē*, while Plains Karbi instead uses the form *ingjong* (§ 6.2.2.1.1).

With respect to syntax, there are two frequently occurring Plains Karbi constructions that are not used in Hills Karbi. On the one hand, the positive existential copula *dō* is often used following a bare stem. It is not clear to my Hills Karbi language consultants what the function of this construction is, but it is very striking to them. On the other hand, habitual aspect is marked by the suffix *-man* in Plains Karbi, where in Hills Karbi simply the bare stem is used.

Within the Plains Karbi variety, there is also some amount of dialectal variation. For example, in the Ri-Bhoi district of Meghalaya, the Bhoi Mynri variant is spoken.

A good resource on the Plains Karbi variety is the trilingual dictionary edited by Teron and Tumung (2007), which includes both Plains and Hills Karbi forms of each lexical item, as well as translations into Assamese and English.

1.4.2 Relationships between the Hills Karbi 'dialects'

There are four 'dialect' labels that are used by Hills Karbi native speakers to identify their own and other people's speech: Amri, Rongkhang (or Ronghang), Chinthong, and Killing. Originally, however, these labels are connected to historical administrative units of the Karbi kingdom and the people that lived in these administrative units. Therefore, while there certainly is a historical connection between an individual's

18 I myself have not carried out any research on the differences between the two dialect groups. Almost everything I report here is second-hand information from my Hills Karbi language consultants.

affiliation to one of these groups and the variety s/he speaks, this is not always the case anymore today.

Since this discussion of these different groups is only for the purpose of sketching out the linguistic landscape, everything said here comes through the lens of linguistic variation and is considered in its relevance to linguistic varieties. I would like to acknowledge that it is shorthand to speak of 'Amri Karbi' or 'Rongkhang Karbi' and that the more accurate way of referring to the people that identify with these names is to say *Amri aso* 'children / inhabitants of Amri' and *Rongkhang aso* 'children / inhabitants of Rongkhang.'

According to Dharamsing Teron (p.c.), the following can be said about these 'dialect' groups:

The three major groups are Amri, Rongkhang, and Chinthong. The Killing group appears to be a part of the Rongkhang group. The Amri group was the first to migrate into the present-day Karbi Anglong area. They split up and some of them went to present-day West Karbi Anglong (i.e., the group discussed in § 1.4.3), while others moved into the plains (i.e., the Plains Karbis, discussed above in § 1.4.1).

The Rongkhang group is mostly located in the southern portion of (eastern) Karbi Anglong, where the district capital Diphu is located. This group became the most dominant and influential group, which is why the Rongkhang 'dialect' is currently considered the standard dialect. Note, however, that linguistically, it is not possible for my language consultants to pinpoint defining differences between Rongkhang and the 'dialects' of (eastern) Karbi Anglong and the area to the south, i.e., Chinthong, and Killing. (The Killing group is geographically centered in places such as Kheroni, Jyrikyndeng, and further into the North Cachar Hills area.)

1.4.3 Hills Karbi: Differences between Rongkhang (Hills Karbi; Diphu) and Amri (Hills Karbi; West Karbi Anglong) dialects

While any particular differences between the Rongkhang dialect and other dialects to the north and the south are not easy to discern for my language consultants (although among themselves, they may identify with different dialect groups, see § 1.4.2 above), there are a number of differences between, on the one hand, Rongkhang and the other putative dialects, and, on the other hand, the Amri dialect in West Karbi Anglong. These differences are not only lexical in nature, but also include two systematic phonological differences. First, the Amri dialect has a sixth phonemic vowel, which is a high to mid-high, front, centralized /ɪ/ (§ 3.2.1). Second, the Amri dialect has preserved both the /ei/ and the /ai/ rhymes, while the Rongkhang dialect has merged them to /ai/ (§ 3.2).

In addition, there are a number of differences in other grammatical domains as well. For example, the *pe-~pa-* 'causative' prefix is seemingly only ever produced as *pa-* in the Amri dialect, i.e. without any allomorphy (see § 3.9.2.1). Another difference

is that the 'afterthought' particle *=he* (§ 16.3.9) is more frequently used in Amri speech, and that the particle *hedī* is prominently used as a marker of narrative style specifically in this dialect (§ 16.3.12).

