lpehr

Languages and Peoples of the Eastern Himalayan Region **(LPEHR) comprises an annual special issue of** ***Himalayan Linguistics*** **together with occasional special publications. Peer-reviewed and open-access, it focuses on linguistic and ethnographic documentation and description in the Eastern Himalaya.**

Languages and Peoples of the Eastern Himalayan Region (LPEHR)

Tangsa-Nocte as a Continuum: A diagnostic feature list for classification of varieties

Kellen Parker van Dam
Universität of Zürich
La Trobe University

Syed Iftiqar Rahman
Chiang Mai University

ABSTRACT

Numerous languages of North East India are classified according to extralinguistic factors such as district and state borders or similarities in culture. This has resulted in highly diverse language varieties — often with little or no mutual intelligibility — falling under a single label despite these considerable linguistic and sometimes cultural differences. Likewise, varieties which are closely related or lacking any meaningful differences may be classified as distinct entities, such as with the Phong variety which is classified as either Tangsa or as Nocte depending on the district in which the speaker resides. Based on an analysis of sound correspondences and lexical variation between varieties, this paper argues the case for treatment of Tangsa-Nocte not as two closely related branches within the Sal languages, as earlier classifications may suggest, but rather as a single dialect continuum. Furthermore, some diagnostic tests are provided in order to help determine placement of newly documented varieties within the larger group.

KEYWORDS

Tibeto-Burman, Sal, Tangsa, Nocte

This is a contribution from *Himalayan Linguistics Vol 20(2) – Languages and Peoples of the Eastern Himalayan Region: 61-74.*
ISSN 1544-7502

Himalayan Linguistics Vol 20 (2) – Languages and Peoples of the Eastern Himalayan Region
 ISSN 1544-7502

Tangsa-Nocte as a Continuum: A diagnostic feature list for classification of varieties[†]

Kellen Parker van Dam
Universität of Zürich
La Trobe University

Syed Iftiqar Rahman
Chiang Mai University

1 Introduction

Within Northeast India, a large number of languages and cultures are organised for official purposes based on the geography and visible cultural practices of the communities in which these language varieties are spoken. In addition to the larger questions of "Naga" as a linguistically meaningful label between the so-called Northern Naga (central Sal) varieties and those of the Central-Southern Naga group — see Post & Burling (2017:228) for further discussion on this topic — the tendency to group based on extralinguistic features has resulted in classificatory problems within the various Northern Naga subgroups as well. This paper focuses on the problem as it relates to two such "groups" of language varieties spoken in and around Changlang, Arunachal Pradesh, namely Tangsa and Nocte.

It is clear from previous work on these varieties (Morey 2019b, van Dam 2018a, Statezni 2013) that Tangsa and Nocte varieties — as well as varieties such as Ollo and Tutsa — share many features. Such features include semantic drift of lexical items not seen more generally in Sal varieties, the use of a reduplicative system of intensifiers not generally in other Northern Naga varieties (van Dam 2018b) and a common tone system with a single shared origin (van Dam 2018a). Rather than having a Tangsa branch and a Nocte branch under which all varieties would be classed, we may instead imagine a prototypical "Nocte"-like set of features on one end of a continuum, with a prototypical set of "Tangsa"-like features on the other end, between which the many different language varieties would fall. Such features include the etyma for certain lexical

[†] Data used for this study has been collected by a large number of people over many years without whose help the work could not be done. This includes Stephen Morey, Mijke Mulder, Dipjyoti Goswami and Martin Dietzel. We are also grateful to our many consultants across North East India and Upper Myanmar and the community linguists among them whose contributions have been invaluable. In particular, thanks to Wanglung Keluim Mossang for assistance in arranging and interpreting of many of the stems being analysed. Thank you as well to the anonymous reviewers who provided comments and corrections. Any remaining errors are our own.

items such as `elephant' as *#bok*[1] vs *#tʃaŋ*, whether or not certain distinctions are attested such as a difference between BLACK and DARK with *#mak* and *#ɲak*, and the phonemic realisation of various onsets and rimes based on a reconstructed proto Tangsa-Nocte.

The purpose for this work is to assist in an ongoing attempt to classify the relationships between all of the Tangsa-Nocte varieties. This is being done both for its own sake, as well as to help better understand the linguistic history and both migrations and contact situations in the unrecorded past. This in turn might help paint a clearer picture of when and from whence these migrations may have occurred. Efforts at historical reconstruction also benefit from having a picture of internal relationships, which in turn will assist in placing Tangsa-Nocte and Nothern Naga more generally within the recent phylogenetic trees of Sino-Tibetan (Sagart et al 2019; Zhang et al 2019) which have not taken the larger group into account.

In the following sections, We will present a brief set of features which can be used as a sort of diagnostic checklist. This is following in the spirit of Jerry Norman's list of features with which he classified the larger Sinitic languages as belong to either a Northern group, a Southern group, or a transitional language between the two (Norman 1988). Additionally, the case will be made for classification of Tutsa and Ollo among the Tangsa-Nocte varieties.

