Alice Vittrant, Justin Watkins (Eds.)
The Mainland Southeast Asia Linguistic Area

Trends in Linguistics
Studies and Monographs

Editor responsible for this volume
Hans Henrich Hock

Volume 314

The Mainland Southeast Asia Linguistic Area

Edited by
Alice Vittrant
Justin Watkins

DE GRUYTER
MOUTON

ISBN 978-3-11-040176-9
e-ISBN (PDF) 978-3-11-040198-1
e-ISBN (EPUB) 978-3-11-040213-1

Library of Congress Cataloging-in-Publication Data
Names: Vittrant, Alice, editor. | Watkins, Justin, editor. | Container of (work): Peterson, David A. (David August), 1968- Bangladesh Khumi.
Title: The mainland Southeast Asia linguistic area / [edited by] Alice Vittrant/Justin Watkins.
Description: Berlin ; Boston : Mouton, [2018] | Series: Trends in linguistics. Studies and monographs ; volume 314 | Includes bibliographical references and index.
Identifiers: LCCN 2018007425 | ISBN 9783110401769 (hardcover) | ISBN 9783110401981 (pdf) | ISBN 9783110402131 (e-pub)
Subjects: LCSH: Southeast Asia-Languages.
Classification: LCC PL3501 .M33 2018 | DDC 495--dc23 LC record available at https://lccn.loc.gov/2018007425

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.

Typesetting: jürgen ullrich typosatz, Nördlingen
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

James A. Matisoff

Preface

This book is a welcome addition to the literature on individual East and Southeast Asian languages, as well as an important validation of the concept of *linguistic area*. The 13 articles treat languages belonging to the five great families of the region (Sino-Tibetan, Mon-Khmer/Austroasiatic, Tai-Kadai [=Kradai], Hmong-Mien [=Miao-Yao], and Austronesian), with an explicit emphasis on the manifestation of particular areal features that have been discussed in the literature.[1]

By any criterion – sheer number of languages, significance for the history and culture of the world – East and Southeast Asia constitute one of the world's most important linguistic areas.[2]

• *Multilaterality, directionality, and concentric spheres of influence*

In a delightfully messy linguistic area like mainland Southeast Asia, where both lexical and grammatical homogenization have occurred on a grand scale, it is no simple matter to determine the genetic origin of such features as phonemic tone, quadrisyllabic "elaborate expressions", "psycho-collocations", classifier systems, emphatic sentence-final particles, adversative passives, grammaticalized verbs functioning as aspectual markers or prepositions, etc. The situation is reminiscent of a group of neighbors living cheek-by-jowl, who often squabble with each other, but who have nevertheless borrowed so many cups of sugar from each other over the years that they have lost track of who owes what to whom.[3]

1 The organization of this volume is thus quite different from that of Thurgood & LaPolla, eds. (2003), where some 40 Sino-Tibetan languages (30+ Tibeto-Burman and 7 dialects of Chinese) are featured, but all from a single great language family, and with few explicit references to cross-genetic similarities.

2 The discussion in this section roughly follows the presentation in Ch. VII (*Diffusional dynamics*) of the manuscript of my book *Languages of Mainland Southeast Asia* (still alas, "in progress"!).

3 Terms like *convergence*, *diffusion*, and *contact* are in themselves neutral with respect to directionality. On the other hand, there does exist a term *(parallel independent development)* which specifically denies that a certain phenomenon in Language A has been influenced by contact with Language B.

James A. Matisoff: University of California, Berkeley
E-Mail: matisoff@berkeley.edu

https://doi.org/10.1515/9783110401981-202

For millennia Mainland Southeast Asia (including China, especially the vast area south of the Yangtze) has been marked by voluntary migrations, forced displacements of populations by war or invasions, and intermarriage among different ethnic groups. One result has been widespread multilingualism, to the point where it is common to encounter people with little or no formal education who are fluent in several languages. Although one's mother tongue *(la lingua del cuore)* occupies a special place in people's hearts, the language one must speak in order to earn a livelihood *(la lingua del pane*[4]*)* is equally essential.

For not all languages are equal in terms of cultural and political prestige. The two most important cultural influences on the Southeast Asian linguistic area as a whole have been from India and China, domains of influence which are conveniently referred to as the *Indosphere* and the *Sinosphere*.[5] Of course these spheres overlap geographically, and the proportion of influence that each has had on a particular culture has sometimes greatly changed through time. Roughly speaking, the Southeast Asian Indosphere comprises northeast India, the Himalayan region, Burma, Cambodia, Thailand, and Indonesia (especially Java and Bali). The learned components of the lexicons of the "great" languages of this area (Mon, Burmese, Khmer, Cham, Thai/Lao, Javanese, Balinese) are from Sanskrit/Pali. The vast Sinosphere includes Korea, Japan, and large swaths of Inner Asia (Mongolia, Xinjiang), as well as regions where Southeast Asian languages are spoken: China south of the Yangtze ("Cisyangtzeana") and Vietnam. Southern China is a veritable "mother soup" of languages (W.L.Ballard, 1983),[6] a "corridor of peoples" (Sun Hongkai 1983),[7] that certainly witnessed the birth of Tai-Kadai and Hmong-Mien (Downer 1963),[8] and probably Austroasiatic and Austronesian as well. The learnèd components of the lexicons of Hmong-Mien and Vietnamese are from Chinese.

It is perhaps inevitable to distinguish between "great" languages, spoken by the majority populations of past or present nation-states, and "humble" languages, spoken by minority groups, and often well on the way to extinction. The present volume strikes a nice balance between these two categories, with eight chapters on "great" languages (Chinese, Burmese, Mon, Khmer, Vietnamese, Thai, Cham, Malay), and five on "humble" ones (Khumi, Na/Mosuo, Mong Leng,

4 In the Southeast Asian context *la lingua del riso* might be more appropriate.

5 These terms were first introduced in Matisoff 1978, "Tonogenesis in Southeast Asia".

6 "Mother soup – a south Chinese recipe for tonometamorphogenesis", *Computational Analyses of Asian and African Languages* (Tokyo) 22: 65–70.

7 [in Chinese] "The languages of the Corridor of Nationalities region of Western Sichuan." *In Research on the Peoples of the Southwest*, No. 1, pp. 429–454. Chengdu: People's Publishing Co.

8 "Chinese, Tai, and Miao-Yao." In Harry L. Shorto, ed., *Linguistic Comparison in South East Asia and the Pacific*, London: School of Oriental and African Studies, pp. 133–139.

Pwo Karen, Wa). Another important dichotomy is related to this distinction, but grounded in geographical/ecological factors: *plains languages* (spoken by majority populations in large contiguous geographical areas in the plains, where irrigated ricefields are possible) vs. *hill languages* (spoken by small relatively isolated groups of people in the hills, who live by practicing "slash-and-burn" agriculture in non-irrigated mountain fields or "swiddens").[9]

On the sociolinguistic level, great languages tend to reflect hierarchically stratified cultures, often with clearly distinguished formal vs. colloquial styles of speaking, including such features as elaborate pronominal systems that encode differences in age, gender, and social status. The languages of the hills are more "democratic", with simple pronouns and few euphemisms for embarrassing bodyparts or bodily functions.

The macro-distinction between Indo- and Sino-spheres is valid as far as it goes, but it is too simplistic when it comes to detailed examination of the influences exerted on particular modern languages. This is because there are a large number of regionally dominant languages that have developed spheres of influence of their own.[10] *Tibetan* has exerted a strong influence on the Tibeto-Burman languages of Nepal and southwest China. *Nepali*, the national Indo-Aryan language of Nepal, has had an even more powerful influence on the TB languages of the country. *Jingpho* (also called Kachin), itself influenced by Shan, Burmese, and Chinese, has strongly affected Burmish languages like Atsi/Zaiwa, Maru/Langsu, and Lashi/Leqi, to the point where they are considered in China to belong to the "Jingpho nationality". *Meithei* (also called Manipuri) is the dominant language of India's Manipur State, and is rapidly supplanting the other TB languages of the area. *Tai* languages (including standard Thai or Siamese, as well as closely related languages like Shan and Lue) have a wide area of influence on minority languages of Thailand, northern Burma, and southwest China.[11]

The languages represented in this volume may be differentiated according to the particular sphere(s) of influence that have affected them:

---

9 See Burling (1965) *Hill Farms and Paddy Fields*; New Jersey: Prentice Hall. This state of affairs has been called "vertical ecological zonation".

10 This is analogous to the situation with respect to Chinese dialects, where there is often a three-tiered system in a speaker's repertoire: the local "patois", spoken at home in intimate situations; the regional standard, used in more formal contexts outside the home; and the national language, used when communicating with people from other parts of the country.

11 SHINTANI Tadahiko (1998) has called this whole region the "Tai cultural area". See, e.g. "The zone of the Golden Quadrangle: the history, language, and ethnic groups of the Tai cultural area." Tokyo: Keiyusha. [in Japanese]

Sinosphere	**Indosphere**	**Tibetosphere**	**Taiosphere**
Chinese "dialects"	Burmese	Mosuo	Pwo Karen
	Cham		Wa
Mong Leng	Khmer		
Mosuo	Khumi Malay		
Thai	Mon		
Vietnamese	Pwo Karen		
	Thai		

Types of lexical borrowing situations
It is generally agreed that lexical borrowing involves the most superficial area of linguistic structure, yet it is by no means a simple matter. Several subtypes should be recognized:

(a) Unidirectional borrowing
Sometimes the direction of borrowing is clear. It is obvious that Japanese **pan** 'bread' is borrowed from Portuguese **pão**, rather than the other way around, since bread was not part of the Japanese diet before European contact. It is clear that Jingpho **yàk** 'difficult, distressful' is a loanword from Shan (cf. Siamese **jâak**), since no native Jingpho words end in **-k** (PTB ***-k** > Jg. **-ʔ**).

(b) Backloans
Much trickier are cases where Language B borrows a word from Language A at one historical period, then later A borrows the same etymon – which may meanwhile have undergone considerable change in the course of its history in Language B – back from B again. The case of the everyday condiment *ketchup* furnishes an interesting example. This word probably originated in a Cantonese compound **ke-chap** 茄汁 'eggplant [i.e. tomato] sauce', perhaps first borrowed into Malay as **kicap** 'fish sauce', whence to English *catsup* or *ketchup*, from where it was adopted as Japanese **kechappu**, then back into Taiwanese Chinese as **ket-chap-pu**, thus returning to its language of origin.

(c) Borrowing from related languages
In cases where the languages in contact are also genetically related, it may be particularly hard to distinguish true cognates from diffusional material. Lahu **ɔ̀-vɛ̀ʔ** 'screw' can easily be shown to be a borrowing from Burmese **wɛ́ʔ-ʔu** (Written Burmese **wak-ʔu**), literally "pig intestine" (so called because of its spiral shape), although most Lahu are unaware of this, since the ordinary Lahu word for 'pig' is **vàʔ** (inherited from Proto-Lolo-Burmese ***wak**), and the head of

the Burmese compound, **ʔu** 'intestine', has been dropped. The Lahu forms **và?** and **vɛ̀?** are thus doublets, one inherited and one borrowed.

(d) Borrowings from a common source
Lexical similarities between languages may be due to borrowing from a common source. Resemblant morphemes in Tai and Vietnamese are largely due to independent borrowings from Chinese. In psycho-collocations (one of the areal features emphasized in this volume), the "psycho-noun" for MIND derives from Pali **citta** in the literary languages of the Indosphere: Written Burmese **cit** (Mod. Bse. **sei?**), Khmer **ceut**, Mon **cʊt**.

(e) Reversibility of the directionality of influence
The vicissitudes of cultural and military history have sometimes led to striking reversals in the direction of influence between languages. English borrowed massively from French for centuries after the Norman Conquest (A.D. 1066), a situation which began to change in the 19th century, and which by now has reached alarming proportions as English words are flooding into the hallowed French lexicon. Japanese borrowed enormously from Chinese for hundreds of years, but this state of affairs changed radically in the latter half of the 19th century, when Japan was exposed to Western influence before China was. Many modern scientific and cultural terms were translated into Japanese by arranging Chinese morphemes (along with the characters used to write them) into novel combinations. Subsequently a good number of these were borrowed back into Chinese[12], e.g. 'university' Jse. **daigaku** > Chinese **dàxué** 大學; 'diabetes' Jse. **tōnyōbyō** > Chinese **tángniàobìng** 糖尿病. In Southeast Asia proper, a similar historical scenario affected the two principal Indospheric Mon-Khmer literary languages, Khmer and Mon. For centuries these two languages exerted strong influence on Thai and Burmese, respectively, only to have the directionality of influence reversed in recent times.

• *Scale of intensity of contact*
The publication of Thomason & Kaufman's seminal book *Language Contact, Creolization, and Genetic Linguistics* (1988), mentioned in the Editors' *Introduction* to this volume, went a long way toward making the study of degrees of language contact more precise. The authors established a scale from 1 to 5 according to the intensity of the contact involved, providing rich exemplifications of each point on the scale ranging from #1 (casual contact; only lexical borrowing) to #5 (very strong contact with heavy structural borrowing).

12 See section (b), *Backloans*, above.

• *Structural borrowing and metatypy*

In the not too distant past, the possibility of grammatical borrowing seemed like a rather daring notion. One of the questions on my doctoral orals (Fall 1964) was "Is structural borrowing possible?" The belief was widespread that morphosyntax is somehow the deepest level of linguistic structure, and thus virtually impervious to outside interference. Nowadays, careful study of borrowing patterns all over the world has shown that "While there may be some aspects of a language's syntax which, because of internal structural cohesion, are especially resistant to foreign interference, the evidence…indicates that syntactic interference is as common as phonological interference" (Thomason & Kaufman, *op. cit.*, p. 118).

The Southeast Asian linguistic area furnishes us with a number of especially striking cases of *metatypy*, i.e. typological change resulting from intensity of contact "to the fifth degree", such that areal influence has cut across genetic relationship. These cases include the following:

	Genetic	**Typological**
Karen	Tibeto-Burman	Tai; Mon-Khmer
Bai	Tibeto-Burman	Sinitic
Vietnamese	Mon-Khmer	Sinitic; Tai
Munda	Austroasiatic	Indo-Aryan
Cham, Rade	Austronesian	Mon-Khmer < Sinitic

Unlike the vast majority of Tibeto-Burman languages, the Karenic and Baic subgroups have acquired non-verb-final word order, under the influence of Tai/Mon-Khmer and Chinese, respectively. Vietnamese has changed from an atonal sesquisyllabic language to a fully tonal monosyllabic one, due to Chinese (and to a lesser extent, Tai) influence. The Austroasiatic Munda languages of eastern India have undergone a change from the non-verb-final syntax typical of AA/MK to a verb-final one, with the verb often accompanied by a string of suffixes alien to the rest of the family, suggesting some sort of Indo-Aryan substratum. The Chamic languages of southern Vietnam have changed from their earlier disyllabic or sesquisyllabic structure to a monosyllabic pattern, perhaps in part due to influence of Vietnamese (which had itself undergone monosyllabization under Chinese influence).