1.4.4 Variation in lexemes

Table 2 provides sample lexemes of which there are two (or more) variants based on vowel alternations. No study has been conducted yet to see whether these alternation patterns somehow align in the speech of individual native speakers or whether the variants are tied to particular dialects. My consultants do not consider any forms more correct than others but just report that all variants are used, and that they have not noticed any sociolinguistic patterning of this variation.

Table 2: Vowel alternations.

Alternation	Gloss	/i/	/e/	/a/	/o/	/u/
i~e~u	'banana'	*phinū*	*phenū*			*phunū*
i~e	'trade'	*bihá*	*behá*			
	'match'	*chináng*	*chenáng*			
	'eggplant'	*hipī*	*hepī*			
	'main people in charge'	*khitirí*	*khetirí*			
	'king(<Ind)'	*richó*	*rechó*			
i~u	'yam'	*phirùi*				*phurùi*
	'snake'	*phirūi*				*phurūi*
	'rat'	*phijū*				*phujū*
e~a	'field hut'		*mendu*	*mandu*		
	'dried fish'		*menthu*	*manthu*		
	'when'		*(ko)mentu, (ko)nemtu*	*(ko)mantu, (ko)namtu*		
e~o	'cotton'		*pheló*		*pholó*	
	'alkaline'		*phelō*		*pholō*	
	'story'		*temó*		*tomó*	
a~o	'girl'			*okarjāng*	*okorjāng*	
a~u	'carry on back'			*bā*		*bū*

While Table 2 shows that many different vowel alternation patterns exist, it should be noted that in almost all cases, the vowel alternation occurs in the first syllable of a

disyllabic word. The only two exceptions are the two words listed last: *okarjāng~okorjāng* 'girl' and *bā~bū* 'carry on back'.

In addition to the simple vowel alternations in lexemes listed in Table 2, there also exists lexical variation in other words, such as *mensopi* or *nemsopi* for 'papaya'.

1.5 Sociolinguistic profile of language endangerment

The UNESCO Atlas of the World's Languages in Danger lists Karbi as a 'vulnerable' language (Moseley 2010). This section contains a brief sociolinguistic evaluation of the current status of endangerment of the Karbi language.[19] It includes a discussion of setting factors; the impact of language contact; issues relating to domains, vitality, and attitudes; and the official policy concerning the language. Along with a summary of the degree of endangerment of Karbi, the last section reports on current, and proposes future, remedial actions to strengthen the status of the language.

1.5.1 Setting factors

A number of setting factors have an impact on the degree of language endangerment. First, the number of speakers is a relevant factor, and Karbi has a relatively large community of native speakers at approximately half a million people (for more information, see §1.1.2). In addition, this section will discuss the following factors: (a) languages represented in education; (b) virtual and physical infrastructure in the community; (c) the relationship between government and language (policy); and (d), how Karbi is represented in mass media.

As far as the **languages representend in education** are concerned, even within the Karbi Anglong and West Karbi Anglong districts, most schools have either Assamese or English as the medium of instruction, depending on whether they are government or Christian schools. Within the last twenty years or so, a few Christian primary schools started to teach in Karbi (and other local minority languages), spearheaded by Frs. U.V. Joseph and Joseph Teron, and textbooks have been developed (along with Br. Benjamin Kathar). However, by now these schools have mostly reverted back to English. There are a few private schools that teach in Karbi, and there has been an effort to translate existing textbooks from Assamese into Karbi. By and large, however, children go to Assamese or English medium schools. At the same

19 This profile was put together as part of a seminar on 'The Sociolinguistics of Language Endangerment' offered by David and Maya Bradley at the 2011 LSA Institute in Boulder, CO. I would like to thank them for the feedback they provided me and for engaging me in thinking about these issues.

time, as of the year 2019, there are now official plans to devise a full curriculum for Karbi language classes from primary to high school.

With respect to **infrastructure**, a very sudden increase in virtual infrastructure in recent years (which was very noticeable even just between 2008–2012), i.e., availability of electricity, TVs, cell phones, and the internet, will likely affect the community. Increase in physical infrastructure has recently improved mobility, and will almost certainly affect the community as well. A lot is currently changing, and the endangerment situation ten years ago was likely substantially different from what the situation will be like in five to ten years from now.

Language shift to Assamese is a lot more common among those Karbis who live in the plains, compared to *Karbi Anglong* Karbis, who live in the hills. This has likely been the case historically (in the last several centuries or so) as well, since there are a lot more Assamese loans in the Plains Karbi variety (§ 1.4.1).