It should be noted that Tangsa both and Nocte have received recognition from the state government of Arunachal Pradesh as recognised Naga subtribes[2], listed in the 2001 census among the roughly one hundred other recognised tribes. Prior to the 2011 census, Tutsa was not recognised as its own subtribe in Arunachal Pradesh; Tutsa speakers were classified within Tangsa and Nocte. More recently, in August 2021, the Wancho[3], Tangsa, Nocte and Tutsa subtribes received more complete recognition as part of the Scheduled Tribes within Arunachal Pradesh, rather than falling under the category of "any Naga subtribes". This was accomplished by the passing of The Constitution (Scheduled Tribes) Order (Amendment) Bill, 2021 by the state's Ministry of Tribal Affairs.

With official recognition of Tutsa as distinct from Tangsa, happening as recently as 2001, we the authors must make clear that our intent in confirming alignment of the Tutsa language within the linguistic continuum here is not meant as a rejection of Tutsa identity or uniqueness of the linguistic varieties. Rather, it is simply a statement of linguistic pedigree. A more inclusive name for the Tangsa-Nocte dialect continuum would be Tangsa-Nocte-Tutsa-Ollo; Tutsa is linguistically very close to “Tangsa” varieties such as Phong, Nocte and Muklom, despite having a distinct political status. In the interest of concision, however, that longer label

[1] Note that proto-Tangsa-Nocte forms are given with the hash symbol (#) here, rather than an asterisk, in order to indicate that these are not yet a fully worked out set of reconstructions. Reconstruction of proto-Tangsa-Nocte is in preparation for future publication.

[2] The terms "tribe" and "subtribe" here are used follow with official recognition as part of the Scheduled Tribes of India. These are official designation in India, and the terms are not meant here in any derogatory manner. While the term is problematic, something which has gotten consierable attention within the anthropological literature, for our purposes here we will use the terms to reflect official state designations.

[3] It must also be mentioned that Wancho is included in Khan Lann (2017) as part of a larger Tangshang group. The classification given in Khan Lann is not without problems, but more importantly, the notion of Tangshang reflects a larger cultural grouping, and not one that is well founded on solely linguistic grounds. Wancho lacks a number of defining features of the Tangsa-Nocte continuum, including an ABB pattern of reduplicated modifiers (van Dam 2018b; 2020; forthcoming) and possibly toneme cognacy, although recent work by Morey & Losu (forthcoming) may show the latter to indeed be shared.

has been eschewed in this paper. It should be taken as the intended reading of what the term "Tangsa-Nocte" as used here actually represents, sacrificing accuracy for brevity in this case.

By addressing Tangsa-Nocte as a dialect continuum, rather than as two or three closely related languages, a more accurate picture of the linguistic situation may be provided. Dialect continuum describes a situation wherein neighbouring varieties of a linguistic group have minor differences with each other -- too minor to significantly inhibit mutual intelligibility -- while the same types of differences become greater in number and more substantial for mutual intelligibility as one moves further along the continuum. Speakers at one end of the continuum would have no difficulty communicating with their immediate neighbours, who in turn would have no difficulty communicating with *their* immediate neighbours, and so on all the way along the region, while those at one end of the continuum would have very little mutual intelligibility with those at the other end of the continuum. This situation arises when a language variety spreads over a relatively large area in which speakers' contact is generally limited to their nearest neighbours. Over time, these intermediate varieties may disappear, or individual varieties may be given official status as the standard, either of which case may break the continuum. Thus by looking at Tangsa-Nocte as such a continuum rather than as discrete branches within the central Sal / Northern Naga family tree, we are better able to address the relationships between varieties and understand how and when different aspects of the current linguistic and geographic situation as come to be. Doing so may allow us to explain the distribution of certain features which do not otherwise follow along more traditional subgrouping lines.

Finally, it should be made clear that in none of the following statements of linguistic relatedness is any claim being made regarding any group, subtribe or larger tribe's legitimacy or illegitimacy, similarity of difference in identity, or any other extralinguistic factors. The following sections are meant to speak only to the question of linguistic similarities, and the authors take no position on the complex politics of the region nor associations of the speakers of the included varieties, nor make any challenge thereof.

1.1 Methodology

Data were collected from a wide range of Tangsa-Nocte varieties. From these data, patterns were determined between those varieties which are uniformly classified as Tangsa — varieties such as Muishaung, Mungre and Cholim — and those classified uniformly as Nocte, such as that of Namsang. These included differences in phonology, lexicon and syntactical patterns.

The majority of data come primarily from wordlists and unpublished sketch grammars collected by Dr. Stephen Morey and his graduate students[4], data gathered by the authors of this paper for ongoing phylogenetic work, as well as data gathered on subsequent trips to the language communities for descriptive purposes. In addition to trips made to Tangsa-Nocte communities in subsequent years, this work has been supplemented by online communications, particularly during the ongoing global pandemic.

