

Overcounting Numeral Systems and Their Relevance to Sub-grouping in the Tibeto-Burman Languages of Nagaland*

Alexander R. Coupe
Nanyang Technological University

This paper presents an appraisal of the most influential genetic classifications that have been proposed for the Tibeto-Burman languages of Nagaland, north-east India, and weighs up the evidence for the validity of a ‘Naga’ branch within Tibeto-Burman. The survey concludes that while phonological and lexical correspondences might be generally useful for establishing the affiliation of these languages to the Tibeto-Burman family, such criteria shed limited light on the problem of establishing lower level sub-groupings. Apart from the Konyak languages, which have been convincingly grouped with Jinghpaw and the Bodo-Garo languages on the basis of lexical innovations, sub-grouping within the languages of Nagaland remains inconclusive for want of robust diagnostic criteria. The paper presents new evidence for an intermediate grouping of languages of the central and southern regions of Nagaland (i.e. the Ao and Angami-Pochuri clusters), the historical-comparative basis for this being typologically rare overcounting numeral systems documented in the languages of these two regions in the late 19th and early 20th centuries. The complexity of overcounting numeral systems lessens the likelihood of their being borrowed, and such patterns have never been reported in contiguous Konyak, Karbi, Zeme, Kuki or Tangkhul languages. The paper concludes that the erstwhile presence of overcounting patterns in Tibeto-Burman languages of central and southern Nagaland must therefore represent an inherited innovation indicative of a close genetic relationship.

Key words: Tibeto-Burman, overcounting, numeral, Naga, genetic classification

1. Introduction

Despite it being more than 150 years since linguistic research first began on the so-

* This paper was written while I was visiting at the Research Centre for Linguistic Typology, La Trobe University. An earlier version was presented at the International Symposium on Sino-Tibetan Comparative Studies in the 21st Century, hosted by the Institute of Linguistics at Academia Sinica, Taipei, in June 2010. I am grateful to the audience for their comments and to the organizers for the opportunity to present this research. The paper has further benefitted from the comments and suggestions of two anonymous reviewers. The usual caveats apply.

called ‘Naga’ languages of north-east India, to this day their genetic classification remains as confused and tortuous to negotiate as the jumbled mountain ranges in which they are spoken. Long-standing political tensions that began in the colonial era and have continued into the post-independence period have rendered much of north-east India inaccessible to outsiders until relatively recently, with the result that there have been limited opportunities to build upon the often patchy historical linguistic data hitherto available for comparative research. Given this historical backdrop, the sluggish advances in genetic classification are not so surprising. It is still the case that very few Tibeto-Burman languages of north-east India have been documented using modern methods of linguistic description, and all classifications—even the most recent—continue to be based almost exclusively on word lists that were collected by amateurs more than half a century ago.

It is relevant to begin this overview by questioning the use of the term ‘Naga’ to refer to a Tibeto-Burman linguistic grouping. One could be forgiven for assuming from the literature that ‘Naga’ is already a valid and well-attested sub-branch of Tibeto-Burman, such is its ubiquity in all existing classifications. Marrison (1967, I:12) presents evidence from a number of sources that points to Assamese *naga* ‘naked’ as being the most likely etymological source of ‘Naga’; by all historical accounts, the term was initially unknown to the peoples to whom it was applied (e.g. see Woodthorpe 1882:57, Soppit 1885:1, and Grierson 1903, 3.2:194). ‘Naga’ came to be used by the British to refer to all the tribal people inhabiting the mountains between the Manipur valley and the North Cachar Hills in the south and south-west, the Patkai range in the east, the Brahmaputra River valley to the west, and the Mishmi Hills to the north. In the latter half of the 20th century, the term assumed an ethno-nationalistic utility when it was adopted for self-reference by a loosely affiliated group of people who, although speaking disparate Tibeto-Burman languages, were nonetheless united in their common struggle against Indian rule. The languages of the people who nowadays identify themselves as Nagas may not necessarily be closely related to each other, but their speakers appreciate that the greater the numbers, the greater the leverage for articulating a community’s political demands in India. Because of this, some authors have rejected the validity of ‘Naga’ as a linguistic label altogether (e.g., Burling 2003:172).