In terms of the **relationship between government and language (policy)**, the Karbis live in their autonomous *Karbi Anglong* and *West Karbi Anglong* districts. *Karbi Anglong* has the *Karbi Anglong Autonomous Council* (and that is also its official name, i.e. it is in English rather than Karbi) (see § 1.1.5). This political autonomy strengthens the status of the language.

As far as **mass media** are concerned, there are a few newspapers in Karbi. The *Arleng Daily* and the *Thekar*[20] are written in Roman script, while there are also newspapers that use the Assamese script. The KLA publishes quite a lot of books in Karbi. There is a local film industry that produces movies and comedy shows (and some documentaries) in Karbi. There also is a Karbi language TV program.

There has been a recent increase in availability of Karbi media on the internet. In particular, there are Karbi blogs and there are a number of songs sung in Karbi (but following the 'mainland' Indian *Bollywood* style) on Youtube.

1.5.2 Domains/vitality/attitudes

This section discusses the domains in which Karbi is used, the overall vitality, and the attitudes that native speakers have towards the language.

With respect to the **domains of language use**, there are two situations of language use within the Karbi community that I have experienced first hand and can comment on. One is life in the *Karbi Anglong* district capital Diphu, the other is life in a Christian village in rural *West Karbi Anglong*. Both sets of experiences stem from the time between 2009–2012. As would be expected, the differences are very noticeable.

In Diphu, especially among the middle class, it is typical for Karbis under the age of 35 to 40 years or so, to be quadrilingual. While Karbi is the native language that is

20 The *Thekar* now also has an online edition: http://thekararnivang.com/.

spoken in the home, it is also common for them to use Assamese, English, and Hindi (probably in that order of frequency) on a regular basis. It is therefore no problem for them to switch to whatever language is shared with their interlocutor(s). If several languages are shared, it is typically with first priority Karbi and then Assamese that are used, but that is not always the case, and people enjoy switching between languages. For example, I have heard Karbi native speakers in their late twenties talk to each other using Karbi, but suddenly switch to Assamese or English words or entire sentences.

In the villages that I have visited, this multilingual situation does not exist to that extent. While it is common for Karbis except for the older generation to speak Assamese to varying degrees, they may only ever use Assamese when they go to the weekly or biweekly market where many sellers only speak Assamese and no Karbi. There is a trend for the younger generation, especially with increases in physical and virtual infrastructure, to be exposed to English and Hindi and to know how to say a few things in these languages. In addition, in the villages in *West Karbi Anglong* with Tiwa-speaking[21] villages nearby, it is also common for Karbis to know Tiwa to varying degrees (and vice versa).

The overall **vitality** is currently good for Hills Karbi. Most commonly, Karbi is transmitted to the younger generation and used in the homes.

The **attitudes** of the Karbi community towards their language are generally positive. I have not met Karbis who did not consider their language important to them. The fact that the KLA (§ 1.1.4) exists also speaks to that: The KLA is a non-governmental organization that survives on private donations; its members volunteer, with no monetary compensation for their work.

1.5.3 Other factors

The major dominant language in the area that many Karbi speakers shift to is Assamese. Shifting to Assamese occurs among all of Assam's minority languages, since it is the main lingua franca for speakers of different minority languages. While English functions as a lingua franca to some degree as well (especially in the Christianized areas), Assamese is more widespread.

1.5.4 Summary of degree of endangerment

The Karbi language currently looks healthy. However, especially the increase in infrastructure is changing so dramatically that consequences will likely become more and more noticeable in the near future. Although the facts are that (a) there are a large

21 Tiwa is a Bodo-Garo language spoken mostly inside the *West Karbi Anglong* district.

number of speakers; (b) the language is almost always transmitted to the children; and (c) the community is politically protected by having their own autonomous district, the dominant Assamese language creates a lot of pressure that many in the community feel. The KLA's largest current concern is the standardization of the writing system. This will serve both the purpose of doing language maintenance as well as result in added prestige.

In accordance with the KLA's concerns, the standardization of the orthography could be a large step to strengthen the language, as it would likely represent a prerequisite for the large-scale development of school materials in Karbi and the use of Karbi in higher education. In fact, as of 2019, there are official plans to create a comprehensive curriculum to teach Karbi language classes in schools.