[4] Notably by Stephen Morey, Mulder (2020), Goswami (2018), as well as ongoing work by Vong Tsuh Shi and others cited below.

63

The following language varieties have been considered for this current study, organised according to their classification in van Dam (2018):

Nocte: Lowang, Namsang, Khasik
Tangsa: Bote, Champang, Cholim, Gaqlun, Hahcheng, Haidley, Jiingi, Joglei, Khalak, Chamchang, Lochhang, Longri, Lungkhi, Maitai, Muishaung, Muklom, Mungre, Ngaimong, Pingkhu, Rangsi, Rëra, Rinkhu, Shangthi, Shecyü, Tikhak, Yongkuk
Tutsa: Dewin, Hanphuk
Ollo: Laju dialect
Cross-classified: Phong, Hakhun

The classification in van Dam (2018) was adapted from Post & Burling (2017). Adaptations were based on widely-held community member impressions of similarity between varieties. Morey (2017:353) provides similar impressionistic groupings given by two such community members.

1.2 Structure of this paper

Starting with Section 2 below, we will present three varieties which have been given conflicting classification between sources or due to sociopolitical factors. In section 3, we will offer a list of features which, in lieu of an as-yet-unfinished phylogeny of Tangsa-Nocte, will provide insight into relatedness of varieties and may help to illustrate potential areas of further study in terms of contact scenarios and historical migrations.

2 Previous classifications of Tangsa, Nocte, and related varieties

There has been little argument in recent decades regarding the classification of Tangsa and Nocte varieties within the larger Sino-Tibetan macro family, nor since Burling (1983) has there been much argument against classification of Tangsa and Nocte, along with Wancho and Konyak, within a single group. While the label of this group has changed from author to author (Northern Naga, Eastern Naga, Konyakian), there is little doubt that Tangsa and Nocte are rightly placed.

Less clear are the discussions of how these various groups, to include Wancho, Konyak, Chang and Phom, related to each other within this group. Marrison (1967) split Northern Naga into two branches, with Tangsa, Nocte and Wancho sharing one branch and Konyak, Chang and Phom on another. This division was made based on phonological grounds, with Marrisons "Tangsa" group languages employing an /r/ onset and in some cases coda /l/ (Marrison 1967:75). French (1983) left Tangsa and Nocte on their own branch. In both cases, the available data on Tangsa and Nocte was largely limited to the Muishvung (Mossang) and Joglei varieties of Tangsa and the Nocte variety spoken around Namsang.

While the literature on Tangsa-Nocte varieties has grown considerably in the past years, with the publication of numerous grammatical descriptions and papers on individual subjects for member languages, with some exceptions (Khan Lann 2017, Morey 2019b), little has been done to resolve questions of internal classification. Descriptions of languages which may be called

transitional, or which are often classified both as Tangsa and as Nocte depending on source, often appear in these recent publications still in this state of limbo. This is not in any way to fault the authors of such works, since it may be outside the scope of a grammar written with heavy time constraints. In some of the more recent literature, such as Boro (2017) discussed in the following section, considerable effort is made to clear up this very matter for the variety being discussed. As it related to the current topic, it is worth discussing some of these varieties in more detail, if only to paint a clearer picture to the reader of the current state of uncertainty in internal classification.

2.1 Hakhun

Boro (2017) provides an excellent grammar of Hakhun, one such variety which is referred to in earlier works as both Tangsa and as Nocte. Quoting from Boro:

> *Hakhun is most similar to the speech variety of another ethnic community known as Nocte. ... Hakhun people maintain that their speech varieties are very similar, and some go a step further and maintain that they are 'essentially the same' language. It seems people are referring to the number of shared vocabulary items in such cases. These two varieties certainly have considerably different grammar, and it becomes obvious when looking at descriptions like Rahman (2016) on Nocte.* (Boro 2017:3-4)

The author of the grammar offers multiple paragraphs throughout the text giving comparisons to Lowang Nocte in order to show the differences. Indeed, many of the similarities between the varieties, such as the use of /t-ʌk/ and /t-ɤʔ/ as first-person singular past-tense markers in Nocte and Hakhun respectively, are found throughout Tangsa-Nocte. Muishvung — a Pangwa Tangsa variety which is not particularly close to Nocte — has this same marker in the form of /t-eɤʔ/. Indeed, Morey (2019b:143) provides an overview of agreement markers for Pangwa Tangsa varieties, for which his Group 1 varieties (Ngaimong, Joglei, Muishvung, Mungre and Maitai) also follow this same pattern of stop finals on the first-person past marker, not found elsewhere in Pangwa.

2.2 Phong

Another example of cross-classification between Tangsa and Nocte is the variety known among its speakers as Phong, and to outsiders often as Pontai. Phong is officially classified by the state government entirely based on the district in which the speakers reside (Dutta 2019). In Changlang District, they are classified as Tangsa, while Phong villages located in neighbouring Tirap district are classified as Nocte. There are aproximately 3000 Phong speakers spread among 10 villages (ibid) speaking a single common variety, for whoch self-reported differences are described as something like a slight accent. As with Hakhun, there is some linguistc affinity to more prototypical Nocte varieties such as Haʔva, to be discussed in a later section below.