This paper undertakes two tasks. The first is to present a synopsis of the most influential genetic affiliations that have been proposed for the ‘Naga’ languages, and to weigh up the cited evidence for these classifications in order to determine if a Naga grouping within Tibeto-Burman can truly be justified. This will serve as an appropriate point of departure for then reviewing the validity of higher level affiliations that seek to link ‘Naga’ languages with other proposed linguistic groupings of the north-east India region. The second task is to determine what features, if any, set the ‘Naga’ languages

apart from the established branches of Tibeto-Burman, and what evidence can be found for sub-grouping within a Naga branch. This will take into account recent research revealing typological characteristics that are almost certainly innovative and thus diagnostic of genetic sub-groupings.

The paper has the following structure. Section 2 provides an historical overview of the problems presented by the classification of Tibeto-Burman languages of the north-eastern border region of India, and provides outlines of sub-grouping proposals as they apply to languages classified as belonging to a putative Naga branch. Section 3 then offers my own contribution to the classification of these languages, which recognizes the value of obsolete overcounting cardinal numeral systems as historical evidence for sub-grouping at the lower levels of taxonomy. Finally, Section 4 discusses the implications of the findings and proposes arguments in support of a new Angami-Ao sub-grouping for languages spoken in central and southern Nagaland.

2. Historical overview of the classification of Naga languages

Four influential classifications as they pertain to the ‘Naga’ languages are considered and compared in this section: Grierson’s (1903-1928) *Linguistic Survey of India*, Shafer’s (1955) ‘Classification of the Sino-Tibetan languages,’ Benedict’s (1972) *Sino-Tibetan: A Conspectus*, and Burling’s (2003) classification titled ‘The Tibeto-Burman languages of Northeastern India,’ which incorporates an earlier (1983) classification of a sub-group named ‘Sal’, based on shared lexical innovations.

2.1 *Linguistic Survey of India* (1903-1928)

The first extensive attempt to classify the Tibeto-Burman languages spoken in the tract of land then known as the Naga Hills was undertaken in the *Linguistic Survey of India* (1903-1928, hereafter LSI) by Grierson & Konow. By Grierson’s own admission (1903, 3.2:194), the sub-groupings of the Naga languages presented in the LSI were essentially unchanged from those originally proposed by Damant (1880), which identified Western, Central and Eastern Naga sub-families on the basis of comparative vocabularies consisting of just thirty lexical items. From the discussion of the sub-grouping methodology, we may infer that Damant relied equally upon the non-linguistic considerations of locality, shared customs and manner of dress to determine affiliations between linguistic communities of the region.¹

¹ Some of Damant’s data is wildly inaccurate. The “Hatigorra” words, for example, bear little resemblance to any Ao variety I have ever encountered and are unlikely to be representative of the Ao language.

In the absence of compelling proof of genetic affiliations established by way of inflectional morphology or other structural characteristics, the LSI's justification for sub-grouping continued to depend as much on geographical proximity as upon any phonological or morphological evidence that could be garnered from the meagre sources of data. A proposed Naga group of languages under an Assam-Burmese branch was subsumed by Western, Central and Eastern sub-groups in agreement with Damant's original classification, with a Naga-Bodo and a Naga-Kuki sub-group being the only new additions. These two new sub-groups were established to account for a number of transitional languages demonstrating features of Bodo and Kuki/Chin languages, in addition to sharing lexical similarities with other languages of the Naga group (these linkages are represented by the broken lines of Figure 1 below). The languages of the Naga-Bodo and Naga-Kuki sub-groups are respectively spoken to the south-west and south of the Naga Hills. Their locations are likely to have put them in contact with Kuki and Bodo languages in the North Cachar Hills and Karbi Anglong districts of present-day Assam, and with Chin languages in the neighbouring state of Manipur, possibly resulting in the diffusion of linguistic features.²