1.6 Previous study of Karbi grammar

There are a small handful of important names in Karbi linguistics, which should be mentioned at the outset of this section. The chronologically first mention should be made to George D. Walker, who published a fairly comprehensive dictionary in 1925. In the mid 1960s, the late French missionary Father Balawan put together a word list with parts of a grammatical description (published as Balawan (1978)). Not much later, Karl-Heinz Grüßner worked on Karbi grammar. Around the same time, Karbi scholar Professor Rongbong Terang published a Karbi dictionary (Terang 1974). Within the last decade or so, substantial contributions have also been made by another important Karbi scholar, Longkam Teron.

1.6.1 Early work on Karbi

Early resources on Karbi include word lists by Robinson (1849), Stewart (1855), and Kay (1904), as well as information provided in sections on Karbi in the Linguistic Survey of India (Grierson 1904). A historical ethnographic description with some references to Karbi grammar as well as several texts with translation was "edited, arranged and supplemented" by Sir Charles Lyall based on notes by the Indian Civil Service officer Edward Stack (Stack and Lyall 1908). Several other early resources on Karbi linguistics are listed by Grüßner (1978: 218–21).

1.6.2 Karl-Heinz Grüßner's work

Karl-Heinz Grüßner worked on Karbi in the early 1970s. Based in Shillong, he collaborated with Karbi native speakers Harrison Langne and Clement Singnar and took a number of trips to villages in West Karbi Anglong to record texts and work with other

native speakers. Grüßner wrote a grammar of Karbi as his PhD dissertation at the University of Heidelberg in Germany, published as Grüßner (1978). The grammar is an excellent resource especially on Karbi phonology and morphology, which is all the more impressive considering the relative non-availability of technical equipment at the time. Grüßner made reel-to-reel audio tape recordings, which he subsequently transcribed for use as examples in his grammar.[22] Grüßner was the first to investigate the tone system of Karbi and he systematically indicated tones in his work. While he was working on Karbi grammar he also began compiling words for what became a 241-page dictionary manuscript, which he never published.

Grüßner's work was mostly based on the speech of Christianized Karbi native speakers from around Tika in West Karbi Anglong. According to some of my language consultants, there are some subtle aspects of the variety Grüßner worked on that partially overlap and are partially different from their own varieties:

For example, Grüßner reports the systematic use of *pe-* 'causative' before monosyllabic roots but the allomorph *pa-* before disyllabic roots (corresponding to parallel allomorphy in the nominalizer *ke-~ka-(~ki-)*), which is a pattern found in Rongkhang speech; however, Grüßner also reports the use of /ei/ instead of only /ai/, which is typical of (Hills Karbi) Amri speech (§ 1.4.3). There also are some lexical differences according to my language consultants. For example, while one of the bamboo parts used to make fire via friction is referred to as *theng-dang* in my consultants' speech (with the first element in the compound being *thēng* 'wood'), Grüßner reports the word to be *me-dang* (with the first element being *mē* 'fire'). These subtle differences are suspected to be part of an (again, subtly) distinct sociolect typical of the Christian villages in the area. While the issue of Karbi varieties cannot be treated here in a comprehensive way (see also § 1.4), it should be noted that certain discrepancies in grammatical description between Grüßner's work and the present work might be due to dialectal/sociolectal differences.

The lasting contribution of Grüßner's work cannot be underestimated. The level of detail and carefully researched description, especially given the historical context when much less was known about the Tibeto-Burman language family, is truly remarkable.[23]

1.6.3 Longkam Teron's work

Longkam Teron's perhaps most well-known contribution is a grammar with the Karbi name *Karbi lamtasam* (Teron 2005a). Significant other contributions include his collections of proverbs (Teron 2005b), and idioms and phrases (Teron 2006). Particularly

22 Grüßner has since digitized his old audio recordings.

23 I also want to take this opportunity to again thank Karl-Heinz for all his support, his enthusiasm, and for freely and fully sharing everything and anything he had that could possibly help me in my endeavours to work on Karbi grammar.

important for further study of the origins of Karbi and the historical contact situation is the collection of words used as part of the (Hills Karbi) song language, *Karbi lamlir achili* (lit., 'the seeds of the Karbi poetical language') (Teron 2008) (see § 1.1.3.3 above).

1.6.4 Other resources on Karbi grammar and lexicon

In 1966, besides Father Michael Balawan, Father John Mariae also produced work on Karbi grammar and lexicon, which was only later published and is now available as a booklet called 'Karbi Self-taught' (Mariae 2007). An overview of the history of research on Karbi grammar is offered by Teron (2011: 148–57).