• *Some areal features emphasized in this volume, and the languages/articles which give the most information about them*

(1) Phonological
tone and register; monosyllabicity; sesquisyllabicity; profusion of vowels; relative paucity of syllable-final consonants

(2) Morphological
psycho-collocations; elaborate expressions; expressives; reduplication; compounding; lack of inflection; affixation

(3) Grammatical categories and syntax
classifiers; nominalization and relativization; pro-drop; emotive sentence-final particles; topic prominence; verb concatenation and serialization; grammaticalization of verbs to prepositions, auxiliaries, TAM particles, etc.; adversative passives

(4) Semantics/Pragmatics
proliferation of terms in certain semantic areas: *rice; cut; carry; dry*; politeness distinctions in pronouns; pronominal use of kin terms; literary vs. colloquial style

The extent to which each of the languages in this volume exemplify these features is roughly indicated in the following charts:[13]

	Tone (T) Register (R)	**Mono-syllabic**	**Sesqui-syllabic**	**Vowel profusion**	**Paucity of codas**
Min Chinese	(T) ++	+	–	+	+
Burmese	(T, R) +	+	+	–	+
Yongning Na	(T) +	+	–	+	+
Khumi	(T) ++	+	+	+	+
Pwo Karen	(T) +	+	+	+	+
Khmer	–	+	++	++	+
Mon	(R) +	+	+	+	+
Vietnamese	(T) ++	+	–	+	+
Wa	(R) +	+	+	+	+
Siamese	(T) +	+	+	+	+
Mong Leng	(T,R) ++	+	(+)	+	++
Cham	(R) +, (T) (+)	+	(+)	+	+
Malay	–	–	–	(+)	–

13 A similar scheme of pluses and minuses is also followed in the Vietnamese paper by Do-Hurinville & Huy Linh Dao.

	Psycho-collocations	Elaborate expressions	Compounds	Redup-lication	Expressives	Affixa-tion	Inflection
Min Chinese	+	+	+	+	–	–	–
Burmese	+	+	+	++	–	+	–
Yongning Na	+	+	+	+	–	+	–
Khumi	(+)	++	+	(+)	–	+	+
Pwo Karen	+	+	+	+	–	(+)	–
Khmer	+	–	–	+	++	+	–
Mon	+	(+)	(+)	+	+	+	–
Vietnamese	++	+	+	+	+	–	–
Wa	+	+	+	+	+	–	–
Siamese	++	+	+	+	+	–	–
Mong Leng	+	+	+	++	++	–	–
Cham	(+)	++	+	+	?	–	–
Malay	++	+	+	+	++	+	+

	Topic prominence	Classifiers	Nominalization & relativization	Pro-drop
Min Chinese	+	+	+	+
Burmese	+	+	+	+
Yongning Na	+	++	?	+
Khumi	+	(+)	+	+
Pwo Karen	+	+	+	+
Khmer	+	(+)	?	+
Mon	+	(+)	?	+
Vietnamese	+	+	?	+
Wa	+	(+)	?	–
Siamese	+	++	+	+
Mong Leng	+	++	?	+
Cham	+	(+)	?	+
Malay	+	++	?	+

	Grammatical-ization of verbs[14]	Verb concatenation and/or serialization	Adversative passives	Emotive final particles
Min Chinese	+	+	+	+
Burmese	+	+	–	+
Yongning Na	++	++	?	?
Khumi	–	–	?	?
Pwo Karen	+	++	?	++
Khmer	+	+	+	?
Mon	+	+	+	–
Vietnamese	+	+	+	++

14 To prepositions, auxiliaries, TAM particles, etc.

	Grammaticalization of verbs[14]	Verb concatenation and/or serialization	Adversative passives	Emotive final particles
Wa	++	+	?	–
Siamese	+	+	+	+
Mong Leng	+	+	+	++
Cham	+	+	?	++
Malay	(+)	+	+	+

	Literary vs. colloquial styles	Politeness distinctions in pronouns	Pronominal use of kinterms	Proliferation of culturally key words[15]
Min Chinese	+	+	+	+
Burmese	+	+	+	+
Yongning Na	–	–	+	+
Khumi	–	–	?	+
Pwo Karen	–	+	+	+
Khmer	+	+	+	?
Mon	+	–	+	+
Vietnamese	(+)	+	+	++
Wa	–	–	+	++
Siamese	+	++	+	+
Mong Leng	–	–	(+)	+
Cham	+	+	+	+
Malay	+	+	+	+

• *Other specially developed/interesting features of the individual languages*

Min Chinese (Chappell)
- complex tone sandhi
- voiced initial stops < older nasals (similar to Lahu)
- strong relationship between relativization and nominalization
- Southern Chinese dialects like Min share many features with Tai, while the Northern ones share features with Altaic

Burmese (Vittrant)
- interesting patterns of reduplication
- dozens of simplex/causative pairs distinguished by aspiration
- rich systems of pragmatic particles and classifiers
- strong Pali component in abstract lexicon and high style

15 For verbs like *cut*, *carry*, *dry* and nouns like *rice*.

Yongning Na/Mosuo (Lidz)
- strong contact with Bodic (Tibetospheric) languages: conjunct/disjunct system, emphasis on evidentiality, agentive marking/non-systemic ergativity
- speakers shifting to Mandarin
- basic tense distinction is between future and non-future
- three semantic types of possession: alienable, inalienable, and inabsoluble (the latter only for body-parts)

Bangladesh Khumi (Peterson)
- widespread use of tone for morphosyntactic purposes
- robust nominal case morphology
- semantically and syntactically complex sequences of post-verbal elements, divisible into several "zones" for analytic purposes
- reasonable amount of verbal inflection, including participant-coding prefixes

Pwo Karen (Kato)
- striking example of metatypy with respect to the rest of Tibeto-Burman: SVO syntax
- relative clauses may appear either before or after their head (as, e.g. in Lahu)
- elaborate systems of serial verbs, both concatenated and separated
- negation via a final particle rather than a pre-verbal adverb

Khmer (Haiman)
- register system has disappeared; infixation as well as prefixation
- a stronghold of sesquisyllabism; strong relationship between reduction of compounds and the creation of sesquisyllables, an ongoing process
- strong history of contact with Thai, which has led to much grammatical convergence as well as many lexical borrowings in both directions
- much word-play, rich ideophones, expressive onomatopoeia

Mon (Jenny)
- register survives, unlike in Khmer
- "class terms" appear as first elements in compounds for fruits, fish, plants, birds, etc. (as in Thai, Mong Leng, and Cham)
- in psycho-collocations there is a semantic difference according to the relative position of the psycho-noun and its mate (again as in Thai)
- many elaborate expressions, ideophones, euphonic compounds

Vietnamese (Do-Hurinville & Dao)
- huge phonological, lexical, and grammatical influence from Chinese, leading to tonogenesis and monosyllabization
- order of the elements in a classifier phrase is like Chinese: Num + Clf + N
- extensive repertoire of verbs for *cutting* and *carrying*
- rich psycho-collocations, especially with the favorite psycho-noun *lòng* 'entrails'

Wa (Watkins)
- variable word-order: VSO as well as SVO
- dual/plural and inclusive/exclusive distinctions in pronouns, as in many Himalayan TB languages
- tripartite compounds of the form *head-noun + transitive verb + object*, e.g. 'iron' ("machine-press-clothes"); just as in Thai, e.g. **phâa-chét-tua** 'towel' ("cloth-wipe-body")
- pro-drop is relatively rare

Thai (Jenny)
- very rich classifier system, although the younger generation is tending to lose some of the finer distinctions
- elaborate distinctions in pronouns, taking account of relative age, status, gender; the formal register of the language includes special locutions used for monks and royalty
- lexical influence from Pali, mediated by Khmer
- polite final particles, with compulsory gender distinctions for the first person

Mong Leng (Mortensen)
- extremely complex tonal system, involving both pitch and voice quality; many syllable onsets, but no codas (except **-ŋ** by one analysis)
- "class terms" in compounds (as in Thai, Mon, and Cham)
- classifiers play an important role in the syntax
- serial verb constructions are highly developed

Cham (Brunelle and Phu Van Han)
- becoming totally monosyllabized, perhaps because of Vietnamese influence
- long written tradition, but script is now rarely used
- Old Cham prefixes and infixes have been lost, replaced by compounds and other periphrastic structures
- has three implosives, including a palatal one (besides the usual labial and dental ones)

Malay (Nomoto & Hooi Ling Soh)

- no tones or registral differences
- highly developed systems of classifiers and psycho-collocations
- adversative passive (as in Thai, Vietnamese, Mon, Khmer, Mong Leng, Hokkien)
- mostly dissyllabic structure, with no sesquisyllabization

***

The expertise of many specialists has contributed to the excellence of this volume, which is destined to become a *livre de chevet* for everyone interested in the fascinating languages of the mainland Southeast Asian linguistic area.

Table of Contents

Maps

Alice Vittrant and Justin Watkins

Introduction: Languages of the Mainland Southeast Asia linguistic area – Grammatical Sketches

1 Areal linguistics

Thomason and Kaufman's 1988 book *Language contact, creolization, and genetic linguistics* had a stimulating effect on the fields of comparative and descriptive linguistics and inspired a number of studies on various topics related to language contact: the relationship between typology and language contact; the effect of language contact on a language's genetically inherited characteristics, and work on mixed and endangered languages. More generally speaking, the increased availability of data relating to language contact has enabled wider-ranging discussion on the nature of language contact and its consequences (see Hickey 2010 for a more detailed account of these subjects).

Within this landscape, our book lies at the crossroads of the following themes:

(1) vulnerable and endangered languages, since some of the languages described here are minority languages losing ground under the linguistic influence of dominant neighbouring languages (see chapters on Cham, Wa); (2) areal typology, since our book is concerned with one area in particular: Mainland Southeast Asia (hereafter MSEA); (3) language contact and genetic affiliation, since the various grammatical sketches lay emphasis on characteristics shared by unrelated languages. This in turn raises the question of how such traits are acquired and how they spread, though neither of these two issues is addressed in this volume, at least from a typological perspective.

Specifically, we address here the issue of linguistic area or *Sprachbund*. This issue is closely associated with language contact and has been discussed extensively over the past fifteen years (see Thomason 2000, 2001; Muysken 2000; Stolz 2002; Heine & Kuteva 2005; Campbell 2006; Bisang 2006a; 2006b; 2006c; Matras et al. 2006; Matras & Sakel 2007; Muysken 2008; and Bisang 2010, *inter*

Alice Vittrant: Aix-Marseille Université / CNRS-DDL (UMR 5596)
E-Mail: alice.vittrant@cnrs.fr
Justin Watkins: SOAS, University of London
E-Mail: jw2@soas.ac.uk

https://doi.org/10.1515/9783110401981-001

alia). Linguistic area, or *Sprachbund* in German, is a concept which was introduced to linguistics in the 1930s by NS Trubetzkoy, in response to the need to account for the linguistic situation observed initially in the Balkans (Sandfeld 1930) and in India (Emeneau 1956; Masica 1976 etc.). In each of these regions both languages with a close genetic relationship and languages from different language families were found. However, despite lacking a common origin, these languages had surprising structural similarities, apparently acquired in part through contact with structural linguistic features that they did not originally possess.

This concept of linguistic area has triggered much debate among linguists interested in language contact. For some authors (e.g. Aikhenvald & Dixon 2001) the difficulty in distinguishing what is inherited through parentage from what is diffused through contact casts doubt on the possibility of establishing genetic parentage in language contact situations, for it may be impossible to determine whether shared traits arise through inheritance, distribution, independent parallel development or by accident.

For other linguists, the concept of linguistic area is viewed in the light of more general work on language change and the constraints on language change associated with language contact (e.g. Gilbers *et al.* 2000; Heine & Kuteva 2005; Aikhenvald & Dixon 2006b; Matras *et al.* 2006; Siemund & Kintana 2008). Studies in recent years concur on the idea that a linguistic area is not a purely linguistic phenomenon, but also brings into play history and culture (Aikhenvald & Dixon 2001: 11–13; Dahl 2001: 1458, Thomason 2001: 104).

> "[Linguistic areas] arise in any of several ways—through social networks established by such interactions as trade and exogamy, through the shift by indigenous peoples in a region to the language(s) of invaders, through repeated instances of movement by small groups to different places within the area." (Thomason 2001: 104)

In other words, the conditions which give rise to language contact (bilingualism, diglossia) are not sufficient in themselves to cause the emergence of a linguistic area. The linguistic communities involved must also share a common culture or a common history—and they are generally aware of this fact. Thus, even if it is not possible to determine when changes in languages occur, it may be possible to identify the factors which favour the emergence of a linguistic area, namely extralinguistic socio-linguistic factors such as culture or social organization, community type, history, politics, geography (for instance 'spread zone' vs. 'residual zone' – cf. Dahl 2001: 1461), population density and diversity, etc.

Finally, there is much published literature on the status of certain geographical areas where languages come to share common properties without being closely related to one another, such as the Balkans (van der Auwera 1998),

Meso-America (Stolz & Stolz 2001), the 'Baltic area' (Koptjevskaja-Tamm 2006), Africa (Heine et Nurse 2008), as well as South Asia (Masica 1994, Ebert 2001).

In recent decades a number of definitions of a linguistic area or Sprachbund have been proposed (see Campbell 2006; Stolz 2006: 33), all of which aim to describe the phenomena of linguistic convergence, common innovations or common retentions, or to identify the specific properties which set a *Sprachbund* apart from other language-contact situations. In general, such accounts invoke the same key concepts, namely (1) a geographical area; (2) the involvement of a number of languages (at least three); (3) shared linguistic characteristics; (4) convergence as a result of contact; (5) convergence not by accident; (6) convergence not as a result of shared heritage (Muysken 2008: 3). However, despite numerous attempts to define the concept precisely, a consensus emerges on the impossibility of identifying universal criteria (Stolz 2002, 2006; Bisang 2006c). Some linguists go so far as to suggest the outright abandonment of the term of *Sprachbund*, referring as it does not to a real object but to a projection by linguists (Stolz, 2002: 260), an *a posteriori* construction "based on the accumulation of residue and borrowed traits, regardless of how and when they came to be shared among the languages involved" (Campbell 2006: 14).

Meanwhile, faced with an unsatisfactory definition which is unlikely to produce a concrete generalization (see Dahl 2001: 1457–8), and because of the impossibility of distinguishing between a situation of linguistic borrowing and one of formation of a linguistic area (Sakel & Matras 2008), other linguists have favoured replacing the notion of linguistic area with less constraining concepts such as 'contact superposition zone' (Koptjevskaja-Tamm & Wälchli 2001) or 'zones of contact-induced structural convergence' (Bisang 2006c).

The problems encountered in seeking to define a linguistic area, however, do not detract from the relevance of studies of changes induced by contact between the languages spoken within a particular geographical area, i.e. areas of linguistic convergence. In the case of Southeast Asia, a region characterized by the presence of five language families and several millennia of contact between the area's linguistic communities, an areal approach is fruitful both for the description of undescribed languages and for typological studies.

2 Establishing the existence of the Mainland Southeast Asian linguistic area

Assessing the extent of our knowledge of the languages of Mainland Southeast Asia, we notice that recent publications on Asian languages are often geneti-

cally oriented, with the exception of Goddard (2005), a nice overview of the linguistic situation in Asia with partial information on each language. A number of recent publications do not address the MSEA *Sprachbund*, for instance Thurgood & LaPolla (2003) on Sino-Tibetan languages; Adelaar & Himmelmann (2005) on Austronesian languages; Diller, Edmondson & Luo (2008) on Tai-Kadai languages, and Jenny & Sidwell (2015) on Austroasiatic languages.

The Mainland Southeast Asian *Sprachbund* inhabits a geographical area stretching from the easternmost fringes of India in the west to China in the east, encompassing the peninsular Southeast Asian states of Burma, Thailand, Laos, Cambodia and Vietnam, as well as peninsular Malaysia. Five different language families are present in the area (Austroasiatic/Mon-Khmer, Tai-Kadai, Hmong-Mien, Sino-Tibetan and Austronesian). It is only relatively recently that this area has been described as a linguistic area or *Sprachbund* (Matisoff 1991, Bisang 1996, etc.), even though now the idea is firmly attested (and uncontroversial) since Enfield 2005.[1] Thus, new general publications on linguistic areas do cite the (M)SEA area as one of the recognized *Sprachbünde* (see for instance Heine & Kuteva 2005: 203), although it was not mentioned in previous works (see Campbell 1994, Feuillet 2001).

In the final analysis, we can say of a linguistic area that "in the absence of a universally valid numerical value of shared isoglosses, language contact situations lend themselves to a classification as a *Sprachbund* if the absolute number of shared isoglosses with no genetic basis among the members of the *Sprachbund* significantly exceeds the number of such isoglosses they have in common with languages outside the *Sprachbund*" (Stolz 2006: 36).

We have adopted an empirical and – we hope – more promising approach towards the study of the MSEA linguistic area. Previous studies such as Matisoff (1986), Migliazza (1996) and Enfield (2005) draw together features that cut across the genealogical phyla in many domains. These works cast light on phenomena or structural properties that cluster around this geographical area, offering an explanation for the distributional asymmetry that cannot be accounted for in terms of linguistic genealogy.

By adopting an approach that is similar to an isogloss approach rather than a geographical, historico-cultural or communicative approach (Stolz 2006: 36), we aim to determine the maximal distribution of single features. To put it in other words, one task of this book is to provide new information about the limits

1 Enfield (2001, 2005) provides a very well-documented argumentation even though he is not alone in having pointed out the strange similarities between unrelated languages in the area. See also Henderson (1965), Matisoff (1986: 75–80), Matisoff (1991), Bisang (1996), Migliazza (1996).

of the contact-induced convergence area (see for instance the chapters on Yongning Na (Mosuo), Khumi or Southern Min) in Southeast Asia, examining the geolinguistic distribution of the given features (or zone of overlap of several such features) that shape the linguistic area.

3 Structure of the book

The structure of this book is intended to be in line with existing typological studies of specific grammatical phenomena such as Kahrel and van den Berg's (1994) work on negation, or cross-linguistic studies such as Aikhenvald & Dixon (2003, 2006), Zúñiga & Kittilä (2010), and Kopecka & Narasimhan (2012) *inter al*. Each of these publications is a collection of chapters which adopt a common format, structure and/or theoretical approach.

The present volume on languages of MSEA has been conceived in a similar way, to allow researchers to do cross-comparisons and to facilitate such comparisons by ensuring that all the chapters have a broadly similar organization and structure and use similar terminology. Our aim in adopting this common approach is to allow specific linguistic phenomena to be studied across a range of languages.

Each contributor was asked to compile a grammatical sketch of a MSEA language following the same guidelines, to allow the reader to navigate easily between and across chapters and languages. A copy of the guidelines which each author was asked to follow is included as an appendix at the back of the book.

Thus, the typological descriptions of the languages are intended to have a common structure to facilitate comparison and to highlight, on the one hand, the unique typological features of the language and, on the other, the features shared with other languages in the area.