Only a few structural innovations were identified within the LSI's Naga branch, one of these being the negative suffixes that characterize the Western sub-group and distinguish it from the Central sub-group, which uses negative prefixes on verb stems. The recognition of this particular morphological feature was probably in part responsible for Sumi being shifted out of Damant's Central Naga sub-family and into the LSI's Western sub-group. Apart from this, most of the morphological characteristics that were recognized—such as the ubiquitous distinct forms for negative imperatives, the derivation of deverbal adjectives and verbal nouns via nominalizing prefixes,³ the presence of morphological causatives, and an alienable~inalienable contrast in the marking of nominal possession—failed to distinguish discrete sub-groups, but nevertheless served to establish a genetic link between the confusing mass of languages spoken in the Naga Hills and surrounding areas, as well as their common Tibeto-Burman heritage. The significance of a shared overcounting cardinal numeral system occurring in Angami (Western sub-group) and Ao (Central sub-group) went unappreciated, although it was recognized and briefly compared (Grierson 1903, 3.2:266). Phonological differences between the Western and Central groups were more apparent, such as the absence or

² It is relevant to mention that the terms 'Kuki' and 'Chin' both refer to one and the same linguistic sub-grouping, so one term would in fact suffice. 'Kuki' refers to members of the sub-group spoken in India, and 'Chin' refers to members of the sub-group spoken in Burma (Myanmar).

³ This feature is not found in Chang (Konyak sub-group), which instead derives verbal nouns and adjectives via a nominalizing suffix *-puw*.⁵⁵

restricted number of coda consonants permitted in languages of the Western sub-group compared with the still limited but greater range of coda consonants permitted in the Central sub-group.

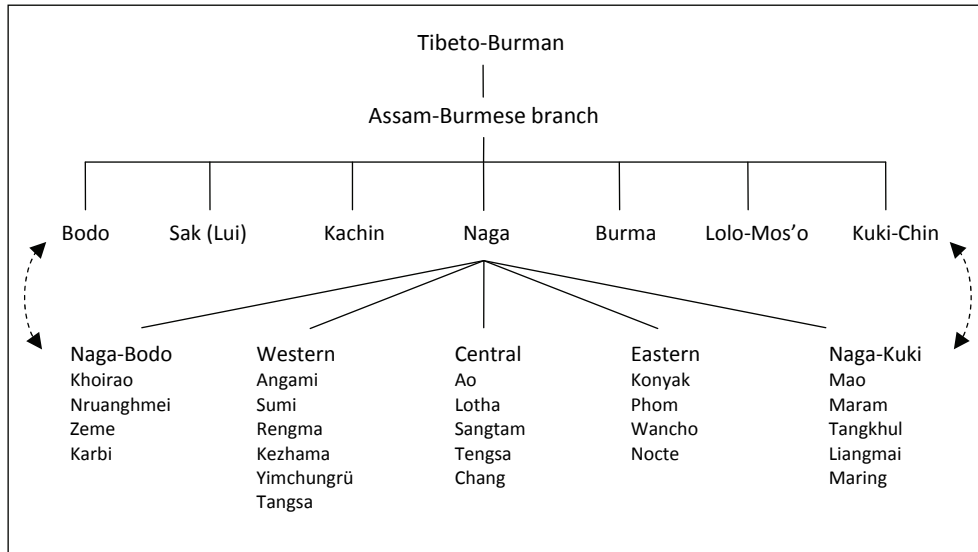


Figure 1: Classification of the Naga languages by LSI (Grierson & Konow 1903-1928). Historical names, toponyms and other inaccuracies have been replaced with current designations used for these languages.