In addition to the resources on Karbi grammar mentioned above, important contributions are also a number of dictionaries. The first comprehensive dictionary (Karbi to English and English to Karbi) was published by Walker (1925).

The Karbi scholar Bidorsing Kro produced a Karbi to Karbi dictionary with explanations in Assamese and English that was first published in 2002, with a second edition that has been published since (Kro 2009). Another dictionary emerging from scholarship from within the Karbi community is Taro (2010). A comparative dictionary of Plains and Hills Karbi by Teron and Tumung (2007) represents an important resource on the differences between the two major dialects of the language.

1.7 Writing in Karbi and conventions followed in this grammar

The Karbi orthography based on Roman script was developed by missionaries in the 19th century. Currently there also exists a Karbi orthography using the devanagari-based Assamese script. However, the *Karbi Lammet Amei* advocates for the use of the Roman script. Although there is a standardized set of letters in the Roman script that adequately represent Karbi phonemes, the orthography is to date not standardized, which is a matter of concern for many in the community, especially, of course, for the members of the *Karbi Lammet Amei*.

The controversial issues arise with respect to (a) word boundaries; (b) capitalization after adding prefixes; (c) syllable boundaries; and, most importantly, (d) the representation of tone.

Regarding (a) word boundaries, the status of clitics is controversial. For example, it is not clear whether the very frequently occurring topic marker *=ke* (§ 14.2.1) should be attached to the last word of the noun phrase that it is phonologically bound to, or not.

With respect to (b) capitalization after adding prefixes, the most frequent issue is the occurrence of the possessive prefix *a-* (§ 7.3.1) on proper names. For example, in the simple noun phrase 'this Karbi woman', the word 'Karbi woman', i.e., *Karbipi*, needs to have the *a-* prefix on it. That is, one could write this noun phrase as *laso*

aKarbipi, with *a-* in lower case and *Karbipi* capitalized, but that looks odd to some members of the community.

The issue of (c) representing syllable boundaries is often (but not exclusively)[24] encountered when a vowel-initial syllable occurs in the middle of a word. This represents a problem because vowel onsets are accompanied by glottalization, which is not represented in the writing system, but is noticeable in this context since it prevents resyllabification (see § 3.3). For example, adding the predicate derivation suffix *-ò* 'much' (§ 9.2.5.1) to the existential copula *dō* results in a pronunciation of [dōʔò]. As a result, many members in the community dislike a representation as *doo*, resorting to either using a dash (i.e., *do-o*) or an apostrophe (i.e., *do'o*).

Finally, (d) the representation of tone has typically been avoided. Neither the Roman script nor the Assamese devanagari-based script has been successfully modified to represent tone. Different proposals have been put forth to do so, but none has been systematic; either only one tone (usually the most salient mid tone that features glottalization) is represented and/or only certain rhymes are marked for tones (and not actually consistently). For example, tone in open syllables that lack a coda consonant has been represented with an <h> coda by some Karbis – however, it is sometimes the low tone that is represented by this orthographic <h> coda (e.g., *neh* '1EXCL'), and sometimes the mid tone (e.g., *meh* 'fire'). Another proposal specifically aiming at the representation of the mid tone in nasal coda syllables was to write an orthographic homorganic stop, e.g., *(a-)tump* for the plural word with a bilabial nasal coda, *(a-)phant* for the non-subject marker with an alveolar nasal coda, or *langk* for 'water' for a velar nasal coda.

Except for where the representation is not systematic or phonological (i.e., the attempts so far at representing tone), all of these issues are mostly political in nature and there is no right or wrong in linguistic terms. In this grammar, I follow the orthography Grüßner (1978) used. This has the following implications for the four orthography problems outlined above:

For (a) word boundaries, clitics are written as one word with the element that they are phonologically bound to. For (b) capitalization after adding prefixes, the above representation is used, i.e., lower case prefix with capitalized proper noun stem (i.e., *aKarbipi*). With respect to (c) syllable boundaries, the apostrophe is used, i.e., *do'o* for the example word from above. Finally, for (d), the representation of tone, Grüßner's approach with diacritical marks is used: the grave for low tone (e.g., low tone *thì* 'die'); the acute for high tone (e.g., *thí* 'snatch, grab'); and the macron[25] for mid tone (e.g., *thī* 'be short').

---

24 The other situation where this issue arises is when a multisyllabic word contains a consonant combination of /pl/, /pr/, /kl/, or /kr/ between two vowels, where the two consonants could be split up as coda plus initial or an open syllable followed by a consonant cluster.