Phong is one of the few cases where varieties which are nearly identical — and in some cases actually identical beyond slight prosodic differences according to speakers' reports — are classified officially as entirely distinct from one another. More common is the case of Hakhun, where classification is more a matter of uncertainty due to a dearth of documented linguistic data.

65

Data gathered for the current study was conducted by the authors in 2015, with additional data provided by Dutta (p.c.). A grammatical description of Phong is in preparation (Dutta, n.d.).

2.3 Lungkhi

A different but similar case arises even when the question of "Tangsa or Nocte?" is not to be found. Within the traditionally-labeled Tangsa varieties, a number of subgroups exist. Many of these have salience among community members rather than being externally applied branches given by visiting linguists. Two major groups within Tangsa are the Rangpang (or Pangwa[5]) and Heimi (van Dam 2018:66f; Morey 2019b:136). Other speech communities, such as Chamkok and Champhang, form additional smaller groups within Tangsa-Nocte. Rangpang speakers, in addition to having a high degree of linguistic similarity, also share certain cultural traits not generally found outside of Rangpang communities, with a few exceptions having resulted from prolonged contact with non-Rangpang groups. These labels — Rangpang and Heimi — are often self-classifications. They are generally consistent with linguistic similarities, but often include cultural and geographical affinities as well.

These classifications are not concrete, and can shift over time. One Tangsa variety, Lungkhi, is classified both as Rangpang and as Heimi, depending on the author. Morey (2019a) provides an account of agreement markers and other features of Lungkhi showing a strong similarity to other Rangpang varieties. In Dewar (1931), Lungkhi is clearly given as a Heimi variety. There are multiple possible explanations for this situation. It may be that the Rangpang/Heimi distinction as it is understood today did not follow the same grouping in 1931. The terms may also have some overlap in referents. It may also be that Rangpang was more salient as a cultural marker based on shared song forms and other cultural characteristics, and therefore did not closely follow the linguistic similaries as it does today. Indeed, Chamkok or Champhang, two closely-related non-Rangpang varieties and ones which differs considerably from most other Tangsa-Nocte varieties, should not under any linguistic grounds be considered Rangpang, but according to our consultants, the speakers of Chamkok have adopted some of the cultural markers of Rangpang communities after long periods of close contact and cohabitation with Rangpang speech communities. And, of course, it is possible that Dewar simply made a mistake.

3 A diagnostic feature list

At the time of writing, a project is underway to produce a phylogenetic tree of Tangsa-Nocte based on cognacy of a considerable wordlist covering a large sample of language varieties.[6]

[5] Thank you to an anonymous reviewer has pointed out that our use of Rangpang here is not entirely consistent with the literature, as both Morey (2019b) and Khan Lann (2017) use the name Pangwa. Our use of Rangpang here is based on discussion with community members who have reporeted that they themselves use Rangpang when refering to the group, and have asked us to do the same. In every sense, however, the two terms are synonymous.

[6] Currently under way at the University of Zürich Department of Language Science, the project is working to develop a detailed phylogenetic tree of all attested Sal languages using previously published linguistic data as well as data from a number of varieties being newly collected for this purpose.

Until this work is complete, there is value in the preliminary investigation undertaken for the sake of this paper. This section provides the assumptions from which the work began, and offers up a sort of diagnostic checklist. Diagnostic, because it may be used to help place otherwise undescribed varieties within the larger Tangsa-Nocte-Tutsa-Ollo family. Thus, rather than being primarily for the sake of placement on a tree, the checklist provides an illustration of the nature of Tangsa-Nocte varieties as existing on a continuum, and not simply as two branches of a higher order grouping.

As a starting point, it is clear that there is a close genetic relationship between the so-called Tangsa and Nocte varieties. It is also clear that in many cases, such as the varieties discussed in Section 2 above, that the distinction between what is Tangsa and what is Nocte is not always apprent or readily agreed upon. While in this and previous publications on Tangsa-Note, I (van Dam) have taken the firm approach of referring to the singular group as "Tangsa-Nocte", as has been seen elsewhere (as in Morey 2019b), when commencing research on this group in 2015, based on earlier literature, it was my understanding that such a clear distinction likely did exist. This was based soley on the way in which Tangsa and Nocte had been treated in earlier literature as two distinct entities. It was never in question, however, that the two had some close genetic affinity despite considerable cultural differences. A number of common etyma are readily apparent even to new arrivals to the subject, and many of the sound correspondences also retain a consistency that seems obvious at first glance.
The main features with which the varieties have been analysed is as follows. The first set are differences primarily in etyma.