Beginning with general information about the language — its affiliation, its geographical location, relevant ethno-linguistic information—the chapters then provide information on (I) phonology, (II) morphology, (III) grammar and syntax, (IV) semantics and pragmatics. Each section examines in turn the features known to be shared by the languages of the *Sprachbund*. Thus, in the section on Phonology, the stereotypical phonological features observed in languages of the area are as follows: complex vowel systems (diphthongs, a large number of vowels, contrastive vowel length), tone or register systems (or a combination of both), a restricted set of final consonants; a restriction on consonant clusters (see Enfield 2005: 186ff). Each contributor was asked to address the question: "To what extent does the language conform to this stereotype (or not)?" by pro-

viding a description of the phonological system and syllable structure, keeping in mind (as a secondary guideline) the correlation between the two as highlighted by Henderson (1965).

In the second section of each chapter, which is on Morphology, the contributors were asked to examine the morphological structure of words, the generally observed tendency toward monosyllabicity, the presence of sesquisyllables, the lack of extensive inflectional morphology, and the use of compounding (rather than derivation). Expressives (psycho-collocation, elaborate expressions) and reduplication are also phenomena examined as shared features across the MSEA area.

The third section, on Grammar and Syntax, is divided into three parts devoted, respectively, to (1) the nominal domain, (2) the verbal domain and (3) the clausal domain. Besides the basic structure of the NP, we also asked for a close investigation of classifier systems (see Grinevald 1999, 2000). Next, MSEA languages being famous for their lack of inflection, the description of the verbal domain relies on notional categories expressed in the verbal phrase, grammaticalization (see Matisoff 1991), and serial verb constructions (see Bisang 1991, Bisang 1996, Durie 1997, Aikhenvald & Dixon 2006, Vittrant 2006). In terms of sentence organization, the basic word order is (S)VO for MSEA languages, except for Tibeto-Burman languages in the area. All languages in the area show a characteristic information structure, being 'topic-prominent' rather than 'subject-prominent'. All languages in the area are also known for widespread ellipsis of definite arguments, regardless of grammatical role. These phenomena are investigated in each language.

Marybeth Clark (1985) also noticed that many MSEA languages share a similar device for asking questions that request an affirmative or negative response rather than other information, that is to say for wh-questions or yes-no-questions. Matisoff (1986: 78) noticed that MSEA languages have a penchant for nominalizing whole sentences without embedding them into any larger unit, typically via a particle, which is also used in citation-form verbs, and which has a relativizing/genitive function in other constructions. These two phenomena are examined by some contributors to the present volume (Chappell, Lidz, Peterson).

Regarding the last section on Semantics and Pragmatics (IV), MSEA languages seem to share basically similar conceptual frameworks about humans and nature (Matisoff 1986: 79). Thus the comparative semantics domain needs to be investigated. As for pragmatics, MSEA languages have systems of sentence-final particles as a basic mode of distinguishing illocutionary force (such as requesting, questioning, persuading, advising, reminding, instructing, etc.), but also for expressing 'propositional attitudes', that is to say emotions of the speaker (such as surprise, doubt, impatience, reluctance, hesitation, etc.). Lastly, some languages may have developed systems for encoding politeness (i.e. formality vs. intimacy with an interlocutor) and honorific systems (to ex-

press respect or reverence towards the referent). All these phenomena are explored in each of the thirteen languages described.

Each grammatical sketch ends with two appendices: a glossed text, and a table summarizing the features examined in the chapter. This glossed text offers a glimpse of the language used in more natural context and a demonstration of the broader expertise of the linguists who compiled them. A series of maps at the end of the book show the geographical distribution (in terms of the location of the language described in this book) of a number of the more linguistic features explored in the tables.

In conclusion, thirteen languages of MSEA are described in this collection. As with any linguistic survey, some explanation of why we chose that number and that particular selection of languages is required.

The languages in our sample were selected for a number of reasons. In part, languages such as Khmer, Wa, Vietnamese, Thai and Hmong are representative of the Mainland Southeast Asian peninsula, which is known to be the geographical core of this linguistic area, from which certain features diffused outwards. In contrast, languages such as Burmese, Khumi, Mosuo and Min are spoken on the periphery of the same area.

Any examination of the traits shared by the languages of Southeast Asia must also include some examples of languages which are representative of other contrasting phenomena, such as:

(a) size of speaker population – i.e. major or national languages (Malay) vs. minor or marginal languages (Cham);
(b) dominant lowland languages (Vietnamese or Khmer) vs. languages of highland minorities (Wa);
(c) languages with a long-established literary tradition (Burmese) vs. languages with a predominantly oral tradition or which are unwritten or seldom written (Mosuo).

These are also factors which have determined the choice of languages in this collection.

Finally, we would have preferred to be able to include, for each language family, at least one major and one minor language, which has unfortunately not been possible for two of the language families, namely Tai-Kadai or Hmong-Mien (a family whose very size makes an objectively major language difficult to identify).

The chapters are organized according to language family: five Sino-Tibetan languages; four Mon-Khmer (Austroasiatic) languages; two Austronesian and one each of Tai-Kadai and Hmong-Mien.

We might reasonably be accused of arbitrariness in our selection of languages; in fact our choice was intended to be neither comprehensive nor sys-

tematically representative of the area. Rather, we wanted to give an impression of the colours and flavour of the region, based on linguistic commonalities and differences. Given that MSEA is a recognized linguistic area (or, a zone of contact-induced structural convergence) in the literature on language contact and Asian languages more generally, our goal was not to prove its existence. Rather, it was an attempt to further explore the boundaries of the area and the path of diffusion of shared linguistic features.

At the same time, the diversity represented in our selection of languages entails a concomitant variability in the depth of the descriptions in the chapters of this book: for certain languages, the authors have the benefit of a wealth of previous linguistic investigations, while for others the authors are amongst the first linguists to describe the language in question. In all cases, however, the authors are actively involved in research on the languages they have described and the data is not taken from secondary literature.

On that note, we are very grateful to the consultants who have collaborated with the authors of each chapter. Without such collaboration or the readiness of linguists to engage in fieldwork on this kind, a book such as this would not have been possible. Lastly, we hope to inspire further work on the many languages of Southeast Asia which have yet to be described.

References

Adelaar, K. Alexander & Nikolaus Himmelmann. 2005. *The Austronesian languages of Asia and Madagascar*. London: Routledge.

Aikhenvald, Alexandra Y. & Dixon R. M. W. (eds.). 2001. *Areal diffusion and genetic inheritance: Problems in comparative linguistics,* Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. & Dixon R. M. W. (eds.). 2003. *Studies in evidentiality*. Amsterdam/Philadelphia: Johns Benjamins.

Aikhenvald, Alexandra Y. & Dixon R. M. W. (eds.). 2006. *Grammars in contact: A cross-linguistic typology*. New York: Oxford University Press.

Aikhenvald, Alexandra Y. & Dixon R. M. W. (eds.). 2006. *Serial verb constructions: A cross-linguistic typology*. New York: Oxford University Press.

Bisang, Walter. 1991. Verb serialization, grammaticalization and attractor posi-tions in Chinese, Hmong, Vietnamese, Thai and Khmer. In Hansjakob Seiler & Waldfried Premper (eds.), *Partizipation*, 509–562. Tübingen: Gunter Narr.

Bisang, Walter. 1996. Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. *Studies in Language* 20 (3). 517–597.

Bisang, Walter. 2006a. Contact-induced convergence: Typology and areality. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, vol. 3, 88–101. Oxford: Elsevier.

Bisang, Walter. 2006b. South East Asia as a Linguistic Area. In Keith Brown (ed.), *Encyclopaedia of language and linguistics*, vol. 11, 587–595. Oxford: Elsevier.

Bisang, Walter. 2006c. Linguistic areas, language contact and typology: Some implications from the case of Ethiopia as a linguistic area. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas. Convergence in historical and typological perspective*, 75–98. Hampshire: Palgrave Mac Millan.

Bisang, Walter. 2010. Areal language typology. In P. Auer & J. E. Schmidt (eds.), *Language and space. An international handbook of linguistic variation – Vol. 1: Theories and methods*, 419–440. [Handbücher zur Sprach- und Kommunikationswissenschaft]. Berlin & New York: Mouton de Gruyter.

Campbell, Lyle. 1994. Grammar: Typological and areal issues. In R. E. Asher & J. M. Y. Simpson (eds.), *Encyclopedia of language and linguistics*, vol. 3, 1471–1474. London: Pergamon Press.

Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In Yaron Matras, April M. S. McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*. New York: Palgrave Macmillan.

Clark, Marybeth. 1985. Asking questions in Hmong and other southeast Asian languages. Linguistics of the Tibeto-Burman area 8 (2). 60–67.

Dahl, Östen. 2001. Principles of Areal Typology. In M. Haspelmath, E. König, W. Österreicher & W. Raible (eds.), *Language typology and language universals*, vol. 2, 1456–1470. Berlin & New York: Mouton de Gruyter.

Diller, Anthony, Jerold A. Edmondson & Yongxian Luo (eds.). 2008. *The Tai-Kadai languages*. London & New York: Routledge.

Durie, Mark. 1997. Grammatical structures in verb serialization. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Complex predicates*, 289–354. Stanford: CSLI Publications.

Ebert, Karen. 2001. Südasien als Sprachbund. In M. Haspelmath, E. König, W. Österreicher & W. Reible (eds.), *Language typology and language universals*, vol. 2, 1529–1539. Berlin & New York: Mouton de Gruyter.

Emeneau, Murray B. 1956. India as a linguistic area. *Language* 32. 3–16.

Enfield, Nick J. 2001. On genetic and areal linguistics in Mainland South-East Asian: Parallel polyfunctionality of 'acquire'. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 255–290. Oxford: Oxford University Press.

Enfield, Nick J. 2003. Linguistic Epidemiology: Semantics and Grammar of Language Contact in Mainland southeast Asia, London: Routledge.

Enfield, Nick J. 2005. Areal linguistics and Mainland Southeast Asia *, Annual Review Anthropology* 34. 181–206.

Feuillet, Jacques. 2001. *Introduction à la typologie linguistique*. Paris: Honoré Champion.

Gilbers, Dicky G., J. Nerbonne & J. Shaeken (eds.). 2000. *Languages in contac*t. Amsterdam & Atlanta: GA, Rodopi.

Goddard, Cliff. 2005. *The languages of East and Southeast Asia: An introduction*. Oxford & New York: Oxford University Press.

Grinevald, Colette. 1999. Typologie des systèmes de classification nominale. *Faits de Langue* 14*, La catégorisation dans les langues*. 101–122.

Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. In G. Senft (ed.), *Nominal classification*, 50–92. Cambridge: Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Heine, Bernd & Derek Nurse (eds.). 2008. *A linguistic geography of Africa*. Cambridge: Cambridge University Press.
Henderson, Eugénie J. A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian language. *Lingua* 15. 400–434.
Hickey, Raymond (ed.). 2019. *Handbook of Language Contac*t. Malden, MA: Blackwell Publishing.
Jenny, Mathias & Paul Sidwell. 2015. *The handbook of Austroasiatic languages* (2vols). Leiden: Brill.
Kahrel, Peter & René van den Berg (eds.). 1994. *Typological studies in negation*. Amsterdam: John Benjamins.
Kopecka, Anetta & Bhuvana Narasimhan. 2012. *Events of putting and taking. A crosslinguistic perspective.* Amsterdam: John Benjamins.
Koptjevskaja-Tamm, Maria. 2006. The circle that won't come full: Two potential isoglosses in the Circum-Baltic Area. In Matras Yaron, April McMahon & Nigel Vincent (eds.). *Linguistic areas: convergence in historical and typological perspective*. Hampshire: Palgrave.
Koptjevskaja-Tamm, Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages: an areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *The Circum-Baltic languages: Typology and contact*, vol. 2: Grammar and typology, 615–750. Amsterdam: John Benjamins.
Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.
Masica, Colin P. 1994. Some new perspectives on South Asia as a linguistic area. In A. Davison & F. M. Smith (eds.), *Papers from the 15th South Asian Language Analysis Roundtable 1993*, 187–200. Iowa City: University of Iowa.
Matisoff, James A. 1986. Linguistic diversity and language contact. In John McKinnon & Wanat Bhruksasri (eds.), *Highlanders of Thailand*. Singapore: Oxford University Press. 56–86.
Matisoff, James A. 1991. Areal and universal dimensions of grammatization in Lahu. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization: Focus on Theorical and Methodological Issues*. London: John Benjamins, vol. 2, 383–453.
Matisoff, James A. 2001. Genetic versus contact relationship: Prosodic diffusibility in South-East Asian languages. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 291–327. Oxford: Oxford University Press.
Matras, Yaron & April McMahon & Nigel Vincent. 2006. *Linguistic Areas: convergence in Historical and Typological Perspective*, Hampshire: Palgrave.
Matras, Yaron & Jeanette Sakel (eds.). 2007. *Grammatical borrowing in cross-linguistic perspective.* Berlin & New York: Mouton de Gruyter.
Migliazza, Brian. 1996. Mainland Southeast Asia: A unique linguistic area. *Notes on Linguistics* 75. 17–25.
Muysken, Pieter. 2000. From linguistic areas to areal linguistics: a research proposal. In D. G. Gilbers, J. Nerbonne & J. Shaeken (eds.), *Languages in contact*, 263–275. Amsterdam & Atlanta: GA, Rodopi.
Muysken, Pieter (ed.). 2008. *From linguistic areas to areal linguistics.* Amsterdam: John Benjamins.
Norde, Muriel, Bob de Jonge & Cornelius Hasselblatt (eds.). 2010. *Language contact – New perspectives.* Amsterdam: John Benjamins.

Sandfeld, Kristian. 1930. *Linguistique balkanique: problèmes et résultats.* Paris: Librairie C. Klincksieck.

Sakel, Jeanette & Matras Yaron. 2008. Modelling contact-induced change in grammar. In Stolz Thomas, Bakker Dik & Palomo Roasa Salas (eds.), *Aspects of language contact – New theoretical, methodological and empirical findings with special focus on Romancisation processes*, 63–87. Berlin & New York: Mouton de Gruyter.

Siemund, Peter & Kintana Noemi. 2008. *Language contact and contact languages*. Amsterdam: John Benjamins.

Stolz, Christel & Thomas Stolz. 2001. Mesoamerica as a linguistic area. In M. Haspelmath, E. König, W. Österreicher & W. Reible (eds.), *Language typology and language universals*, vol. 2, 1–77. Berlin & New York: Mouton de Gruyter.

Stolz, Thomas. 2002. No *Sprachbund* beyond this line! On the age-old discussion of how to define a linguistic area. In P. Ramat & T. Stolz (eds.), *Mediterranean languages. Papers from the MEDTYP workshop*, 259–81. Tirrenia, June 2000. Bochum: Brockmeyer.

Stolz, Thomas. 2006. All or nothing. In Yaron Matras, April McMahon & Nigel Vincent (eds.), *Linguistic areas: Convergence in historical and typological perspective*, 32–50. Hampshire: Palgrave.

Thomason, Sarah G. 2000. Linguistic areas and language history. In Dicky Gilbers, John Nerbonne & Jos Shaeken (eds.), *Languages in Contact*, 311–327. Amsterdam & Atlanta: GA, Rodopi.

Thomason, Sarah G. 2001. *Language contact: An introduction*, Edimbourg: Edinburgh University Press.

Thurgood, Graham & Randy J. Lapolla (eds.). 2003. *The Sino-Tibetan Languages*. London: Routledge.

van der Auwera, Johan. 1998. Revisiting the Balkan and Meso-American linguistic areas. *Language Sciences* 20. 259–70.

Vittrant Alice. 2006. Les constructions des verbes en série. Une autre approche du syntagme verbal en birman. *Bulletin de la Société de Linguistique de Paris* 101 (1). 305–367.

Vittrant Alice. 2010. Aire linguistique Asie du Sud-Est continentale: le birman en fait-il partie ? *Moussons* 16. 7–38.

Zúñiga, Fernando & Seppo Kittilä. 2010. Benefactives and Malefactives. Typological perspectives and case studies. Amsterdam: John Benjamins.

David A. Peterson

Bangladesh Khumi*

Introduction

Bangladesh Khumi is a member of the Kuki-Chin branch of Tibeto-Burman spoken by around 2,000 people in the Chittagong Hill Tracts of Bangladesh. There are two mutually intelligible dialects of the language in Bangladesh, and there are closely related Khumi varieties with several tens of thousands of speakers in adjacent parts of Burma. This description is based on the northern dialect of Bangladesh Khumi, centered around Ruma Bazaar in the Bandarban Hill Tracts. The observations here are based primarily on an extensive text corpus with material representing a wide variety of discourse genres. Numbers following examples indicate location in the text corpus; examples without such indications are marked as elicited.

Members of this ethnolinguistic group are highly multilingual, many of them commanding near fluency in Marma, a local variety of Arakanese. They also show variable, but often advanced proficiency in Bawm, Mru, and Tripura. Of these, Marma is far and away the most significant from a linguistic standpoint, with a profound impact on lexicon and syntactic constructions, and probably ultimately on syllable structure and possibly even the tonal system and vowel allophony of the language; other languages the Khumi speak are not as influential.