The Eastern sub-group was considered transitional between the other Naga sub-groups and Singpho (Kachin), based on lexical affinities it was found to share with the latter, and probably also its geographical proximity to languages of the Kachin group in neighbouring Burma (Myanmar). A noteworthy morphological characteristic that set it apart from the other Naga sub-groups was a hierarchical system of cross-referencing found on the verbs of some Eastern Naga languages, which also distinguished those languages from the Bodo group of languages. This is not a feature of all members of the Eastern group and may be restricted to just those languages spoken in the Tirap and Changlang Districts of Arunachal Pradesh. If future research reveals that the forms of the agreement markers are cognate, then this may constitute an innovation and thus evidence for low-level sub-grouping within the Eastern languages. The alternative view is that pronominal indexing is evidence of a shared retention, in which case it provides only weak evidence for establishing a sub-group. Whether or not verbal agreement should be reconstructed for Proto-Tibeto-Burman remains a bone of considerable contention in Tibeto-Burman historical linguistics (e.g. see Bauman 1974 and DeLancey

1989 for arguments for, and LaPolla 1992 for arguments against) and a resolution of the opposing points of view is yet to be reached.

2.2 Classification of the Sino-Tibetan Languages (Shafer 1955)

The next major attempt at classification was made by Shafer (principally 1955), who relied upon lexical and phonological correspondences to identify six major divisions of Sino-Tibetan (see Figure 2 below). Although his classification included a Daic division, Shafer suspected that Daic languages were only very distantly related to Sino-Tibetan (1955:97). These languages have been demonstrated to constitute a separate (Tai-Kadai) family in subsequent research (Benedict 1972, Li 1977, Ostapirat 2000).

Each division of Sino-Tibetan was identified by an ending in *-ic*. These were subsumed by sections ending in *-ish*, and below the sections were branches that terminated in either units or just lists of language names. Shafer's Kukish section of the Burmic division linked the Kuki/Chin languages of the Indo-Burmese frontier to the LSI's Western and Central Naga sub-groups and the Naga-Kuki sub-group in a total of twelve branches. A major divergence from the LSI's classification was the recognition of an independent Baric division bifurcating into a Barish section (containing five branches) and a Nagish section (constituting a single branch). The Barish section included the LSI's Bodo group of languages predominantly spoken on the Plain of Assam and the transitional languages of the LSI's Naga-Bodo group, while the Nagish section isolated the LSI's eastern sub-group.

As with the classification of Tibeto-Burman languages by the LSI, Shafer's work is long on lists of cognate lexical items but short on a methodology that recognizes isoglosses, innovative sound changes, or other credible comparative evidence that might justify his profusion of sub-groupings. In 'The Naga branches of Kukish' (Shafer 1950) the largest grouping in his classification of Sino-Tibetan and the one he held to be the most important due to its retention of archaic forms, correspondence sets for initials and prefixes are not provided 'for lack of space' (1950:471). The reader is instead presented with the phonetic correspondences of rhymes from which the forms of Proto-Kukish have been reconstructed, but the data is not presented in such a way that would give justification to setting up the twelve branches of Kukish. In the same article (p.475) he confusingly refers to a 'Southern Naga Branch' that includes Lotha, Thukumi [=Sangtam], and Yatsumi [=Yimchungrü], yet there is no such branch in his Kukish section, as demonstrated by its absence in Figure 2 below.