1. The most basic etymon used for GRASS, *#nam* vs *#ʦiŋ* vs *#mwat*
2. The etymon for TREE, *#baŋ* vs *#pul*
3. The etymon used for ELEPHANT, *#pok* vs *#ʧaŋ*
4. The etymon used for WHITE, *#pwoŋ* vs *#loŋ*
5. The etymon used for BLACK, *#ŋjak* vs *#mak*
6. The etymon used for DRINK, *#ʤok* vs *#njaŋ* ⨳ *#ljaŋ*[7]

In some cases these are likely semantic shifts. Based on modern varieties in which multiple stems are found with differences in meaning, it is likely that GRASS *#ʦiŋ* was a more general term for grasses, and *#nam* referred specifically to wild jungle grasses, and sometimes the jungle more generally. This is the case in varieties such as Cholim and Joglei. In Muklom, both forms are also found, however *#ʦiŋ* referring to weeds and other small plants and *#nam* is “grass”.

Another example is ELEPHANT, where the *#ʧaŋ* stem is found widely in Tibeto-Burman and reconstructed in French (1983) as **Ɂ-gla.ŋ*, while *#pok* appears to be an innovative stem.

Further historical reconstruction and phylogenetic work is needed to conclusively determine the origins of some of the other terms which differ so frequently. It is speculated that DRINK *#ʤok*

[7] The ⨳ symbol here indicates two stem forms which reconstruct for a single concept, with distinct but related phonological shapes. Reconstructions for a handful of etyma indicate frequent alternations, often between /n/ and /l/, between nasal onsets, or between obstruent and nasal codas, the latter often found today as suppletive forms of verb stems.

67

may derive from a stem originally meaning to suckle, but at the time of writing this is not yet clear.

Additional features are based on morphological structure of some words:

7. Presence of absence of productive plural pronoun markers, e.g. from THREE *#dim* attached to a singular stem to make it plural, rather than distinct stems for plural and singular pronouns.
   The proto-Tangsa-Nocte pronominal system was likely one which distinguished singular and non-singular stems, the latter also distinct for representing clusivity, in which the unmarked non-singular stem may have been interpreted more strongly as a dual form, as described for Jinghpaw by Pujitski (1968), or as plural. Explicit marking was then used to specify, most often with numerals TWO or THREE, but sometimes with terms meaning "group" as in modern Muishvung.
   Thus, in some varieties such as Kotlum, the inclusive first-person non-singular stem *#naŋ* stands alone to indicate first-person plural, while in others such as Champhang, the plural-marking suffix — in this case THREE — is required and considered a part of a single polysyllabic lexeme.

8. The form of the term for SEVEN. A handful of related but distinct terms are attested, involving either *#nit* ⨯ *#mit* or *#ʃi*. The term in a given variety may be a combination of the two, in either order, or only one of the two either on its own or with a general numeral prefix — often schwa or /a/.
   In varieties such as Ngaimong and Muishviung, SEVEN is *#mit.ʃi*. In Muklom and Phong, it is *#ʃi.nit*. Lunkhai and the Lowang Nocte variety has simply #ə.nit, as do varieties such as Rangsi and Shangthi.

Finally, phonological changes are taken into account. These are generally less reliable in many cases, since phonological innovations may arise independently in similar contexts in otherwise distantly related varieties. However, the following show consistency for specific etyma.

9. Stems sharing the same initial as HAND and MACHETE begin with /d/ rather than /j~ʒ/
10. Stems sharing the same initial as GRASS and RAW begin with /ʃ/ or /h/ rather than /t^h/
11. Stems sharing the same initial as GOOD and WASH begin with /s/ rather than /x~h/

In reconstructing the proto-form of Tangsa-Nocte, a large degree of consistency is found to justify such an approach.

Morphosyntacic criteria must also be considered, such as agreement markers. In Morey (2019a), three features of Rangpang (Pangwa) grouping are given. Quoting from Morey (2019a:142)

1. *The presence of verb agreement,*
2. *The lack of a comprehensive hierarchical marking and inverse marking, and*

3. *A mix of preverbal and postverbal elements to mark the Tense/Aspect/Modality and Polarity categories in some varieties.*[8]

Of these three, the category of verb agreement is held in common with Nocte languages, with the Hakhun group …, with the Muklom-Havi group, with the Tikhak group, and with Phong. However the second feature, lack of comprehensive hierarchical marking, can distinguish Pangwa from each of these other groups except the Tikhak which also lacks hierarchical marking. Finally, the presence of some preverbal elements in the marking of some Tense/Aspect categories, such as future and negative in at least some varieties, sets the many of the Pangwa group aside from the other Tangsa-Nocte languages.

It should be noted that morphosyntactic considerations have been largely ignored in this paper up to this point. The topic of grouping based on agreement markers has been addressed in detail in Morey 2019a. Furthermore, at the time of writing, morphosyntactical features are not yet adequately incorporated into the larger phylogenetic study upon which this work is based. Discussions with Dr. Morey have taken place in order to determine how best to address this, and we hope to address the issue as work continues.