The Khumi in Bangladesh, according to their oral record, migrated in a number of waves west from a more easterly point starting several generations ago. There is still a moderate degree of fluctuation in their location, with families drifting back and forth from one side of the border to the other as living conditions change.

* My research on Khumi has been sponsored by the Fulbright Foundation, The Max Planck Institute for Evolutionary Anthropology, and National Science Foundation grant BCS-0349021 Many thanks are due to Lelung Khumi and other members of the Khumi community in Bangladesh who have made this research possible. The contents and format of this paper have benefited greatly from comments provided by the editors and an anonymous reviewer.

David A. Peterson: Program in Linguistics, HB 6220, 307 Reed Hall, Dartmouth College, Hanover, NH 03755 U.S.A., E-Mail: david.a.peterson@dartmouth.edu

https://doi.org/10.1515/9783110401981-002

In terms of its position within Kuki-Chin, Khumi is usually regarded as belonging to a southern division of the family (Grierson 1904, So-Hartmann 1988, VanBik 2009). It is unclear that this subgrouping is warranted, however, and the language may instead represent an early branching from Proto-Kuki-Chin, as implied by Bradley's 1997 classification. Peterson 2017 alternatively suggests that Khumi belongs to a grouping, Khomic, or Southwestern, which together with the Southeastern and Northeastern groups forms a Peripheral Kuki-Chin branch. Khumi is nevertheless clearly a Kuki-Chin language, sharing, besides abundant lexical and grammatical cognates, diagnostic features like the change of Proto-Tibeto-Burman **s-* to *t^h-*, the innovative 1S pronoun root **kaj*, and probably even the system of verbal ablaut found elsewhere in the subgroup, although it has a distinct manifestation in Khumi. Otherwise, Khumi is not a prototypical Kuki-Chin language, especially given its relatively impoverished system of participant coding, its distinct nominal case morphology, and in unique paths of development for various verbal categories.

1 Phonology

As a Mainland Southeast Asian language, the phonological system of Khumi has a number of prototypical areal features: tonal contrasts (which involve quantity and phonation cues); a reasonably large number of vowel contrasts, including diphthongs; a usual array of consonantal contrasts; and typical restrictions on syllable structure.

1.1 Suprasegmental phonology: tone and register

Khumi exhibits standard sorts of suprasegmental distinctions for a tonal Southeast Asian language. There are five tones, which on monosyllables in isolation are phonetically realized as in (1), the only known minimal quintuplet:

(1) underlying tones:

1	low falling	*së^1*	'scales'
2	low checked	*së^2*	'kilogram'
3	high checked	*së^3*	'pound rice for a second time'
4	rising (with fall at end)	*së^4*	'begin, rise (of the sun)'
5	high falling	*së^5*	'sound of seeds being scattered'

Tones 4 and 5 are clearly contour tones. Any of the five tones may occur with all syllable types, excluding the half-syllable associated with sesquisyllabic elements.

Two other things about the phonetics of these tones bear mentioning. First, syllables with checked tones are about half as long as those with the other three tones, so duration is an important aspect of tone in this language.[1] Also, the two checked tones are usually accompanied by glottalization towards the end of their syllables, meaning that phonation is also a significant tonal cue, alongside pitch and duration. Khumi might thus be viewed as exhibiting register phenomena to a certain extent, although the primary and most consistent tonal cue is pitch.

There is a further marginal tone, extremely high and long, which might qualify as a sixth lexical tone, but it also appears possible to describe it simply as an intonational phenomenon associated with intensification. See the discussion of demonstratives below (section 3.1.1) for a concrete example.[2]

Going beyond the basic tonal analysis, specific manifestations of the underlying tones and their interaction with each other yields a fairly convoluted system, but one which would appear to be typical for Kuki-Chin languages. (See, for instance, Hyman and VanBik 2004 and Watkins 2013 for some recent treatments of tonal interactions in other languages.) The instantiation of lexical tones in Khumi depends heavily on the morphosyntactic environment of nominal and verbal roots; furthermore, the tonal instantiation itself often is the sole indicator of particular high-frequency grammatical categories, to the exclusion of segmental material which might otherwise be involved in signaling those categories. For a fairly straightforward example, see (38), where the speaker marks 'locative' tonally on the first word rather than using the (also acceptable) segmental locative marker, *=a*^{l}. The second word in that example hypothetically has a tonal instantiation of 'negative', except that this predicate never occurs in the affirmative anyway.

Henderson 1965 discusses similar morphological uses of tone in other Southeast Asian languages (and see also Coupe 2007: 64–5 on the alternative tonal instantiation for Mongsen Ao's agentive case marker), but to my knowledge no other language of the area is reported to exhibit as widespread a use of

1 An alternative analysis might therefore be to posit a vowel length distinction, and reduce the number of underlying tones to three (high falling, rising, low, but this would raise problems for the description of tonal variants, discussed in what follows.

2 Solnit describes a highly comparable phenomenon in Kayah Li (1997: 27–28), suggesting this may have a wide areal spread. Several of Khumi's immediate neighbors also appear to exhibit it.

tone for marking morphosyntactic information as Khumi does. It seems likely that careful analysis of other Kuki-Chin languages will reveal comparable intricacies, however.

The phonetic characteristics of the tone variants for the basic tones 1–5 are summarized first, with phonetic descriptions, in (2). (2) also summarizes the morphosyntactic contexts where these variants appear for given nouns and verbs with particular underlying tones. There may be as yet undetected grammatical contexts which induce tonal variants, but these are the ones which have been identified so far.

(2) contextual variants of underlying tones:

tonal variant:	**phonetic description:**	**morphosyntactic contexts:**
6	checked low, but distinct from tone 2	genitive for nouns with underlying tones 1 and 2; negative for verbs with underlying tones 1 and 2
7	high checked, but higher than tone 3	genitive for nouns with underlying tone 3
8	high checked, but with less glottalization than other checked tones	imperative for verbs with underlying tones 3 and 4
9	rapidly rising (without a fall)	genitive for nouns with underlying tone 4
10	gradually rising (without a fall)	locative for nouns with underlying tones 1 and 4; irrealis for verbs with underlying tones 1 and 4
11	high level	locative for nouns with underlying tones 2, 3, and 5; irrealis for verbs with underlying tones 2 and 3
12	extremely high checked	genitive for nouns with underlying tone 5; negative for verbs with underlying tones 3 and 4

A conceptualization of the system focussing more specifically on underlying tones and their variants in particular morphosyntactic contexts is given in (3).

(3) underlying tones and their variants:

underlying	genitive	locative	V citation	negative	irrealis	imperative
1	6	10	1	6	10	2
2	6	11	3	6	11	2
3	7	11	3	12	11	8
4	9	10	4	12	10	8
5	12	11	n/a	n/a	n/a	n/a

As can be seen more clearly by the representation in (3), there are usually neutralizations for different underlying tones in the specific morphological categories. In fact, were it not for different behaviors of underlying tones 2 and 3 in the negative and imperative forms, the distinction between the high and low checked tones in verbs would be undetectable, as it is neutralized in verb citation forms (both are instantiated as tone 3). The lack of variants of tone 5 for negative, irrealis, and imperative, and the lack of a V citation form for that tone arises from the fact that no tone 5 words are verbs, and so never exhibit these verbal categories.

The tonal system exhibits further complexities when nouns are marked with other case and discourse clitics, or if verbal complexes have one or more post-verbal element besides the negative and irrealis markers. Examples in the paper indicate essentially surface-level tones, which frequently deviate from underlying ones (and also from the non-lexical tones given in the chart above) due to the operation of a few tone rules which I will not attempt to illustrate in detail here.

To take one relatively simple example, though, in sentence (10) of the accompanying text, the nominalizer *=nö*^{3} surfaces with tone 2 rather than tone 3 in the sequence *niw*^{1}*-thay*^{11}*-nö*^{2} because it is preceded by the 'potential' element *-thay*^{3}, and a sequence of 3-3 undergoes a change to 3-2. A further complication in this form is the realization of tone 3 on *-thay*^{3} as tone 11 due to its occurrence in non-final position (see the immediately following discussion).

The examples in this paper do not represent a surface-level process of peak-delay, whereby syllables with tones 1 (low falling) and 2 (low checked) preceded by a rising tone syllable (tone 10, which may result from underlying tone 4) are realized as tones 5 (high falling) and 3 (high checked), respectively. This peak delay is detectable instrumentally; speakers appear to hear such sequences in terms of their underlying tones, however. In non-final contexts, tone 1 surfaces as a relatively low unchecked tone without the normal fall that tone 1 has in final contexts. Tone 11 is similarly used for high-level tones in non-final contexts, which may be due to underlying tone 3 or tone 5; slight differences have been noted for these in some cases, which may suggest a further distinction needs to be made, ultimately, but I keep things as simple as possible here.

1.2 Segmental phonemes: consonants and vowels

Khumi's consonant inventory, seen in (4), is typical for Southeast Asia, although it differs from that of languages in the immediate vicinity in certain respects.[3] It has standard sorts of voicing and aspiration distinctions in its system of stops, but unlike many area languages (in particular, Marma), there is no voiced/voiceless distinction for sonorants. *k* and *k*^{h} have a fairly uvular realization in contexts preceding low back vowels. *ʃ* is in free variation with *s*, a pattern also seen in a number of neighboring languages. *dʒ* and *j* are also in relatively free variation, except where *j* forms part of a consonant cluster.

(4) consonant inventory:

	bilabial	**labio-dental**	**alveolar**	**palato-alveolar**	**velar**	**glottal**
stops						
unaspirated	p		t		k	ʔ <'>
aspirated	p^{h} <ph>		t^{h} <th>		k^{h} <kh>	
voiced	b		d			
nasals	m		n		ŋ <ng>	
fricatives	(f)	v		s~ʃ <s>		h
affricates			tɬ <tl>	tʃ <c>		
			t^{h}ɬ <thl>	dʒ <j>		
approximants			r, l	j <y>		

In addition to the above consonants, original Proto-Kuki-Chin **u* has led to some interesting developments. In many instances this vowel has diphthongized to *iw* (e.g., PKC **buʔ* > Khumi *biw*^{2}). In most cases where the vowel has remained *u*, however, it has led to labialization of the immediately preceding consonant, and this labialization has further evolved into frication approximating *v* or *f*, depending on the voicing and other laryngeal characteristics of the associated consonant; this frication carries over into the vowel, yielding something quite distinct from a simple affricate. Even with only about two dozen instances *without* this frication, there are a handful of minimal pairs which show the distinction to have been phonemicized (e.g., *hu*^{5} 'there' vs. *hfu*^{5}

3 Orthographic conventions used here for representing sounds where the relationship may not be clear are given in angled brackets following the sounds. Otherwise, these charts represent distinct phonemes with the symbol corresponding to their main phonetic variant.

'bamboo'), but exactly what phonemes should be recognized is problematic. Since this appears to be more a feature of the vowel at this point, one analysis would be to posit a separate vowel, resulting in two *u*-like vowels. A disadvantage of this is that there is a remarkably small number of words having a straightforward *u* phoneme, without the accompanying frication, but an overwhelming number exhibiting the variant with it. On the other hand, treating it as a feature of the consonant would nearly double the consonant inventory, with a skewed distribution resulting from the source of frication. Perhaps the best solution would be to treat these cases as *Cv* or *Cf* consonant clusters, reflecting their origin in labialization/affrication. However, in designing a practical orthography for Khumi, it is clearly more advantageous to treat this frication as automatic, and somehow signal its absence in the relatively small number of exceptions; for purposes of this paper, I will simply ignore the difference, as there are only a few instances where the frication is absent.[4]

There are also several interesting things to note about nasals in Khumi. First, one salient consonant sound not represented in the inventory above is a doubly-articulated bilabial-velar nasal. This occasionally occurs as a variant of a syllable final velar nasal reflecting an etymological bilabial nasal, e.g., the word 'house', *üng*^{1}, where *ng* reflects **-m*, sometimes occurs with this variant bilabial-velar nasal double articulation. Next, for unclear reasons, there are certain roots which optionally have nasals in one or more of their codas (cf. *biw*^{1}*tlëng*^{5}*~biw*^{1}*tlë*^{5} 'rice package', with an optional final nasal, and *ay*^{1}*tlëng*^{1} 'men's carry-basket', where the final nasal is obligatory.) Finally, Khumi has a number of forms which show an alternation between a final syllable *mu* or a syllabic bilabial nasal, e.g., the variants *mu*^{3}*~m̩*^{3} 'blow'.

The Khumi vowel inventory is given in (5).

(5) vowel inventory:

	front	**central**	**back**
high	i		u
high-mid	e̱ <ë>	ɘ <ö>	ɤ<ü>,o
low-mid	e		ɔ <å>
low	æ <ä>		a

diphthongs: iw, ew, ɘj <öy>, uj <uy>, aj <ay>, ɔj <åy>

Noteworthy in this chart is the richness of vowel and diphthong contrasts, including two non-front unrounded vowel qualities (approximating ɘ and ɤ). The

4 The word *Khumi*, in fact, also exhibits this frication: *khfu*^{11}*mi*^{1}.

only remarkable vowel allophony is found before syllable-final nasals; in this context non-low vowels (and diphthongs) tend to exhibit more lowered and lax phonetic realizations (e.g., *e* tends to be realized as *ɛ* or even *æ* and *u* tends to be realized as *ʊ*). Coda nasals do impart a certain amount of nasalization to preceding vowels, although the degree of nasalization varies according to speaker and word, and since there are clearly audible nasals for most speakers, it does not appear that the proper analysis of nasalization is as a property of vowels rather than syllable-final nasals.

1.3 Syllable structure

The Khumi syllable canon can be characterized as in (6):

(6) (C(v))C(C)V(N)+tone

Syllables have an underlying tone. The only possible final consonant (with the exceptions discussed earlier) is *ng*. The first (optional) *C(v)* portion of the schema in (6) represents the half-syllable found in sesquisyllabic forms. Orthographically, these half-syllables are indicated by an apostrophe between the half-syllable consonant and the following full syllable; no vowel is written in the half syllable: *p'liwng*^{4} [pə^{1}liwŋ^{4}] 'heart'. Rarely, these half syllables combine with a main syllable that has an initial consonant cluster. Initial consonant clusters themselves are also relatively rare and tightly constrained in terms of their composition: the first element must be a stop and the second must be *l*, *r*, or *j* (although not all combinations are attested in the available data.) By contrast, the consonants of the minor and major syllables of a sesquisyllable are largely unconstrained, though there is a clear tendency for them to have distinct places of articulation.

2 Morphology

The domain of morphology in Khumi and many other area languages is vexed by lack of coincidence between grammatical and phonological words (Dixon and Aikhenvald 2002, Hall et al. 2008). While the vast majority of morphemes in Khumi are monosyllabic, sesquisyllabic, or disyllabic, and could serve as independent phonological words, assuming a phrase-level analysis of tonal sandhi phenomena, many of them nevertheless are morphosyntactically bound, and so take on a highly structured, position class-like status; this is especially the case

for post-verbal elements, as will be discussed further with the exposition of specific verbal categories below. While it would be possible to describe these as particles, it seems preferable, insofar as possible, to analyze them as morphosyntactic affixes or clitics, which may or may not constitute independent words from a phonological perspective, resorting to a separate category of particle only when absolutely required.

2.1 Word structure

With this background in mind, Khumi exhibits classic Mainland Southeast Asian monosyllabicity in the sense that it tends strongly to have monosyllabic or sesquisyllabic formatives (i.e., markers of inflectional information) which could be analyzed as standing alone as separate phonological words. It should be kept in mind, however, that a monosyllabic analysis belies the fact that the elements in question usually occur in highly fixed positions with respect to each other and from a morphosyntactic standpoint are bound: they may never occur in isolation, as free elements may, given an appropriate context.

Adopting Bickel and Nichols' (2007) approach to traditional morphological typology, many of the formatives Khumi has are isolating and agglutinative, and they typically have separative exponence: they tend not to encode multiple inflectional meanings simultaneously. Morphosyntactic word structure is highly complex, and given the semantic richness of grammatically bound adverbial elements to be discussed shortly, actually verges on polysynthetic.[5] The major deviation from these characterizations is the tonal variants of roots and other formatives used instead of discrete segmental expression for specific inflectional catetories discussed above (irrealis, locative, negative, etc.), which leaves the language with a decidedly less isolating and more fusional appearance.

Compounding is fairly standard for this part of the world. In Khumi, two-member compounds are typically right-headed (if endocentric) and are phonologically distinguished by the treatment of their first (and sometimes second) element. There are various possibilities, but typically, the first element bears less stress than the second and does not show the full array of tonal distinctions found on its last syllable compared to non-compound occurrences of that member; tonal distinctions on the last syllable of the first member are generally reduced to high vs. low, although tone 4 (rising) in compounds sometimes shows

5 Cf. Coupe 2007's category of 'lexical' suffixes in Mongsen Ao (298 ff.).

a peak delay that bleeds over into a following syllable, altering the tone of that syllable: e.g., *kno*^{4}+*kho*^{4} 'ear+hole' is realized as *kno*(rising with peak delay)-*kho*^{5}.