25 Note that in his dictionary manuscript, Grüßner also sometimes used the circumflex for the mid tone, e.g., representing 'be short' as *thî*.

In this grammar, examples are offered with both a word line and a morpheme line, where tone is only indicated in the morpheme line but not in the word line. The details of the representation of tone is further discussed in § 3.5.9; the details of the representation of data in general in this grammar (including different types of brackets) is further discussed in § 2.4.

1.8 Organization of this grammar

This grammar is organized as follows. Chapter 2 discusses the theoretical framework and methodological approach employed in this grammar, as well as the data that were collected and produced as part of this research, and that are used as a basis for the grammatical description.

Chapter 3 is dedicated to Karbi phonology, a large portion of which concerns the Karbi tone system, which poses difficulties for thorough description due to its low functional load.

Karbi word classes are the topic of Chapters 4–6. Chapter 4 deals with the major word classes of nouns, verbs, and discusses the status of property concept terms. Chapter 5 goes over the subclasses of nouns, while Chapter 6 provides an overview of other word (sub)classes, including pro-forms, verb subclasses, and minor word classes, such as adverbs and numerals.

This leads into the morphologies of the two large word classes of nouns and verbs. While nominal morphology is treated in Chapter 7, the discussion of verbal morphology is divided up into Chapter 8 with a general overview and the presentation of pre-root morphology, and Chapter 9, which deals with post-root morphology.

In Chapters 10 and 11, syntactic issues concerning the noun phrase and predicate constructions are discussed.

Chapter 12 is solely dedicated to nominalization and nominalization-based constructions, as nominalization represents a major structural device with functional application in a number of different domains of grammar (including simple derivation of nouns from verbs, noun phrase modification, monoclausal predicate constructions, as well as clause combining).

The status of clausal participants and the ways in which they may be marked is the topic of Chapters 13 and 14.

A discussion of clause types and clause combining, including non-declarative speech acts, is offered in Chapter 15.

Finally, Chapter 16 provides an overview of the major constructions that have functions on the level of the larger discourse.

2 Methodology and data

This chapter deals with the methodology and general approach to grammar writing that underlies this grammar, as well as some relevant points about the collection and organization of the data that this grammar is based on.

Most of the existing literature on linguistic data management has been published within the fields of language documentation (Himmelmann 1998; Gippert, Himmelmann, and Mosel 2006; Himmelmann 2006a; Woodbury 2011) and the description of (especially endangered) languages (Austin and Sallabank 2011). While this grammar of Karbi is primarily aimed to be a descriptive resource, attempts were made to incorporate the insights from the recently emerged (or, as some would argue, revived) field of 'language documentation' (or 'documentary linguistics').

This chapter begins with a discussion of the general approach and theoretical framework underlying this grammar (§ 2.1). In § 2.2, an overview of the corpus is offered. Aspects of primary data collection are discussed in § 2.3, while § 2.4 covers the ways in which the data have been annotated and are presented in this grammar.

2.1 Approach and theoretical framework

The approach taken in this grammar contains three major components. First, it is firmly rooted in a functional-typological framework. Second, it embraces collaboration with the language community as the best approach for data collection and analysis. Third, the analysis presented in this descriptive grammar gains explanatory force through a historical-comparative perspective on the grammatical constructions that are discussed.

2.1.1 Functional-typological framework

First and most importantly, this grammar is based on a functional-typological framework. This approach permeates all aspects in the design, data collection, and analysis and write-up.

For the design and data collection aspects, this framework puts an overarching emphasis on a data-driven and inductive, as well as data-rich approach to document how Karbi is actually spoken in a wide variety of natural uses of the language. As a result, the vast majority of examples that are offered in this grammar to illustrate a particular point come from naturally produced speech rather than elicitation via translation from English. Elicitation serves an important purpose in supplementing information that did not happen to be provided in data from natural speech, but it should always be treated with the necessary caution.

The implications of using this framework for the analysis and argumentation consist of the understanding that there most typically are functional motivations for patterns and that there are also functional motivations for exceptions to patterns.

For example, Karbi classifiers generally do not mark a distinction between singular and plural: the same classifier is used when counting one or many items of a particular kind. This is functionally motivated because classifiers only ever occur with numerals to form classifier-numeral words in Karbi (§ 5.1). Therefore, having different classifier forms to distinguish singular from plural would be redundant.