In addition to the features mentioned above, a number of changes are also to be found which do not otherwise assist in any effort at subgrouping or determining relatedness. As an example, the onset of words including tiger, firewood, eight and bear, reconstructed as *#ʧ*, has become /x/ in Cholim but not in closely related varieties such as Longri, and has become /h/ in Pinkhu and Rasa, neither of which are considered particularly close to Cholim on any other grounds. Likewise, other terms with onset *#ʧ* such as gold, bed and elephant *#ʧaŋ* have not undergone this change in these three varieties, either pointing to an incomplete lexical diffusion coincidentally found in only those three, or some onset which differed in some way from *#ʧ* but which merged at some earlier state. Such complications in reconstruction notwithstanding, the correspondences described in points 7–9 above show much more considerable consistence both between varieties and internal to varieties.

Table 1 shows a sampling of these features[9] for a selection of language varieties across the Tangsa-Nocte group, chosen to represent a range of possibilities and for having considerably more data available than some other groups. It would have been good to add Chamkok, for example, but at the present time there is insufficient data. A plus (+) indicates the presence of the feature, a (–) indicates the absence, such as in Muishvung where *#pok* is not the stem used for ELEPHANT and instead *#ʧaŋ* is used. A combination of plus and minus (±) indicates both stems are used, and a null sign (∅) indicates an entirely different stem is used, such as Tikhak using a different stem, *mɤ˧k^{h}ɯ˥*, derived from proto-Tibeto-Burman **m-gwi*, also found in other less-closely related varieties such as Konyak and the Jinghpaw languages. In Table 1, "agr." refers to the presence of agreement marking in general, while "hier." refers to hierarchical or inverse

[8] Also absent from the table due to visualisation constraints

[9] The stymon for GRASS has not been included here due to insufficient data for the included varieties. The general trend appears to be that #tsiŋ is most frequent among Rangpang varieties, #mwat < PTB *m(r)(u/a)k is more common in southern varieties spoken in Myanmar. #nam frequently occurs in southern varieties spoken in India and as part of compounds in Rangpang.

69

marking on the verb, both features suggested previously in Morey (2019a:142) as significant in identifying varieties as belonging to the Pangwa subgroup.

		Haʔwa	Hakhun	Phong	Tutsa	Muklom	Gaji	Tikhak	Lungkhi	Muishvung	Cholim	Khasik
ELEPHANT	*#puk*	+	+	+	+	+	+	Ø	–	–	–	+
TREE	*#baŋ* ^{i}	+	+	+	±	+	–	–	–	–	–	–
WHITE	*#poŋ*	+	+	+	+	+	–	–	–	–	–	Ø
BLACK	*#mak*	–	–	+	+	+	Ø	–	Ø	–	–	–
DRINK	*#ʤok*	+	+	+	+	–	–	–	–	–	–	
HAND	*d-*	+	+	+	Ø	–		–		–	–	–
GRASS	*h-*	+	+	–	–	–	Ø	–	Ø	–	–	±
WASH	*s-*	+	+	–	–	–	–	–	Ø	–	–	–
SEVEN	not *ɦi*	+	+	–	–	–	Ø	–	Ø	–	–	+
agr.		+	+	+	+	+		+	+	+	+	
hier.		+	+	+	+	+		–	–	–	–	

Table 1: Presence (+) or absence (-) of given features in a sampling of varieties.

While this table provides only a brief glimpse at the larger collection of data, the pattern is apparent: On one end are varieties such as Cholim and Muishvung, both belonging to the Rangpang Tangsa category mentioned in Section 2, which may be taken to be prototypically "Tangsa"-like varieties. On the other end are varieties such as Haʔwa, a Nocte variety with considerable documentation spoken in and around Borduria on the border of Assam and Arunachal Pradesh. The majority of other varieties, however, do not stick too close to these prototypes.

Tikhak is of some particular interest, as tradition holds that they were among the very first Tangsa communities to migrate from Myanmar into India. As a result of this and due to the locations in which they settled, there is a long history of interaction between Tikhak speakers and members of the Singpho (Jinghpaw-Kachin) communities in the region, resulting in the adoption of many Singpho customs by Tikhak speakers. Tikhak is not, however, a Rangpang variety, and in terms of other linguistic factors is quite different from Muishvung and does not share any considerable degree of mutual intelligibility.

Tutsa is also of some interest, due to having been classified by the government of Arunachal Pradesh as a Tangsa variety prior to the state census of 2011, at which time the varetiy was given recognition as separate from Tangsa. Based on the limited factors shown in the table above, it shows more similarity to Nocte and semi-Nocte varieties such as Phong.

Muklom is squarely in the middle, although as mentioned previously, the strength of the phonemic realisations of onsets for HAND, GRASS and WASH are perhaps not to be taken to be more valuable than they may be worth.