Besides exocentric and endocentric compounds, there are coordinate compounds (e.g., *am*^{1}-*am*^{1}*po*^{1} 'mother-father=parents', *jay*^{11}-*bä*^{5} 'elder sibling-younger sibling=siblings'). More specialized compound types like psycho-collocations and elaborate expressions will be taken up in the next two sections.

2.2 Psycho-collocations

A handful of psycho-collocations have been recorded for Khumi, although the phenomenon does not appear to be as prevalent as it is in other area languages. There may be monomorphemic roots which come close to expressing some of these concepts, but they are probably not entirely equivalent. Most of the recorded psycho-collocations involve the word *p'liwng*^{4} 'heart', as seen in (7), although there are a few for some other body parts (e.g., *möy*^{3} 'eye' and *k'no*^{4} 'ear').

(7) heart:

p'liwng^{4} *bi*^{4}	lit. 'heart hot'='unhappy, sad, upset'
p'liwng^{4} *håy*^{3}	lit. 'heart good'='generous'
p'liwng^{4} *a*^{1}*dü*^{1}	lit. 'heart equal'='harmonious'
p'liwng^{4} *a*^{1}*ngay*^{3}	lit. 'heart MID-like'='like, be pleased with'
p'liwng^{4} *a*^{1}*pha*^{1}	lit. 'heart sufficient'='happy, contented, satisfied'
p'liwng^{4} *khay*^{2}	lit. 'heart keep'='hold a grudge'
p'liwng^{4} *dång*^{4}	lit. 'heart think.about'='pay attention, think'
p'liwng^{4} *a*^{1}*ma*^{1}	'sad, upset' (second element perhaps originally 'lost')
p'liwng^{4} *a*^{1}*phöng*^{4}	lit. 'heart suspect'='worry'
p'liwng^{4} *phra*^{2}	lit. 'heart destroyed'='worry, upset'
p'liwng^{4} *khå*^{1}	lit. 'heart strong'='feel sure, certain, secure'

(sometimes an elaborate expression for heart, *p'liwng*^{4}-*p'thüng*^{3} (='heart-liver') is used in these constructions rather than just 'heart')

(8) eye:

möy^{3} *a*^{1}*dåy*^{2}	lit. 'eye MID.lure'='attracted to, interested in'

(9) ear:

k'no^{4} *ha*^{2}	lit. 'ear porous'='able to hear easily'
k'no^{4} *döy*^{6}-*lä*^{3}	lit. 'ear die-NEG'='lie awake, unable to fall asleep'

2.3 Elaborate expressions

Khumi makes extensive use of elaborate expressions (described in detail by Peterson 2010). These are quasi-reduplicative compound structures which are used in place of simpler structures for stylistic effect. The simpler structure is included in the more elaborate one. Use of elaborate expressions conveys the impression of speaking eloquence and is generally regarded as a desirable discourse feature. Given the performative nature of narrative, elaborate expressions are more likely to be encountered in storytelling than in procedural discourse or conversation, although they do occur with lower frequency in other text genres. Some typical examples are listed in (10).

(10) elaborate expressions
  a. *tu^{1}pli^{1}-tu^{1}pla^{1}* 'reduplicate elaboration-box'='box'
  b. *s'rung^{1}-s'rå^{1}* 'nonce elaboration-tobacco'='tobacco'
  c. *tuy^{1}-may^{1}* 'water-fire'='water'
  d. *mi^{1}-may^{1}* 'reduplicate elaboration-fire'='fire'
  e. *ka^{11}si^{1}-tå^{11}kë^{1}* 'star-nonce elaboration'='star'

Elaborate expressions in Khumi may differ from those found elsewhere in Southeast Asia (see Haas 1964 on Thai or Matisoff 1973 on Lahu for classic characterizations) in terms of their structure. While some elaborate expressions conform to a classic A-B-A-C type structure, as in (10a–b), others deviate from this pattern. (10c–e), for instance, either involve only two syllables, or involve four syllables which do not include a shared syllable.

2.4 Reduplication

The elaborate expressions just considered involve a reduplicative structure in the default case (i.e., when no more specific elaboration strategy overrides the default.) Besides this, there are certain verbal categories (adverbial elements and verbal classifiers, discussed in 4.2.2) which can occur in reduplicated form, with no readily discernible semantic effect. Next, the ideophone category (see Peterson 2013a) has numerous fully reduplicated elements (e.g., *kri^{1}kra^{1}-kri^{1}kra^{1}* 'sound of walking on a bamboo deck', *li^{1}lång^{5}-li^{1}lång^{5}* 'motion of bamboos waving in the breeze'.) Finally, there does appear to be limited use of reduplication for certain reciprocal (*lewng^{11}-lewng^{11}-rë^{1}*, lit. 'person-person-COUNT') or distributive contexts (as in (11), where *cung^{11}ngay^{1}* '100' is repeated, if not reduplicated.)

(11) *k'ni[11]bi[11]* *k'ni[11]nay[3]* *thiwng[10]=a[1]* *t'ko[11]=nö[2]* *alö[1]-ngay[11]=te[1]*
sunshine.LOC rain inside=LOC go.IRR=NMLZ need-want=EVID

n'='ë[10] *si[1]'ewng[11]=mö[2]* *thuy[11]-pë[1]* *tang[1]ka[2]* *cung[11]ngay[1]*
QUOT=REAS (name)=FGD say-BEN money 100

cung[11]ngay[1] *khay[11]-pë[1]-yo[3]*
100 leave-BEN-IMPV
'They said they needed them [umbrellas] for walking in the sunshine [speaker mistake], the rain, Si'ewng said, so I left them 100 taka each.' (41.128)

Otherwise, reduplication is not used for any established function in the language.

More important than reduplication proper is basic parallelism, which seems to be a driving force in Khumi as in many other Southeast Asian languages (see, for instance, Solnit 1995.) (12) below shows one such case for Khumi, where the speaker starts his sentence with two words meaning 'tomorrow'.

(12) *vå[1]khång[11]=lö[1]* *kh'dang[10]=lö[1]* *ay[11]ni[2]* *måy[1]-co[1]*
tomorrow=TOP tomorrow=TOP 1D.INCL wild.boar-child

lew[11]=kh'=bo[3] *n'=pë[1]=te[5]* *am[1]nå[10]=a[1]*
catch=HORT=REAL QUOT=say=EVID younger.sibling=LOC
' "Tomorrow, tomorrow let's catch a wild boar piglet," he said, to his younger brother.' (34.86)

This does not have the structure of an elaborate expression, which usually would involve compounding of the two elements in question. Instead, the first word is from the northern dialect and the second is from the more southerly Thanchi Khumi variety spoken in Bangladesh, which speakers of the northern Ruma Khumi dialect also are familiar with. The second instance of 'tomorrow' really adds nothing to the content of the sentence, but by virtue of its parallelism, the repetition is stylistically more satisfactory to speakers.

3 Grammar and Syntax

In gross typological terms, Khumi is consistently verb-final with a highly complex verbal structure.[6] Rigid though it is in this ordering, as often is the case in verb-final languages, nominal expressions may occur fronted in a topicalization or left dislocation position, or post-posed as an afterthought. In addition to limited prefixal morphology, which may code S/A (=subject) participants, the verb root may be surrounded by an array of valence-affecting formatives (see Peterson 2013b), some of which clearly arose from historically independent verbs, a process reminiscent of verb serialization. Nevertheless, there is really nothing resembling the productive serialization type often found in Mainland Southeast Asian languages (see Aikhenvald 2005). The language also exhibits fairly limited numeral classifier phenomena. It has frequent argument ellipsis under appropriate discourse recoverability. In terms of alignment, Khumi displays roughly nominative-accusative and primative/secundative (Haspelmath 2005) case marking, regulated by considerations of P/R (=object) specificity. Relativization makes use of nominalization without relative pronouns. These and related topics will be taken up in more detail in this section.

3.1 Nominal domain

In terms of the types of things and order of elements within the noun phrase, Khumi is a typical Tibeto-Burman language for the most part. For nominals, there is collective marking rather than a strict plural, although the same marking indicates plural in the independent pronouns. Numeral classifiers occur but are limited in their coverage and distribution, and the language has a small set of clitic case-marking post-positions and discourse status indicators. In addition to the segmental case markers, however, a noteworthy tendency already mentioned above is that the two highest frequency case markers, genitive and locative, often are expressed by the tone of the last syllable of the case-marked NP (excluding discourse status markers, which are more peripheral).

6 In this section and in subsequent discussion, I make use of the by now familiar semantico-syntactic designations S (single participant of an intransitive clause), A (more agent-like participant of a transitive clause), and P (less-agent-like participant of a transitive clause). Less familiar is the designation R, used for the less-patient or more-recipient-like participant of a ditransitive clause (see Haspelmath 2005).

3.1.1 Basic structure of the NP

The structure of Khumi NPs is fairly complex, so I break this discussion into selected portions of it. A maximally schematic representation for the structure of Khumi NPs is given in (13); the subparts other than the head (premodifiers, post-modifiers, and clitics) will be discussed in more detail in what follows.

(13) overall NP structure:
[premodifier(s)] head N [post-modifier(s)]=[clitics]

Of course, a noun phrase containing all or even most of these elements at once is relatively rare; (14) gives a few examples of some of the more complex NPs attested in our large text corpus.

(14) a. *cång[1]thing[1]kiwng[5]* *n'=pë[1]=nö[3]* *ha[1]-rë[1]=lö[1]*
(name) QUOT=say=NMLZ one-COUNT=TOP
'(as for) the one called Cawngthingkiwng' (2.3)

b. *c'po[1]* *lewng[5]* *t'riw[1]-rë[10]=ya[1]*
son CLF human six-count=LOC
'six sons [marked with a locative case clitic]' (3.28)

c. *a[1]ke[2]* *kang[1]-tha[5]* *ha[1]-rë[1]*
knife NMLZ-new one-COUNT
'a new knife' (3.46)

d. *a[1]vang[6]* *üng[1]* *dewng[1]* *phay[1]-rë[1]*
village.GEN house CLF HOUSEHOLD thirty-COUNT
'thirty households of the village' (35.64)

The internal sequencing of the premodifier portion of the NP is as in (15), though there appears to be some flexibility in the positioning of the demonstrative.[7]

(15) premodifer(s):
[demonstrative possessor relative clause]

7 For instance, direct elicitation suggests that a particularly heavy relative clause can be preferentially dislocated, yielding a sequence relative clause-demonstrative-head N. So far no text data sheds light on this issue, however.

There are three basic demonstratives: *h'ni*[3] 'speaker proximal/addressee distal', *tu*[1]*ni*[3] 'speaker distal/addressee proximal', *hu*[1]*ni*[3] 'speaker/addressee distal', containing the deictic elements *hi*[2], *tu*[2], and *hu*[2]. To these could be added the super-distal *huuni*[3], pronounced with a longer vowel and the tone that was mentioned earlier (section 1.1) as a possible sixth lexical tone; as noted before, though, this tone always involves a nuance of intensification, as in super-distal vs. merely distal, and perhaps might simply be regarded as an intonational contour applied in cases where such intensification is desired rather than as involving a distinct lexical tone.

Adjectival concepts are expressed in one of two ways in Khumi. First, there are verb-like roots which occur nominalized by the generalized nominalizer *=nö*[3] in prenominal position, constituting what is essentially a relative clause. (See section 3.3 below for examples of prototypical relativization.) Such relativizations technically also may occur in post-nominal position, although this is infrequent in natural discourse. This possibility is reflected in the (relative clause) portion of the post-modifier expansion seen in (16).

(16) post-modifer(s):
[adjective/(relative clause) classifier-numeral-count/quantifier(-count)]

Secondly, represented as 'adjective' in (16), there are a large number of verbal or adjectival concepts which have corresponding lexicalized nominalizations formed with a prefix *ka*(*ng*)-; if such a nominalization exists, it may serve as an adjectival modifier *following* the noun. (See (14c) above for an example.) The remaining post-modifer elements are classifiers and associated elements, or quantifiers, which will be discussed further in the next section.

Besides the collective marker, *=cë*[5], NP clitics include case markers and various indicators of discourse status:

(17) clitics:
[collective=locative/genitive/other case markers=
foregrounder=topic/focus]

The high frequency case markers of Khumi are given in (18), along with some examples where they may be seen in context, where possible:

(18)	**case marker:**	**meaning:**	**example:**
	='ë[1], or tonal instantiation	genitive	(36b), (45)
	=a[1], or tonal instantiation	locative	(12), (23)
	=håy[1]	comitative/instrumental	text, sentence 1

=ma[1]	ablative	(19b)
=ma[4]	more restricted locative, often with allative sense	text, sentence 19

Members of this category with extremely low text-frequency include *=cang*[3] 'standard of comparison' and *=rö*[2] 'limitative', meaning 'up to N'. Other spatial notions are primarily encoded using relational nouns loosely compounded with a noun and marked by a locative or ablative case marker, as in (19a-b).

(19) a. *N* *teng*[10]*=a*[1]
N side=LOC
'to/at/etc. N'

b. *N* *thiwng*[4]*=ma*[1]
N inside=ABL
'from inside N'

There are two primary members of the discourse status marker category, which will be introduced briefly here but are discussed more fully in a section 3.3.2. *=lö*[1] is a high-frequency topic marker. *=mö*[3], on the other hand, correlates highly with focus. However, since a different element, *=pö*[1] 'also', has clearer focussing functions, and *=mö*[3] is more varied in its use, we treat the latter as a marker of foregrounding. Here, I assume notions of topic and focus along the lines of those assumed in Lambrecht 1994 and Van Valin and LaPolla 1997. The notion of foregrounding I assume, as distinct from focus, is similar to that developed in Chelliah 2009.

Compared to other Tibeto-Burman languages, Khumi is highly typical in terms of the word order attested at the basic level of the noun phrase. Its main clause word order being verb final, it has postposed case markers and postpositional relational noun phenomena, and since its marker for standard of comparison is a case marker, it also has standard-marker ordering, as usual for a verb-final language. It shows variability in placement of relative clauses and adjective-like expressions, but has genitives preceding the head noun internal to NPs. The only respect in which its NP-internal word order is somewhat atypical is with respect to numeral classifiers (see section 3.1.2 below).

Not surprisingly, noun phrases may also contain an independent pronoun in place of a noun. The paradigm of independent pronouns is given in (20).

(20) Independent pronouns:

	1st person inclusive/exclusive	**2nd person**	**3rd person**
singular	*kay*^{1}	*nang*^{1}	*ni*^{3}
dual	*ay*^{11}-*ni*^{2}/*kay*^{1}-*ni*^{3}	*nang*^{1}-*ni*^{3}	*ni*^{11}-*ni*^{2}
plural	*a*^{1}-*cë*^{5}/*kay*^{1}-*cë*^{5}	*nang*^{1}-*cë*^{4}	*ni*^{11}-*cë*^{5}

Noun phrase-internally, pronouns occur in the same position as the head noun, but there are generally restrictions on the remaining elements that may appear. Pronouns only co-occur with case markers and discourse-status markers. Demonstratives (illustrated above) are built from one of the deictic elements mentioned earlier and the third person pronouns; often demonstratives serve in place of a third person pronoun.

3.1.2 Classifier (CLF) device

Given general Kuki-Chin and also Mainland Southeast Asian languages' expected propensity to make use of numeral classifiers, Khumi has a surprisingly small inventory of true classifiers. The ones that exist are typical in terms of their semantics, such as *lewng*^{5} 'humans', *jiwng*^{5} 'long, thin things', *dång*^{3} 'sheets of paper', *dewng*^{1} 'houses', *kiwng*^{1} 'trees', etc., but many typical classifier categories appear to be missing (such as a dedicated classifer for animals.) Many of the formal classifier elements are simply measure words (e.g., *së*^{2} 'kilogram', *pho*^{1} 'basketful'), and so do not clearly contribute to the total, and a number of them are simply noun copies (e.g., *på*^{3} 'flowers', *a*^{1}*thay*^{3} 'fruits'). In discourse, it is quite normal for numeral classifiers simply to be left out.

There are also some fundamental syntactic differences between the classifer constructions seen in Khumi compared to those found elsewhere in Mainland Southeast Asia (cf. the possible ordering of elements identified by Barz and Diller 1985). The basic template is given in (21).

(21) N classifier numeral-COUNT

This template may be illustrated by (22a). The element glossed here as 'COUNT' is a central element in the construction; it appears not only in classifer construction, but also in a number of adverbial contexts involving quantification (see below for further occurrences). In fact the element is not identical to the verb 'count', which is *rë*^{2}, but it seems probable that they are related.

(22) a. *c'rang*[4] *jiwng*[5] *pang*[11]*-rë*[1]
ceremonial.knife.type CLF LONG five-COUNT
'five *c'rangs*' (elicited)

b. *c'rang*[4] *pang*[11]*-rë*[1]
ceremonial.knife.type five-COUNT
'five *c'rangs*' (elicited)

c. *c'rang*[4] *ha*[1]*-rë*[1]
ceremonial.knife.type one-COUNT
'one *c'rang*' (elicited)

d. *c'rang*[4] *jiwng*[11]*-rë*[1]
ceremonial.knife.type CLF LONG-COUNT
'one *c'rang*' (elicited)

e. **c'rang*[4] *jiwng*[5] *ha*[1]*-rë*[1]

The classifier element can generally be left out, as in (22b), such that it is really the element *rë*[1], which supports the numeral rather than the classifier. With the numeral *ha*[2] 'one' (and numbers including it), there is a further structural simplification in that either the numeral is present, with no classifier (22c), or the classifier is combined directly with *-rë*[1], without a numeral (22d), but not both: sequences like (22e) do not occur.