However, there is one case in which there actually are two forms that are used for classifying the same entity: humans (or personified animals or objects in folk stories, § 5.1.4). For humans, there are two classifiers, the singular classifier *nūt* and the plural classifier *bàng*. On a first level of explanation, there are two forms for the human classifier, because *nūt* actually is a borrowing from Khasi (Joseph 2009). On a deeper level of explanation, however, it is functionally motivated that the classifier (set) for humans is more special than most if not all other classifiers and is also the most frequently used classifier (set) in the corpus.

The functional-typological approach takes function and use seriously. From there, as this approach considers language a tool for a language community, the typological component is relevant: It motivates why some aspects of language are shaped similarly cross-linguistically, which is due to the shared aspects of human life everywhere in the world. But it also motivates why other aspects of language are shaped differently cross-linguistically, which is due the aspects of human life that are different in different parts of the world, just as there exist different cultures.

2.1.2 Collaboration with the language community

This grammar project was initiated by the Karbi community through the *Karbi Lammet Amei* (KLA; § 1.1.4), and was therefore collaborative in nature from the beginning. Due to this close collaboration with the KLA, numerous aspects of data collection and processing were enhanced.

For example, the KLA (and specifically my closest collaborator Sikari Tisso) identified speakers that were able to tell particular stories and provide particular information. Due to the KLA's own interest in the success of the project, they (and again, most importantly Sikari Tisso) also greatly helped with all aspects of the practical realization and facilitation of the project, which always represents a major and time-consuming task in fieldwork (which at times is even impossible for an outsider linguist without collaborators from within the community).

2.1.3 Historical-comparative perspective

The historical-comparative approach to linguistic explanation understands grammatical constructions as being shaped by their historical origins, which can often be investigated by conducting comparative research.

For example, there are main clause constructions in Karbi that feature the *ke-* prefix that functions as the main nominalizer in the language, without having an inflected element to render the clause finite (§12.7.3). The historical perspective taken in that section aims to explain how that the *ke-* prefix in those constructions can still be diachronically analyzed as the nominalizer, even though it has to be analyzed as a different element synchronically. A specifically comparative component to explanation is part of the analysis for the diachronic nominalization construction to indicate focus. This construction can be accounted for with recourse to a copular form that does not exist in Karbi but does exist in the related "Kuki-Chin" branch of Tibeto-Burman.

2.2 Corpus

2.2.1 Data types

The data that were collected and generated for this grammar can be grouped into three subcorpora, as laid out in Table 3. They are (1) data based on individual recording sessions; (2) context-free elicitation data; and (3) the lexical database. All of these data are – or are based on – spoken data. Additionally, three other types of data were available: (4) data from Grüßner's (1978) work and specifically his transcriptions and translations of audio data he had collected, which, therefore, are also spoken data, (5) an extensive dictionary manuscript that Grüßner provided me, containing 240 pages of morphemes with indication of tone, and (6) data extracted from written texts, mostly school textbooks.

Table 3: Subcorpora of Karbi data.

Subcorpus	Content
1. Texts / Recording session data	Audio/Video recordings (and images) along with their transcriptions, translations, and analyses
2. Elicitation data	Elicitation based on grammatical, phonological, or lexical topics
3. Lexical database	Database including all morphemes occurring in recordings
(4. Grüßner's (1978) data)	Transcriptions / translations of audio recordings; individual examples
(5. Grüßner's dictionary manuscript)	Manuscript of a dictionary of 240 pages with tones represented
(6. Written text data)	Mostly from school textbooks, but also from other published sources

The most important data type are the **data from recording sessions**, as they represent the most natural use of the language (§ 2.1.1). Most of the time and effort spent on this grammar research were devoted to the transcription, translation and analysis of the speech recordings. As detailed in Appendix D, which shows the metadata for those texts that were fully transcribed, translated, analyzed and further annotated, the primary data that these texts are based on consist of a total of 01:21:36 (hh:mm:ss) of media files that include video, and a total of 00:40:02 of audio-only recordings, for a total of two hours. As for the data derived from these media data, the two hours of spoken language are represented by a total of approximately 12,500 transcribed Karbi words.