Khasik also requires some discussion, left on its own to the right-hand side of the table. Known more genrally as Khapa, Khasik is a Nocte variety which is considered by speakers both of Khasik and other Nocte varieties to be considerably different. It also serves as the language of traditional song for other culturally-Nocte groups in the region. While showing some features in common with other varieties generally classified as Nocte, it notably lacks the /s/ onset for WASH or the /h/ onset for grass, although for the latter some variation occurs. Considerable differences between Khasik and other Nocte varieties may be due to some difference in origin, possibly placing it linguistically closer to Konyak and Wancho than Tangsa, despite the cultural significance held by Khasik as the poetic register among Nocte speakers (van Dam & Das, forthcoming).

For Lungkhi and Gaji, both sampled from communities in Myanmar, the terms given for grass, wash and seven all came from different stems than the other varieties shown in the table, with GRASS from *#mwat*, WASH as /mi/ or /lu/ compared to *#hwal* as found in other varieties, and SEVEN as *#ʃaŋ*. These match other languages also spoken in the area such as Khalak and Ringkhu, pointing to the variance as a possible areal feature. Khalak is believed to be linguistically quite similar to Lungkhi (Morey 2019), complicating matters further.

Finally, it should be pointed out that in the case of a distinction between BLACK and DARK with *#mak* and *#ɲak*, it is not something which can be said to be a prototypical Nocte-like featureprototypical Nocte-like feature, but rather one which appears to cluster around these transitional varieties, spoken in the geographic area between Rangpang and Nocte varieties. The source of this is as yet uncertain, but this distinction is not commonly found either in Rangpang varieties or in the more traditionally classified Nocte varieties. It may be an innovation, however both forms, *#mak* and *#ɲak*, are reconstructed as allofams of a common stem for Tibeto-Burman (Matisoff 2003). It may also be that this distinction in form and meaning does exist in some of these languages, but was not attested in the sources consulted for the present study.

4 Conclusion

Further work remains to be done. This paper serves as only a first step in a much larger attempt to both classify and reconstruct an ancestral form for Tangsa-Nocte varieties. However, from the work that has been done and presented here, the following can be stated with some high degree of certainty.

First, Tangsa and Nocte as linguistic groups do not have such clearly defined borders either in terms of features or in terms of member varieties that they should be considered two distinct branches within a higher-order Northern Naga grouping. Instead, based on internal diversity within the so-called Tangsa and Nocte varieties, they form a continuum of features which may show tendencies to be more Tangsa-like or more Nocte-like. In some cases, the differentiation of a variety based on these features may be complete, as is seen between Haʔwa and Cholim at least for the features given above. However, most varieties will fall somewhere in between these two extremes. An expanded set of features may show either more grouping at the ends, or continued distribution along a gradient.

Second, in addition to what appears to be extremes on a continuum, additional clusters appear, such as with the case of terms for WASH and SEVEN among some varieties spoken in Myanmar, either as borrowings from another non-Tangsa-Nocte variety, or as innovations what

71

have spread in the immediate vicinity. Due to the difficulties of conducting fieldwork in Myanmar at the present time, this too must wait until a later date. However, based on a cursory investigation into the data available on the Sino-Tibetan Etymological Dictionary (STEDT), no clear source of the Tangsa-Nocte form for SEVEN *#ʃaŋ*, although WASH *#lu* has clear counterparts in Karbi and Central-Southern Naga, and grass *#mwak* is well-attested from proto-Tibeto-Burman **m(r)(u/a)k* (STEDT) and found throughout Burmic varieties including Lisu, Lahu and Burmese.

4.1 Future and ongoing work

As documentation improves, both in terms of ongoing work by scholars working in the area but also an increase in the number of community members contributing their own knowledge to the literature, further progress can be made in answering some of these questions. In the meantime, ongoing research with the available data will continue.

At the time of writing, it can at least be stated with a high degree of certainty that Tangsa-Nocte, or rather Tangsa-Nocte-Tutsa-Ollo, can be reasonably treated as a single entity within Northern Naga can be treated as such, and that continued analysis will shortly reveal the degree to which these varieties can be shown to be related to Wancho, Konyak and other neighbourhing members of the central Sal family.

5 APPENDIX

Table 2 shows the original data behind Table 1. Much of the data comes from published sources – cited above – which has been supplemented by additional fieldwork and discussion with speakers.

		Haʔwa	Hakhun	Phong	Tutsa	Muklom	Gaji	Tikhak	Lungkhi	Muishvung	Cholim	Khasik
ELEPHANT	*#puk*	+ pok	+ poʔ	+ pok	+ pɔk	+ puːk	+ boʔ	Ø	–	–	–	+ puk
	#ʧaŋ								$ʧ^{h}$a	ʧɔ	tsa	
	#m.gui			mukuj				mukui				
TREE	*#baŋ* [i]	+ baŋ	+ bɤ	+ vaŋ	± baŋ	+ bʌŋ	–	–	–	–	–	–
	#pul				pu		p^{h}un	pu	pi	pul	pi	
WHITE	*#poŋ*	+ pu	+ po	+ poŋ	+ puŋ	+ puŋ	–	–	–	–	–	Ø $tʃ^{h}$u
	#loŋ						lo	luŋ	luʔ	lɯ	ljɔ	
BLACK	*#mak*	–	–	+ mək	+ mak	+ mʌk	Ø	–	Ø	–	–	–
	#njak	ɲak	ɲak					ɲak		ɲɐuk	ɲak	njak
	#kha						k^{h}a		k^{h}a			
DRINK	*#ʤok*	+	+	+	+	–	–	–	–	–	–	
HAND	*d-*	+ /d/	+ /d/	+ /d/	Ø /l/	– /ʝ/	– /j/	– /ʤ/	– /j/	– /ʒ/	– /z/	– /ʒ/