Classifiers are used only with numerals; they are not usable anaphorically and they are also not generally used to form indefinites, except for *lewng*[5] 'person', which may sometimes have an interpretation of 'somebody'. The *rë*[1] 'count' element, however, is used in a limited way to form indefinites (*h'tang*[1]*-rë*[1] 'to such an extent', *may*[11]*dë*[1]*-rë*[1] 'a little bit') and occurs with some types of quantification, e.g., *lo*[2] (month) *tlöyng*[11]*-rë*[1] 'every month', or *vä*[1]*-rë*[1] 'every', although it is not obligatory in all such contexts.

We may further note that while the numeral classifier appears to be a relatively poorly instantiated category in Khumi, and certainly one which is often ignored or simplified in discourse, there is a separate and prominent classification phenomenon in the language in the form of high-frequency *verbal* classifiers. These primarily concern the relative size of participants involved in an event. See Peterson 2008 for a full description; a brief example is given in (23).

(23) *...a[1]cë[5]* *h'ni[11]* *ë[1]-bå[1]=khö[11]lö[1]* *pyå[10]=mi[1]*
1P.INCL PROX.DEM.LOC push-AUGVCL=COND ABLE=DUBIT
'...if we push this (a large stone), perhaps we'll be able.' (28:28)

The post-verbal element *-bå[3]* in (23), an augmentative verbal classifier, indicates that the stone (here not overtly expressed) is large in size, in addition to other complex semantic nuances that verbal classifiers often contribute. If the speaker were instead to use the corresponding diminutive verbal classifier, *-bö[3]*, they would be calling attention to the small size of the stone.

3.2 Verbal domain

3.2.1 No inflection

Unlike many other languages of Mainland Southeast Asia, there is a reasonable amount of verbal inflection in Khumi. Least controversially, the language exhibits participant coding (described in Peterson 2002) in the form of prefixes which cannot be analysed as unbound pronominals from either a phonological or a morphosyntactic standpoint.

The status of other marking as inflectional or not depends on whether other bound elements are counted as inflection. As discussed earlier, while an analysis of many morphosyntactically bound elements as independent phonological words is possible for this language, and so many potentially inflectional elements would not be phonologically bound and hence non-inflectional, this ignores the otherwise bound morphosyntactic behavior of the elements in question.

In this section, while recognizing that certain post-verbal elements have somewhat more mobility than other elements, I outline a model for the Khumi verbal complex that recognizes several zones and rigidly ordered positions within those zones. The approach sketched out here is inspired in part by Kari's 1989 analysis of Athabaskan prefixal morphology, as considered for Mongsen Ao by Coupe 2007, and Muysken's 1986 "mode" analysis of Quechua suffixal morphology.

In terms of semantics, some of the elements involved are highly lexicalized or restricted in conjunction with particular predicates, and so not strictly speaking inflectional. They nevertheless are interspersed with more clearly inflectional elements and must be treated in a unified manner with them.

Besides this intertwined nature of more derivational and clearly inflectional formatives, there are discontinuous dependencies between elements, suggest-

ing that the morphological connections between them are tighter than simple concatenation. In particular, the negative marker, which always induces glottalization of an immediately *preceding* syllable, can have a *non-local* effect on the preceding verb root. This is most evident with certain roots having tone 1 in their citation form and the vowel *-a*, like the verbs *ca*^{1} 'eat' and *la*^{1} 'take'. When the negative is present, separate stem alternants, *co*^{6} and *lo*^{6}, or variants with non-final unchecked tones (see, for instance, example (29) below), are required; it does not matter whether the negative is immediately adjacent to or several formatives distant from the root.

3.2.2 Verbal categories

Minimally, a verb consists of a root, although more morphology often occurs in the verbal complex. Indicative verbs in either subordinate or main clauses can exhibit relatively complex structures. Other than indicative morphology, the verbal complex may contain elements marking categories like imperative, hortative, and optative. Interrogative sentences differ in form depending on whether they are polar or non-polar. Generally, interrogatives may contain indicative morphology embedded in them more readily than other non-indicative sentence types may.

The overall structure of the verbal complex requires the recognition of at least the zones seen in (24), which are described in more detail in A–E below. Following an introduction to the structure of these elements, I will make some observations about the types of elements which occur.

(24) prefixes-ROOT-invariably ordered suffixes-
transitional-variably ordered=clitics

A. The *prefix zone* includes participant coding, or middle – i.e., multi-purpose detransitivizer – (*a-*) elements, as well as a causative (*p-*/*t-*), treated in Peterson 2013b. These elements are rigidly ordered and the causative and middle are lexicalized in some cases.

(25) prefix zone:
PARTICIPANT CODING- CAUSATIVE-ROOT
MIDDLE-

The following generalizations apply to participant marking, when it is used. If the S/A is first person, *ka(ng)-* is prefixed to the verb. Otherwise, *ang-* is used. It

is clear that this participant marking ignores a number of the distinctions that are important for the independent pronouns seen earlier: number and inclusive/exclusive. This preverbal participant coding is also often omitted. It occurs fairly frequently in conversation and in reported speech, where first and second person participants are common. However, it only rarely appears in third-person narrative, where it has a subtle perspective-changing effect discussed in Peterson 2002. Participant coding is mutually-exclusive with the middle marker.

B. The *invariably ordered suffix zone* contains a group of ten or eleven bound, rigidly ordered positions situated closest to the verb root. Altogether these involve around 150 highly abstract elements, for the most part falling into several relatively specific semantic types.

(26) invariably ordered suffix zone:

ROOT -BEN- ASSOCIATIVE-DIR-VCL-PRIOR-ADV1-ADV2-ADV3-ADV4-IMPFV -CAUS-

- causative/applicative (CAUS *-hay*^{3})/benefactive/malefactive (BEN *-pë*^{1})/associative (*-håy*^{1}) are all fairly standard in terms of their behaviour; possibly also *-vë*^{4} 'substitutive applicative' (see Peterson 2013b.)
- directional (DIR): *-k'lå*^{4} 'upwards' and *-k'thiw*^{3} 'downwards'.
- (verbal) classifier (VCL): at least (perhaps significantly more than) 100 paired elements carrying all manner of idiosyncratic nuances, but with a central notion of largeness/smallness of a participant, usually S/P (Peterson 2008).
- prioritive (PRIOR): *-ma*^{4} 'first'.
- adverbial: ADV1: usually associated with tightness or completeness, including *-thlöyng*^{1} 'tightly, fast, stuck', *-ceng*^{1} 'tightly, full', *-då*^{4} 'incompletely, a bit, for a while', *-jö*^{11}*'ay*^{2} 'randomly, haphazardly'. One might be tempted to analyze this and Adv2–4 as aspect; certainly aspect is involved, but often much subtler adverbial nuances are invoked, defying analysis in terms of simple aspectual distinctions.
- ADV2: *-prå*^{1} 'immediately, suddenly; desperately, with immediate need', *-båy*^{3} 'all', *-du*^{4} 'in great number', *-lew*^{2} 'at a distance', *-mä*^{3} 'neatly, carefully', *-u*^{1} 'plural participant', *-tüng*^{3} 'carefully, clearly'.
- ADV3: *-khö*^{3} 'in the way', *-may*^{3} 'a bit, partly', *-m'lä*^{4} 'too much', *-sä*^{1} 'repeatedly, clearly'.
- ADV4: *-kä*^{3} 'more, comparative', *-seng*^{3} 'frustrative (due to a mistake)', *-hu*^{3} 'frustrative (due to unexpected absence of something)', *-pay*^{3} 'too much, excessively', *-play*^{3} 'at all, ever, experiential perfect', *-vay*^{3} 'experiential perfect', *-rüng*^{3} 'in advance, already'.

- imperfective (IMPFV): highly ramified category whose members all have vaguely imperfective aspectual nuances; most can occur in a reduplicated form (*-ü*[4], *-të*[4], *-reng*[1], *-thüng*[3], *-mu*[4], *-hë*[1], *-lü*[4], *-väng*[3], *-r'sö*[3], *-rü*[3]) (Note that a reduplicated form is not diagnostic for this class, as some elements from other classes – e.g., *-båy*[3] 'all', *-du*[4] 'in great number', and, most notably, verbal classifiers – may also show a reduplicated form.)

Direct elicitation suggests that for some of these groups, multiple members of a category may co-occur. If this is borne out by naturalistic data, further 'subzones' or additional positions may need to be recognized; e.g., the sequence benefactive-causative has been elicited directly, but does not occur in the text corpus; the opposite ordering may not be elicited and is likewise unattested.

C. The *transitional zone (free particles)* is a group of some of the elements just listed under the imperfective class which show a relatively less bound status than other such elements, especially in that they, or clearly related elements, may occur in preverbal position, in addition to their common post-verbal uses. The elements in question include *të*[4], *väng*[3], and possibly some others. For instance, consider (27a–b), where the postverbal element *-të*[4], usually with iterative semantics, may be preposed to the verbal complex in circumstances where the speaker wishes to place some kind of emphasis on the iterativity, or to attach some kind of special attitude (e.g., annoyance) to the fact of the iterativity.

(27) a. *thuy*[11]*-t'të*[1]
say-ITER
'keep saying' (elicited)

b. *t'të*[4] *thuy*[3]
say
'keep saying' (elicited)

However, elements of the transitional zone are mutually incompatible with elements in the imperfective class, such that it would appear the imperfective class has arisen via the gradual accretion of suffixes at the end of the invariably ordered suffix zone. These elements therefore have a dual nature: they block the occurrence of other imperfective elements, but are themselves inherently more mobile than the other elements of the invariably ordered suffix zone.

D. The *variably ordered zone (bound particles)* includes a number of elements which are all apparently exclusively post-verbal, but which exhibit variable orders with each other, possibly with scopal differences. The basic elements are shown in (29), and an instance of scopal ordering differences is given in (28a–b), where the relative positioning of the inceptive aspect marker and the negative marker has clear consequences for the meanings of the forms.

(28) a. *co[1]-rö[12]-lä[2]*
eat-INCEPT-NEG
'he didn't start eating' (elicited)

b. *co[6]-lä[11]-rö[2]*
eat-NEG-INCEPT
'he's started to not eat' (elicited)

- INTENSIFIER: *-ka[11]mo[1]*.
- ASPECT: *-yo[3]* 'generalized imperfective', *-vuy[3]* 'perfective, but with a frequent sense of surprise/irritation', *-pång[1]* 'neutral perfective', *-rö[3]* 'inceptive', *-r'ra[4]* 'durative, keep on Ving', *-yo[11]yo[2]* 'iterative', *-yå[3]* 'durative', *-råy[4]* 'perseverative'.
- ADVERBIAL: *-bë[4]* 'again, next', *-täng[3]~-t'läng[3]* 'again', *-vë[4]* 'almost' (homophonous but distinct from the substitutive applicative), *-yo[11]ya[2]* 'reluctantly', *-ngöm[1]* 'secretly', *-ngang[3]* 'with great difficulty'.
- NEGATIVE: *-lä[3]*.
- EVALUATION: *-vöyng[1]* 'sympathy', *-lå[3]* 'surprise'.

(29) variably ordered zone:
INTENSIFIER-ASPECT~NEGATIVE~EVALUATION

E. The *clitic zone* is a rigidly ordered group of elements which exhibit looser selectional restrictions in the sense that they can occur after nominals and other lexical categories in addition to their common post-verbal use. These elements co-occur quite readily.

(30) clitic zone:
irrealis=nominalizer=adverbial=realis=evidential=deictic

- IRREALIS: *=a[1]*, a high-frequency element mostly used for indicating future events, but also required marking in desiderative and purposive constructions. This marker evidently derives from the locative case clitic, as reflected in their identical tonal patterning.

- nominalizer: =*nö*[3].
- ADVERBIAL(S): a small number of items, including =*khü*[3] 'just, only', =*pray*[1] 'intensivizer'.
- REALIS: =*bo*[3].
- EVIDENTIAL: =*te*[5], a straightforward hearsay evidential element.
- DEICTIC: =*he*[1], =*to*[1], =*ho*[1], clearly related to independent deictic elements (*hi*[2], *tu*[2], and *hu*[2]); these also have subtle effects on evidentiality and tense interpretations, as typical of such elements in Kuki-Chin languages (cf. Barnes 1998).

(31) reiterates the main parameters distinguishing the elements of the verbal complex other than the root.

(31) summary of the combinatory characteristics of post-verbal elements:

	potential for displacement	variable position	non-selectivity
prefixes	–	–	–
invariable suffixes	–	–	–
transitional (free) particles	+	–	(–)
variably ordered (bound) particles	–	+	–
clitics	–	–	+

Displacement refers to the ability of the element to occur in a position other than post-verbally, illustrated above in (27). Items exhibiting variable position are able to be ordered in multiple ways with respect to each other, sometimes with semantic consequences. Such variability was seen in (29). Non-selectivity refers to the fact that certain items may show an affinity for more than just a verbal base; this applies most clearly to the elements regarded as clitics. The first two types of item are the most canonically affix-like. It would appear that there are certain items in the transitional (free) particle category which are on the verge of grammaticalizing into invariable suffixes, but these are nevertheless less bound. Variably ordered and clitic elements are also less bound for different reasons.

In a more abstract conceptualization, the verb complex might be viewed as containing a verbal core (comprising the verb root, prefixes and invariably ordered suffixes), and more peripheral auxiliary-like entities (the variably ordered elements); the transitional elements in between are more like free adverbs, and the clitic elements are not strictly part of the verbal complex, as indicated also by their ability to attach directly to nominal constituents.

To generalize in terms of verbal categories, it is first apparent that a large number of elements included here are more of a derivational nature, with richly adverbial semantics. Many of these adverbial elements, as well as the classifiers, are highly restricted in lexical terms. Although they are bound elements, this selectivity leaves them with a more lexical or derivational effect than an inflectional one. Similarly, clearly valence-affecting elements like the causatives and benefactive/malefactive are certainly to be treated as derivational.

Quite a few of the more peripherally-ordered elements, though, could be treated as inflectional in nature, including members of the imperfective class, the variably ordered aspect and adverbial elements, the negative marker, and many of the clitics (realis, irrealis, evidential, etc.) The participant coding prefixes are also clearly inflectional in nature, even though they aren't required for every occurrence of a verb.

In sum, then, while many of the morphosyntactically bound post-verbal elements do not appear to represent prototypical inflection in terms of their meaning and regularity, a good number of them do. Khumi thus does not fit the no-inflection Southeast Asian language prototype as well as other languages. However, since the more clearly affixal elements (the prefix zone and the invariably ordered suffix zone) are mostly more derivational in nature, and the elements occurring further to the periphery (the free and bound particles, as well as the clitics) are more inflectional in nature, in some sense Khumi does have a typical Southeast Asian treatment of such elements, which, as could be expected, are less bound.

3.2.3 Serial verb constructions (SVC)

Khumi exhibits only highly limited phenomena resembling verb serialization. This is different from other Kuki-Chin languages which by and large do allow for at least limited verb-verb compounds reminiscent of more prototypical serialization (see, e.g., Chhangte 1993 for Mizo).

The language does have numerous post-verbal elements which have evidently grammaticalized from formerly serialized verbs, however, such as the benefactive marker (*-pë*^{1}), which must have resulted from a serialized verb of giving (*pe*^{1} 'to give'), as most such markers do. In addition, there are a number of semi-bound modal elements, which have grammaticalized meanings distinct from their lexical usages:

(32)

element in bound usage:	**source verb meaning:**
-thay[3] 'general (internal) potential'	'hear'
-p'yå[4] 'impeded (external) potential'	'finish'
-tla[1] 'obligative'	'fall'
-ngay[3] 'desiderative'	'like'
-khëng[1] 'try'	'look'

A small number of similar elements have not undergone semantic shifts between their independent and bound uses: *-kå*[3] 'learned potential', *-nga*[1] 'get (to)', *-håy*[3] 'be good to V', *-cung*[4] 'finish, perfective'. There is also a small, still more loosely bound category which can be regarded as a true auxiliary class (e.g., *a*[1]*håy*[1] 'do together with', *a*[1]*say*[5] 'do again', etc.)

3.3 Clausal/sentential organization

Most non-main clauses in Khumi are loosely subordinated by means of clause final markers, expressing relations such as sequential (*=b'lö*[1], *=kha*[1]), conditional (*=b'lö*[1]), counterfactual/conditional (*=khö*[11]*lö*[1]), concessive (*='ë*[10]*pö*[1], *=kha*[11]*pö*[1]), reason *(='ë*[10], *=nay*[11]*dewng*[11]*kho*[5]*wa*[1]), anterior (*=ma*[1]*b'lö*[1]), and a few rarer elements. These bear a resemblance to chaining or converbial forms, but on the whole are not as invariant in form or bound as prototypical converbs are (e.g., as described for Mongsen Ao by Coupe 2007). In particular, the subordinate clause sometimes bears Khumi's general nominalizer, *=nö*[3] or *=nay*[3] (see the discussion immediately following), or this nominalizer may be absent, with only the marker of subordination present. The first of these possibilities is illustrated for the anterior marker by example (33), and the second is shown in (34).