Context-free elicitation data were collected on phonological and grammatical topics, although grammatical topics were mostly investigated through textual examples and elicitation based on those. An example of context-free elicitation used for this grammar were elicited clauses that included comparative constructions or clauses that included indefinite quantifiers in order to learn more about these particular grammatical constructions. Phonological elicitation was mostly aimed at the Karbi tone system, e.g., eliciting roots that undergo morphophonemic changes due to adding certain prefixes, or elicitation of words with similar tone patterns.

The **lexical database** has been built up parallel to the text database of recording session based data due to the way the software Toolbox, which was used for grammatical annotations, works. Therefore, the lexical database includes all morphemes that occur in the recorded texts with some additional opportunistic data entries. It is a basic lexical database, which will require further work to be usable as a dictionary. As of now, entries generally only consist of the Karbi morpheme, an English gloss, a part of speech label, and in some cases additional grammatical, lexical, or cultural notes.

2.2.2 Access to the corpus

The core text database of all 18 texts (for metadata see Appendix D) which this grammar is primarily based on is freely accessible online in the *Himalayan Linguistics Archive* (Konnerth and Tisso 2018). This document provides the full transcription, translation, and morpheme-by-morpheme glosses of all of these texts.

In addition, all media files are available in the Endangered Languages Archive (ELAR) (Konnerth and Tisso 2019).

2.3 Primary data collection

Data collection was carried out over a total of 15 months consisting of five phases: January–March 2009 (phase 1); February–May 2010 (phase 2); January–March 2011 (phase 3); September–December 2011 (phase 4); September–December 2012 (phase 5).

Specifically, phase 1 mostly consisted of word elicitation aimed at Karbi phonology and simple sentence elicitation without recording much. During phase 2 a festival to celebrate Karbi culture took place in the local capital Diphu. This festival brought a number of village elders to Diphu, about ten of who agreed to be recorded, performing a variety of genres. We were able to record them in a recording studio in Diphu and collected a lot of primary data in the form of recording sessions involving both audio and video data. These data are mostly folk stories as well as some procedural texts. A lot of song data were also recorded during phase 2, which, however, have not been analyzed yet since Karbi song language is entirely different from the ordinary language and requires further research (§ 1.1.3.3).

During phase 3 some additional phonological data were collected. As part of phase 4, a one-week recording trip was carried out in November 2011, which resulted in a number of recorded interviews, conversations, narrations of local histories and folk stories, most of them pertaining to a research project dedicated to investigating the status of women in Karbi society. During the final stage of phase 5, a few other texts were collected, including an on-line narration of the Pear Story (Chafe 1980).

2.3.1 Acknowledgment of Karbi speakers that facilitated data collection

I would like to acknowledge the participation and support from the following people: Maloti Rongpharpi, Kache Rongpharpi, the Assistant Teachers at Lorulangso LP School: Punyalata Ingtipi and Rani Teronpi, Maggie Katharpi, Chandra Kanta Terang, Puspa Engtipi, Kahan Terangpi, and Dr. Janta Pator (Diphu); Burnesh Milikpi, Keshop Terang, Jugal Timung, Rajen Kro, Dim Teron, and Hemari Rongpi (Jyrikyndeng); Aren Ingti, Pretty Ingtipi, and Kamal Chandra Kro (Dongkamokam); Chandra Sing Tisso (Podumsarpo); Harsing Ronghang, Ronghang Lindok and his assistants (Ronghang Rongbong); Longsing Tokbi (Amtereng); Bidyaram Rongpi (Hamren); Dhansing Terang, Kare Rongpipi, and Harsing Kro (Boythalangso); Anjan Teron and Dhaniram Ingleng (Bhoksong); Modon Kro (Rambangla); Sarthe Phangcho (Putsari Hindu arong); Mahin Phangcho (Umpanai); Hem'ari Ingjai (Pharkong Abi); Sika Hansepi (Men Terang); Kache Kropi (Balijuri); Chandra Sing Teron; Kasang Teronpi (Dingso Terang); Rongbang Teron and Seng Tisso (Ujandongka); Welisbon Ronghipi (Umswai Model); Longsing Bey (Murap, Umlaper); Joysing Tokbi (Laru aum, Umlaper); the people of Sohliya (Meghalaya); Owen Terang and the people of Marmein (Meghalaya); Dhiren Ingti and Dhiren Ronghang (Kamrup); Khayasing Hanse (Bokoli); Sarhon Ronghang, KLA; Sangvai Teronpi and Hangmiji Hanse (Hongkram); Dr. Mansing Rongpi, MLA, Dispur; and Father Joseph Teron, Don Bosco. Kardom.