GRASS	*h-*	+ /h/	+ /h/	– /ʧ/	– /tʰ/	– /tʰ/	Ø	– /ʧ/	Ø	– /tʰ/	– /ʧʰ/	±
WASH	*s-*	+ /s/	+ /s/	– /ʧ/	– /ʧ/	– /x/	– /h/	–	Ø	– /h/	– /h/	–
SEVEN	not *fi* *mt-shi* *[ə]nit*	+ [i]ŋit	+ ŋat	–	–	–	Ø mət.ʃi	–	Ø [tə] nit	– mə.ʃi	– mit.ʃi	+
	shi-nit			si.net	si.ne t	sʌ.nʌt		sə.nat				

Table 2: Expansion of Table 1 with phonological data.

REFERENCES

Boro, K., 2017. A grammar of Hakhun Tangsa. (Doctoral dissertation, University of Oregon).

Burling, R. 1983. "The Sal languages." *Linguistics of the Tibeto-Burman Area* 7 (2):1–32.

Dewar, T.P., 1931. Naga tribes and their customs. *Census of India 1931*, *11*, pp.267-295.

Dutta, N., 2019. Verb agreement in Phong. *Himalayan Linguistics*, *18*(1).

French, W. T., 1983. Nothern Naga: A Tibeto-Burman. City University of New York.

Goswami, D., 2018. Preliminary phonology of Rera, a Tangsa variety of Northeast India. *Journal of the Southeast Asian Linguistics Society*, *11*(2), pp.xlv-lxxv.

Lann, K., 2017. *A phonological comparison of Shecyü and Mungre and its contribution to a common Tangshang Naga orthography* (Doctoral dissertation, MA dissertation, Department of Linguistics, Payap University, Chiang Mai).

Marrison, G.E., 1967. *The classification of the Naga languages of north-east India* (Doctoral dissertation, School of Oriental and African Studies. University of London.

Matisoff, J. A. 2003. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.

Morey, S., 2017. Tangsa. Sino-Tibetan Languages, ed. by Graham Thurgood and Randy LaPolla. London and New York: Routledge.

Morey, S., 2019a. Pangwa Tangsa agreement markers and verbal operators. *Himalayan Linguistics*, *18*(1).

—, 2019b. The Tangsa-Nocte languages: An introduction. *Himalayan Linguistics*, *18*(1).

Mulder, M., 2020. *A descriptive grammar of Muklom Tangsa* (Doctoral dissertation, La Trobe).

Norman, J., 1988. *Chinese*. Cambridge University Press.

Post, M.W. and Burling, R., 2017. The Tibeto-Burman languages of Northeast India. *The Sino-Tibetan Languages*, pp.213-242.

Rahman, S. I. 2016. Verbal Agreement and Grammatical Description of Haʔva Nocte. MA Thesis, Payap University.

Sagart, L., Jacques, G., Lai, Y., Ryder, R.J., Thouzeau, V., Greenhill, S.J. and List, J.M., 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences*, *116*(21), pp.10317-10322.

73

Statezni, N. 2013. *Fifty-five dialects and growing: Literacy and comprehension of vernacular literature among the Tangshang Naga in Myanmar*. MA dissertation, Graduate Institute of Applied Linguistics, Dallas.

STEDT Database. *The Sino-Tibetan Etymological Dictionary and Thesaurus*. http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl (last accessed on 2021-06-07)

Van Dam, K. P., 2018a. The tone system of Tangsa-Nocte and related Northern Naga varieties. La Trobe University

—, 2018b. A cross-varietal description of modifiers of basic colour terms in Tangsa-Nocte. *North East Indian Linguistics8*, pp.47-64.

—, 2020. The syntax of intensifiers in Muishaung. *Linguistics of the Tibeto-Burman Area*, *43*(1), pp.19-36.

Van Dam, K. P. & B. Das. 2021. *Reduplication of intensifiers in Ollo (Laju) and Kasik (Khapa Nocte)*. Presented at the HLS Workshop on ideophones and interjections, September 2021. Paris.

Zhang, M., Yan, S., Pan, W. and Jin, L., 2019. Phylogenetic evidence for Sino-Tibetan in northern China in the Late Neolithic. *Nature*, *569*(7754), pp.112-115.