(33) *niw*[1]***=nö***[11]***=ma***[1]***b'lö***[1] *a*[1]*ti*[3] *k'sewng*[1]*=mo*[4] *p'nö*[11]*-thay*[12]*-lä*[3]
see**=NMLZ=ANT** what flower=QUEST know-POT-NEG

tvåy[11]*ampo*[1]*=wö*[1] *do*[11]*=khö*[11]*lö*[1] *håy*[11]*=a*[1] *n'=pë*[1]*=te*[5]
lover=ENDEARING pick=COND good=IRR QUOT=say-EVID
'When (after) she saw it, she didn't know what kind of flower it was, and she said, "Oh, love, if you pick it, it would be good."' (7.24)

(34) *nuy*[1]*-bå*[1]***=ma***[1]***b'lö***[1] *ni*[11]*cë*[5] *thew*[1]*-så*[1]
push-AUGVCL**=ANT** 3P come out-AUGVCL
'After they pushed it they came out.' (28.30)

Markers of adverbial subordination also may occur in conjunction with a dummy predicate (identical in form to the nominalizer, *=nö*[3], or its variant form *=nay*[3], or either of these combined with a deictic element) as clause-initial stand-alone conjunctions (see example (35)).

(35) *h'nay*[11]*ma*[1]*b'lö*[1] *süng*[11]*-yo*[2] *c'nå*[10] *kang*[10]*-p'yå*[6]*-lä*[11]*=bo*[2]
then bring-IMPFV child.LOC forbid/prevent-POT-NEG=REAL
'Then, they took them along. They couldn't prevent the children [from coming].' (33.144)

Relative clauses are even more firmly grounded in nominalization, usually marked by the element *=nö*[3]. (36) gives examples where the target of relativization bears various roles (A vs. P vs. obliques) internal to the relative clause.

(36) a. A target:
kay[1] *h'ni*[3] *ngo*[1] *abay*[1]***=nö***[3] *s'ra*[10]*=a*[1] *niw*[1]
1S DEM fish cut=**NMLZ** doctor=LOC see
'I saw the doctor who cut the fish.' (elicited)

b. P target:
kay[1] *h'ni*[3] *s'ra*[6]*='ë*[1] *abay*[1]***=nö***[3] *ngo*[10]*=a*[1] *niw*[1]
1S DEM doctor=GEN cut=**NMLZ** fish=LOC see
'I saw the fish that the doctor cut.' (elicited)

c. instrument target:
kay[1] *h'ni*[3] *s'ra*[6]*='ë*[1] *h'ni*[3] *ngo*[1] *abay*[1]***=nö***[3] *hay*[1]*co*[10]*=a*[1] *niw*[1]
1S DEM doctor=GEN DEM fish cut=**NMLZ** knife=LOC see
'I saw the knife the doctor cut the fish with.' (elicited)

d. locative target:
kay[1] *h'ni*[3] *s'ra*[6]*='ë*[1] *h'ni*[3] *ngo*[1] *abay*[1]***=nö***[3] *üng*[10]*=a*[1] niw[1]
1S DEM doctor=GEN DEM fish cut=**NMLZ** house=LOC see
'I saw the house the doctor cut the fish in.' (elicited)

Note the uniform treatment of the A participant in (36b-d) as a possessor. In addition, a couple of other nominalized forms are found less-frequently supporting relativization, including the dedicated locative nominalizer *-ra*[1] seen in (37), where it marks a headless relative.

(37) *uh* *ne^{11}=yö^{1}* *ngam^{1}ngampo^{1}* *åm^{1}-ra^{1}=lö^{1}*
INTERJ old.sister=ENDEAR parents exist-LOC. NMLZ=TOP

pnö^{11}=nö^{1}=cö^{4}=ö^{1}
know=NMLZ=INSIST=ENDEAR
‘ “Oh elder sister, I know where your parents are living.” ’ (1.70)

3.3.1 Ellipsis of arguments

Like many other area languages, Khumi is extremely permissive when it comes to ellipsis of arguments which are recoverable from discourse context. As a simple example, (38) was spoken in response when someone had asked whether the speaker had any skewers of grilled pork on hand:

(38) *hi^{11}* *ø* *bö^{6}*
here.LOC (skewers of pork) exist.NEG
‘There aren’t any here.’

In the following narrative text example, the consultant who assisted in translation amazingly identified the zero anaphor associated with the predicate ‘exist’ with *s’i^{1}plang^{1}* ‘sand bank’, which had not been mentioned for about twenty intervening clauses!

(39) *tläm^{3}* *khåy^{1}tewng^{5}=mö^{3}* *p’yung^{11}-pë^{1}=pray^{1}=lö^{1}* *nang^{1}pö^{1}*
suddenly (name)=FGD urinate-MAL=INTENS=TOP DISC.PART

bö^{6}-lä^{11}=te^{1} *a^{1}yö^{1}* *miwng^{2}* *nüng^{2}* *lüng^{11}-täng^{11}=te^{1}*
exist-NEG=EVID man cubit two rise-AGAIN=EVID
‘Suddenly Khåytewng peed a lot on her, you know; it (ø=the sandbank) didn’t exist, man! It (ø=the water) rose again by two cubits.’ (42.115)

However, zero anaphoric links are occasionally inadequate, as seen in the following exchange between a storyteller and a listener, in which there is apparently more than one possible referent for the zero.

(40) A: *kang^{1}khüng^{11}-ka^{1}ngay^{11}=ma^{4}* *a^{1}tung^{1}-ka^{1}-t’nga^{11}=te^{1}*
finally-ELAB=LOC pound-AUGVCL-AT.LAST=EVID
‘Finally he pounded him to death with difficulty.’

B: *tuy¹nang¹⁰=a¹*
water.spirit.LOC=INTERR
'The water-spirit?'

A: *tuy¹nang¹⁰=a¹*
water.spirit=LOC
'The water-spirit.' (42.133-5)

In this narrative context, a man is battling a water-spirit, and A's statement makes it unclear which participant is victorious. B asks if it was the water-spirit that was defeated in the second line, and A confirms B's intuitions.

3.3.2 Information structure

Besides the extensive argument ellipsis for discourse-recoverable entities just illustrated, Khumi has a dedicated topic marker, *=lö¹*, with a discourse use approximating that of well-known topic markers like Japanese *=wa*. This element occurs in one of the discourse marker slots already described above under basic noun structure (see section 4.1.1). In addition, the marker *=mö³*, which occurs in a less peripheral NP slot, serves to focus its NP in some cases. Since *=mö³* is actually compatible with *=lö¹*, and since different elements are responsible for focus under other circumstances (most notably, *=pö¹*, which otherwise means 'also'[8]), it is analyzed not as a focus marker, but as a marker of foregrounding. If both markers occur in a single NP, their order is invariably *=mö³=lö¹*.

The following sentences illustrate the basic focus types identified by Lambrecht (1994) in appropriate discourse contexts, and the occurrence of the markers in question. In the first sentence, we see a predicate focus structure, in response to a question like 'What did the hill doctor do?' As expected in such structures, a topic may be marked by the topic marker, as in (41).

(41) *s'ra¹(=lö¹)* *biw²* *ca¹-vuy³*
hill.doctor(=TOP) rice eat-PFV
'The hill doctor ate (rice).' (elicited)

In (42), the entire sentence is in focus, as in response to a question like 'What happened?'

8 This clitic is similar in at least certain respects to the particle treated in Konnerth 2012.

(42) *s'ra[1]=mö[3]* *biw[2]* *ca[1]-vuy[3]*
hill.doctor=FGD rice eat-PFV
'The hill doctor ate (rice).' (elicited)

Finally, in (43), the *=mö[3]*-marking and cleft-like structure allow narrow focus on a participant, in this case the hill doctor, as a contradiction to a hypothetical claim that someone other than the hill doctor ate.

(43) *mhm,* *s'ra[1]=mö[11]=tew[2]* *biw[2]* *ca[1]-vuy[11]=nö[2]*
interj hill.doctor=FGD=COP rice eat-PFV=NMLZ
'No, it was the hill doctor that ate (rice).' (elicited)

For further details on the use of these markers, the reader is directed to Peterson 2011, where they are given a more exhaustive treatment.

3.3.3 Other observations

Khumi has a robust use of so-called 'stand alone' nominalization commonly seen in Tibeto-Burman languages (see Matisoff 1972, Noonan 1997), where a nominalized clause appears as if it were independent, and at least in Khumi, often with special emphatic effect:

(44) a. *nay[11]b'lö[1]* *thang[11]bë[1]=te[5]* *sam[1]rüng[1]* *ka[1]müng[5]*
then nightingale=EVID chili red

ca[10] *jew[1]=nö[3]*
eat.IRR come=NMLZ
'Then a nightingale came to eat red chilis.' (1.75)

b. *lam[1]phay[2]* *ke[3]* *në[1]na[1]* *ke[3]* *lam[1]phay[2]* *ke[3]* *në[1]na[1]* *ke[1]=nö[3]*
weaving.belt bite skirt bite weaving.belt bite skirt bite=NMLZ
'They (dogs) bit the weaving belt and they bit her skirt, they bit the weaving belt and they bit her skirt.' (8.195a)

c. *a[11]=mö[2]* *klay[10]=nö[2]*
chicken=FGD scratch=NMLZ
'The chickens scratched her.' (8.195b)

In theory, this construction arose in Khumi via simplification of another commonly attested construction: *V=nö³ tla¹*, seen in (45), which has an interpretation basically indistinguishable from a simple past tense.

(45) *nay¹¹b'lö¹* *ca¹-vuy¹¹=b'lö¹* *k'ni⁵* *thung⁵* *p'lü⁵=b'lö¹*
then eat-PFV=SEQ day three four=SEQ

t'vöng¹¹=pö¹ *a¹dång¹⁰=nö²* *tla¹* *ni³=mö⁶* *c'niw¹⁰=a¹*
bear=ALSO miss=NMLZ happen 3S=REFL.GEN daughter=LOC
'...then, he (tiger) ate her (bear's daughter), and after three or four days, bear also, missed her, his own daughter.' (3.65)

In this more complex construction, *tla¹*, literally 'fall', appears to be used in a sense of 'happen', such that the entire construction means something like 'it happened that V'.

4 Semantics and pragmatics

4.1 Common semantic domains

4.1.1 Food terminology

In the food domain, there is a basic division between *biw²* 'cooked rice' and *ang¹* 'curry', or what accompanies the rice. As expected, there are many different rice-related terms. Aside from specific terms for different varieties of rice grown by the Khumi, which are quite numerous, the following are several relatively basic terms. However, some of these are obviously compounded.

(46) selected rice-related terminology:

biw²	'cooked rice, food'
co⁵	'uncooked, unpounded rice'
co¹¹ngang⁵	'rice which is still wet but in the process of drying'
co¹¹hung¹	'rice which still has its outer covering on it'
co¹¹nöyng¹	'rice with the outer covering removed'
co¹¹m(u)¹	'rice (seed)'
cang¹tha⁵	'new rice, first rice harvested from the swidden'
cang¹¹tlöyng¹	'sticky rice'
cang¹hüng⁵	'old rice, rice carried over from one harvest to the following year'

ca¹liwng⁵	'sticky rice'
pew¹pew⁵	'puffed rice'

This hardly scratches the surface of the lexical domain related to rice.

4.1.2 Expressions for 'cutting', 'carrying', 'drying', 'directional'

Khumi differs from a number of closely related and certain other Mainland Southeast Asian languages in lacking a large number of directional elements (see DeLancey 1980, So-Hartmann 2009). However, for other lexical domains it is typical in having numerous clearly delineated actions. These include domains of cutting, carrying, wearing, hitting/beating, digging, and washing. I include just a couple of partial lists in (47) and (48):

(47) carrying:

pho³	'carry with a headstrap, in a basket, on the back'
süng³	'carry in the hand'
a¹pu¹	'carry over the shoulder'
a¹tläng⁴	'carry on the head or back, balance on the shoulder'
c'këng⁴	'carry or hold under the arm, at the side'
t'ke²	'carry behind the ear'
a¹jewng²	'carry on the shoulder (of two people)'
pew¹	'carry on the back (e.g., a child)'
t'pång⁴	'carry under or in the arm'
tång⁴	'carry on the back in a basket'
vä¹	'carry a hanging object'
t'va¹	'carry on the shoulder (e.g., a carry-bag)'

(48) wearing:

a¹na¹	'wear on the lower half of the body'
am¹co²	'wear in the ear (e.g., a flower or earring)'
åy⁴	'wear around the neck'
a¹jeng⁴	'wear on the head (e.g., a turban)'
a¹büng²	'stick or wear in the hair'
a¹khiw¹	'wear on the upper half of the body'
a¹diwng²	'wear on the head'
a¹t'ka¹	'wear on the face'
am¹cu²	'wear on the finger or wrist'

4.2 Pragmatics and discourse

4.2.1 Final particles

In keeping with its Southeast Asian areal status, Khumi is rich in final particles indicating all sorts of subtle subjective evaluations or emotional content. Glosses for most of these can only be tentative. They include: *=mi*[1] 'dubitative', *='ö*[1] 'endearing', *=ba*[1]*~=be*[4]*~=bä*[4] 'softening', *=e*[1] 'affirmative', *=cö*[4] 'insistent', *=ma*(*ng*)[1] 'intensified affirmative', *=n'ang*[5] and *=l'o*[5], which are both roughly like an English tag question in which the speaker is not actually trying to get confirmation but rather to get the listener to believe what they are saying.

4.2.2 Politeness

Politeness is not expressed in Khumi by any of the usual means found in major Southeast Asian languages: there are no pronominal distinctions, special verbal morphology, or strict lexical choices used to indicate relative speaker status or deference.

Khumi discourse does exhibit the common South and Southeast Asian practice of using kinship terms in an extended manner for persons who are not members of one's actual family, e.g., 'grandson' used by an elderly person for males belonging to the generation of one's own grandchildren. The lexicon also distinguishes carefully between whether a person/addressee is a member of one's own clan or not. However, both of these are arguably separate phenomena from systems for grammatical encoding of politeness distinctions.

Imperatives are either bare (i.e., marked by a verb root or extended verb complex with the appropriate tonal variant on the final syllable of the complex), or they may involve a variety of further devices, such as a final particle (see section 4.2.1) for mitigating the abruptness of the imperative. This also is not a manifestation of politeness per se, but according to speakers an abrupt imperative rather than a more extended request form would be less likely to be used by persons of inferior social status towards those of a higher standing, or by a younger person to an elder.

5 Summary

All things considered, Khumi is clearly a prototypical Southeast Asian language closely allied with the Sinosphere in the vast majority of respects. Nothing shows this status better than its phonology, especially in terms of its sesquisyllabic structures, its fairly extensive vowel inventory, and its exceedingly complex system of contour tones and their interactions. While it has–albeit only minimal–participant-coding on its verbs, deviating from the Southeast Asian norm, it has ideophonic elements, widespread use of elaborate expressions and other parallel structures, a number of reduplication types, and at least a smattering of psycho-collocations and numeral classifiers.

Khumi has no true serialization, but it has structures which clearly arose from verb-verb compounding, which in turn presumably grew out of serial-like constructions. In terms of word structure, it clearly has monosyllabic or sesquisyllabic phonological words. It makes liberal use of argument ellipsis, and its lexicon is partitioned in ways that are familiar in the Southeast Asian context.

The ways in which Khumi deviates from the Southeast Asian prototype hinge clearly on its family membership more than its areal position. Its verb-final syntax making use of postpositional case markers and discourse status indicators, and converb-like clause combining strategies are squarely Tibeto-Burman in character. The more complex picture of word structure which consideration of the issue from the perspective of morphosyntactic wordhood reveals is highly typical for Kuki-Chin and other nearby portions of the family.

Given its position on the border between Southeast Asia and South Asia, and its central position within Tibeto-Burman, this generous mix of areal and genetic heritages is only to be expected.

Abbreviations

1	first person
2	second person
3	third person
ABL	ablative
AFFIRM	affirmative
AGR	agreement
ALL	allative
ANT	anterior
APPL	applicative
AUG	augmentative
AUGVCL	augmentative verbal classifier

AUX	auxiliary
BEN	benefactive
CAUS	causative
CLF	classifier
COLL	collective
COND	conditional
D	dual
DEM	demonstrative
DIMVCL	diminutive verbal classifier
DUBIT	dubitative
ELAB	elaboration (in an elaborate expression)
EMOT	emotive
ENDEAR	endearing
EVID	hearsay evidential
EXCL	exclusive
FGD	foregrounder
GEN	genitive
HORT	hortative
IDEO	ideophonic element
IMPER	imperative
IMPV	imperfective
INCL	inclusive
INSIST	insistent
INST	instrumental
INTENS	intensive
INTERJ	interjection
INTERR	interrogative
IRR	irrealis
MAL	malefactive
LOC	locative
NEG	negative
NMLZ	nominalizer
P	plural
PERS	perseverative
PFV	perfective
POT	potential
PROX	proximal
QUEST	question
QUOT	quotative
REAL	realis
REAS	reason
S	singular
SEQ	sequential
TOP	topic

References

Aikhenvald, Alexandra Y. 2005. Serial verb constructions in typological perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions: A cross-linguistic typology*, 1–68. Oxford: Oxford University Press.

Barnes, Jonathan. 1998. Tsuu khaa tii hla?: deixis, demonstratives and discourse particles in Lai Chin. *Linguistics of the Tibeto-Burman Area* 21 (1). 53–86.

Barz, R. K. & A. Diller. 1985. Classifiers and standardization: Some South and South-East Asian comparisons. in David Bradley (ed.), *Papers in Southeast Asian Linguistics No. 9, Pacific Linguistics*, 155–184. Canberra: Australian National University.

Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description – Vol 3: Grammatical categories and the lexicon*, 2nd ed., 169–240. Cambridge: Cambridge University Press.

Bradley, David. 1997. Tibeto-Burman languages and classification. In David Bradley (ed.), *Papers in Southeast Asian linguistics No 14: Tibeto-Burman languages of the Himalayas,* 1–72. Canberra: Australian National University.

Chelliah, Shobhana. 2009. Semantic role to new information in Meithei. In Jóhanna Barðal & Shobhana Chelliah (eds.), *The role of semantics, pragmatics, and discourse in the development of case*, 377–400. Amsterdam: John Benjamins.

Chhangte, Lalnunthangi. 1993. *Mizo syntax*. PhD dissertation, University of Oregon.

Coupe, Alexander R. 2007. *A grammar of Mongsen Ao*. Berlin & New York: Mouton de Gruyter.

Delancey, Scott. 1980. *Deictic categories in the Tibeto-Burman verb*. PhD dissertation, Indiana University.

Dixon, R. M. W. & Alexandra Y. Aikhenvald. 2002. Word: A typological framework. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word: A cross-linguistic typology*, 1–41. Cambridge: Cambridge University Press.

Grierson, G. A. (ed.) 1904. *Linguistic survey of India, vol. III: Tibeto-Burman family (3 parts)*. Calcutta: Office of the Superintendent, Government Printing.

Haas, Mary. 1964. *Thai-English student's dictionary*. Palo Alto: Stanford University Press.

Hall, T. A., K. Hildebrandt & B. Bickel. 2008. Introduction: Theory and typology of the word. *Linguistics* 46. 183–192.

Haspelmath, Martin. 2005. Argument marking in ditransitive alignment types. *Linguistic Discovery* 3 (1). https://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/280?htmlOnce=yes.

Henderson, E. J. A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15. 400–434.

Hyman, Larry & Kenneth VanBik. 2004. Directional rule application and output problems in Hakha Lai tone. *Phonetics and Phonology, Special Issue, Language and Linguistics* 5. 821–861.

Kari, James. 1989. Affix positions in the Athapaskan verb complex: Ahtna and Navajo. *International Journal of American Linguistics* 55 (4). 424–454.

Konnerth, Linda. 2012. Additive focus and additional functions of Karbi (Tibeto-Burman) *=tā*. *Berkeley Linguistics Society* 38. 206–221.

Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.

Matisoff, James A. 1972. Lahu nominalization, relativization, and genitivization. In John P. Kimball (ed.), *Syntax and semantics* (1), 237–257. New York: Academic Press.
Matisoff, James A. 1973. *The grammar of Lahu*. Berkeley: University of California Press.
Muysken, Pieter. 1986. Approaches to affix order. *Linguistics* 24. 629–643.
Noonan, Michael. 1997. Versatile nominalization. In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type*, 373–394. Amsterdam: John Benjamins.
Peterson, David A. 2002. (published 2006) On Khumi verbal pronominal morphology. *Berkeley Linguistics Society* 28 [special session]. 99–110.
Peterson, David A. 2008. Bangladesh Khumi verbal classifiers and Kuki-Chin 'chiming'. *Linguistics of the Tibeto-Burman Area* 31 (1). 109–138.
Peterson, David A. 2010. Khumi elaborate expressions. *Himalayan linguistics* 9 (1). 1–20.
Peterson, David A. 2011. Core participant case marking in Khumi. *Linguistics of the Tibeto-Burman Area* 34 (2). 73–100.
Peterson, David A. 2013a. Aesthetic aspects of Khumi grammar. In Jeffrey P. Williams (ed.), *The aesthetics of grammar: Sound and meaning in the languages of mainland South-east Asia*, 219–236. Cambridge: Cambridge University Press.
Peterson, David A. 2013b. Affecting valence in Khumi. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency: In honor of Johanna Nichols*, 171–193. Amsterdam: John Benjamins.
Peterson, David A. 2017. On Kuki-Chin subgrouping. In Picus S. Ding & Jamin Pelkey (eds.), *Sociohistorical linguistics in Southeast Asia: New horizons for Tibeto-Burman studies in honor of David Bradley,* 189–209. Leiden: Brill.
So-Hartmann, Helga. 1988. Notes on the Southern Chin languages. *Linguistics of the Tibeto-Burman Area* 11 (2). 98–119.
So-Hartmann, Helga. 2009. *A descriptive grammar of Daai Chin*. STEDT Monograph 7, UC Berkeley.
Solnit, David. 1995. Parallelism in Kayah Li discourse: Elaborate expressions and beyond. *Berkeley Linguistics Society* 21 [special session]. 127–140.
Solnit, David. 1997. *Eastern Kayah Li: Grammar, Texts, Glossary*. Honolulu: University of Hawai'i Press.
Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press.
VanBik, Kenneth. 2009. *Proto-Kuki-Chin: A reconstructed ancestor of the Kuki-Chin languages*. Berkeley: STEDT Project, UC Berkeley.
Watkins, Justin. 2013. A first account of tone in Myebon Sumtu Chin. *Linguistics of the Tibeto-Burman Area* 36 (2). 97–127.

Appendix 1: Summary of linguistic features

Legend
+++ the feature is pervasive or used obligatorily in the language
++ the feature is normal but selectively distributed in the language
+ the feature is merely possible or observable in the language
– the feature is impossible or absent in the language

	Feature	+++/++/+/–	§ ref. in this chapter
Phonetics	Lexical tone or register	+++	§1.1, p.13
Phonetics	Back unrounded vowels	+	§1.2, p.18–19
Phonetics	Initial velar nasal	+	not discussed explicitly
Phonetics	Implosive consonants	–	not discussed explicitly
Phonetics	Sesquisyllabic structures	+++	§1.3, p.19
Morphology	Tendency towards monosyllabicity	++	§2.1, p.19
Morphology	Tendency to form compounds	+++	§2.1, p.20
Morphology	Tendency towards isolating (rather than affixation)	++	§2.1, p.20
Morphology	Psycho-collocations	+	§2.2, p.21
Morphology	Elaborate expressions (e.g. four-syllable or other set patterns)	+++	§2.3, p.22
Morphology	Reduplication generally	++	§2.4, p.22
Morphology	Reduplication of nouns	+	not discussed explicitly
Morphology	Reduplication of verbs	+	not discussed explicitly
Grammar	Use of classifiers	+	§3.1.2, p.28
Grammar	Classifiers used in counting	+	§3.1.2, p.28
Grammar	Classifiers used with demonstratives	–	–
Grammar	Adjectival verbs	+++	§3.1.1, p.26
Grammar	Grammatical number	+	§3 and §3.1.1, p.24
Grammar	Inflection of verbs	++	§3.2.1, p.30
Grammar	Use of tense/aspect markers	+++	§3.2.2, p.32–36
Grammar	Use of verb plural markers	+	§3.2.2, p.32
Grammar	Grammaticalization of GET/OBTAIN (potential mod. resultative/perfect aspect)	+	§3.2.3, p.37
Grammar	Grammaticalization of PUT, SET (completed/resultative aspect)	–	not discussed explicitly
Grammar	Grammaticalization of GIVE (causative, benefactive; preposition)	+++	§3.2.3. not discussed explicitly for causatives, though this is the probable source for prefixal causative, p.36
Grammar	Grammaticalization of FINISH (perfective/complete aspect; conjunction/temporal subordinator)	+	§3.2.3, p.37

	Feature	+++/++/+/–	§ ref. in this chapter
Grammar	Grammaticalization of directional verbs e.g. GO / COME (allative, venitive)	–	not discussed explicitly; Khumi has limited directional phenomena without clear grammaticalization sources; it also has fossilized remnants of a probably Proto-Kuki-Chin-level prefixal venitive, which had COME as its source
Grammar	Grammaticalization of SEE, WATCH (temptative)	+	§3.2.3, p.37
Grammar	Grammaticalization of STAY, REMAIN (progressive and continuous, durative aspects)	–	not discussed explicitly
Grammar	Serial verb constructions	–	§3.2.3, p.36
Syntax	Verb precedes object (VO)	–	–
Syntax	Auxiliary precedes verb	–	–
Syntax	Preposition precedes noun	–	§3.1, p.24
Syntax	Noun precedes adjective	+	§3.1.1, p.25
Syntax	Noun precedes demonstrative	–	§3.1.1, p.25
Syntax	Noun precedes genitive	–	–
Syntax	Noun precedes relative clause	+	§3.1.1, p.25–26
Syntax	Use of topic-comment structures	+++	§3.3.2, p.40 & also p.27
Syntax	Ellipsis of arguments known from context	+++	§3.3.1, p.39
Lexical semantics	Specific terms for forms of rice	+++	§4.1.1, p.42
Pragmatics	Use of utterance-final pragmatic particles	+++	§4.2.1, p.44
Pragmatics	Encoding of politeness	–	§4.2.2, p.44
Pragmatics	Encoding of honorifics	–	–

Appendix 2: Interlinearized Text

Treeshrew and Owl

1. *tew^{11}su^{1}=håy^{1}* *bew^{1}ku^{10}=håy^{1}* *am^{1}nay^{11}=nö^{1}=te^{1}=ba^{1}*
   treeshrew=COM owl=COM be.friends=NMLZ=EVID=EMOT
   'Treeshrew and Owl were friends.'

2. *a¹håy¹=nö¹¹=te¹*
   hang.out=NMLZ=EVID
   'They hung out together.'

3. *ang¹lo¹¹-p'vay¹¹* *ce¹=bo¹¹=te¹*
   girl-visit.IRR go=REAL=EVID
   'They went visiting girls (=courting.)'

4. *ang¹lo¹¹-p'vay¹¹* *ce¹=kha¹⁰* *lang¹⁰* *tko¹¹=kha¹⁰* *a¹rung¹⁰=nö¹=te⁵*
   girl-visit.IRR go=time.LOC road.LOC go=time.LOC converse=NMLZ=EVID
   'When they went courting, when going along the road, they talked with each other.'

5. *tew¹¹su¹=mö³* *(ö)* *bew¹ku¹⁰=wo¹* *ay¹¹ni-²* *ay¹¹ni²* *ang¹lo¹¹-pvay¹¹*
   treeshrew=FGD uh owl=VOC 1D.INCL 1D.INCL girl-visit.IRR

   ce¹=nö¹=a¹ *nö¹=te⁵*
   go=NMLZ=QUEST QUOT=EVID
   'Treeshrew said, uh, "Oh, Owl, are we, are we going courting?"'

6. *ay¹¹ni²* *m'nö¹=mö³* *cång¹=nö¹¹=a¹* *nö¹=te⁵*
   1D.INCL how=FGD (mistake) be.sold=NMLZ=INTERR QUOT=EVID
   '"How will we be sold? (=are we going to be handsome enough?)" he said.'

7. *tew¹¹su¹=mö³* *(ö)* *nang¹=lö¹* *kay¹=lö¹* *l'bewng¹* *p'suy¹⁰-ka¹mo¹='e¹* *vay¹⁰*
   treeshrew=FGD uh 2S=TOP 1S=TOP mouth pointy-INTENS=AFF later.LOC

   ang¹lo¹¹=cë¹¹=mö² *döyng¹¹=b'lö¹* *thuy¹¹-pë²* *kay¹=lö¹* *k'lung⁴* *kå¹¹=nö²*
   girl=COLL=FGD ask=SEQ say-BEN.IMP 1S=TOP song POT=NMLZ
   'Treeshrew said, uh, "You, I have a very pointy mouth. Later, if the girls ask, say that I know how to sing songs."'

8. *(ö)* *a¹lung⁵* *öyng¹¹-kå¹=nö²* *b'lüng⁴* *öyng¹¹-kå¹=nö¹=nay¹ra⁴* *l'bewng¹*
   uh gourd.flute play-POT=NMLZ flute play-POT=NMLZ=REAS mouth

   p'suy¹⁰=nö¹=tew¹=ba¹ *nay³* *thuy¹¹-pë²=ba¹* *bew¹ku¹⁰=wo¹* *nö¹=te⁵*
   pointy=NMLZ=COP=EMOT QUOT say-BEN.IMP=EMOT owl=VOC QUOT=EVID
   ' "Say that it's because I can play the gourd flute and the flute, that my mouth is pointy, Owl," he said.'

9. *n'b'lö*[1] *bew*[1]*ku*[10]*=mö*[2] *thuy*[11]*-pë*[1]*-bë*[10]*=bo*[1]*=te*[1] *tew*[11]*su*[10]*=wa*[1] *(ö)* *kay*[1]*=lö*[1]
then owl=FGD say-BEN-NEXT=REAL=EVID treeshrew=LOC uh 1S=TOP

möy[3] *lëng*[11]*-ka*[1]*mo*[1]*=nö*[3] *vay*[10] *ang*[1]*lo*[11]*=cë*[11]*=mö*[2] *thuy*[11]*=b'lö*[1] *nang*[6]
eye big-INTENS=NMLZ later.LOC girl=COLL=FGD say=SEQ 2S.GEN

ahåy[1]*=lö*[1] *möy*[3] *lëng*[11]*-pray*[1]*=ba*[1] *m'nö*[1]*-m'nay*[11]*=lew*[1]
friend=TOP eye big-INTENS=EMOT how-how=QUEST

h'nay[3] *möy*[11]*=lö*[1] *nö*[1]*=te*[5]
thus eye=TOP QUOT=EVID
'Then Owl said to Treeshrew, uh, "I have really big eyes. Later, if the girls say, 'Your friend has really big eyes. What's up with such eyes?'"'

10. *nay*[11]*b'lö*[1] *prë*[1] *vä*[1] *khu*[1] *vä*[1]*-rë*[1] *niw*[1]*-thay*[11]*=nö*[2] *nay*[3] *möy*[11]*=nay*[1]*ra*[4]
then country all ELAB all-count see-POT=NMLZ thus eye=REAS

kay[6] *möy*[3] *lëng*[11]*=nö*[2] *nay*[3] *thuy*[11]*-pë*[2]*=ba*[1] *tew*[11]*su*[1]*=wö*[1] *nö*[1]*=te*[5]
1S.GEN eye big=NMLZ QUOT say-BEN.IMP=EMOT treeshrew=ENDEAR QUOT=EVID
'Then he said, "They are such eyes that I can see all over the country. That's why my eyes are big, you say, Treeshrew."'

11. *h'nay*[3] *h'nay*[3] *lew*[1]*-a*[1]*dew*[1] *kha*[10] *tä*[1]*-yo*[11]*=bo*[1]*=te*[1]
thus thus word-agree time.LOC go-IMPFV=REAL=EVID

ang[1]*lo*[11]*=teng*[10]*=a*[1] *ang*[1]*lo*[11]*-p'vay*[11]*=nö*[2]
girl=GOAL=LOC girl-visit=NMLZ
'When they agreed this way and that, they went to the girls' place, they went courting.'

12. *nay*[3] *ang*[1]*lo*[11]*=teng*[10] *töng*[10]*=ma*[1]*b'lö*[1] *ang*[1]*lo*[11]*-p'vay*[11]*=nö*[2] *(ö)*
thus girl=GOAL.LOC arrive=ANT girl-visit.IRR=NMLZ uh

am[1]*rü*[1]*=nö*[11]*=bo*[2]
get.ready=NMLZ=REAL
'So, when they arrived at the girls' place, they got ready to court them.'

13. *ang*[1]*lo*[11]*=cë*[11]*=teng*[10] *a*[1]*tä*[10]*=ma*[1]*b'lö*[1] *ang*[1]*lo*[11]*=mö*[2] *döyng*[11]*=bo*[1]*=te*[1]
girl=COLL=GOAL.LOC sit=ANT girl=FGD ask=REAL=EVID

tew[11]*su*[1]*=wa*[1] *tew*[11]*su*[1] *(a)* *bew*[1]*ku*[10]*=teng*[11] *döyng*[11]*=bo*[1]*=te*[1] *bew*[1]*ku*[10]*='ö*[1]
treeshrew=LOC treeshrew er owl=GOAL.LOC ask=REAL=EVID owl=ENDEAR