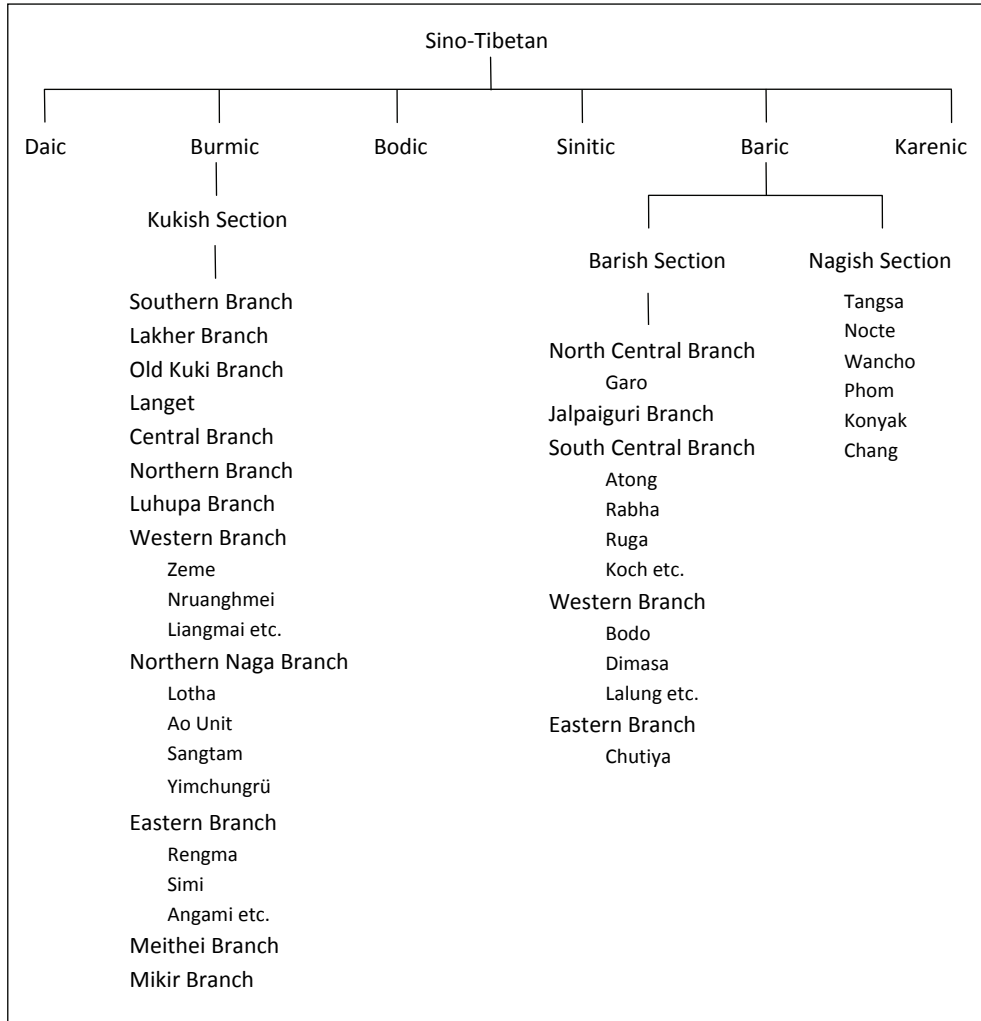


Figure 2: Classification of the Burmish and Baric divisions of Sino-Tibetan by Shafer (1955)

Shafer was critical of Konow's recognition of a Naga group in the LSI and claimed that '[t]he Naga languages are all Kukish except for the northeastern-most, which is Baric' (1955:104, fn.45). In the same annotation he provides a figure that demonstrates his interpretation of the interrelationships holding between his Kukish languages of the north-east. The first six branches listed under the Kukish Section in Figure 2 are considered to be closely related "core" languages of Kukish with little evidence of divergence, and consequently are held to be representative of Proto-Kukish. The second

six branches all share evidence of phonetic decay and are semantically and morphologically divergent. From this one could deduce that Shafer's sub-grouping in the Kukish Section was essentially based on his perception of how divergent any language was from what he considered to be core Kuki.

In summary, we might conclude that despite the copious compilations of phonological correspondences amassed by Shafer, his new classification was in all likelihood predominantly based on lexical correspondences and the geographical locations of languages, in common with the previous classifications by Damant (1880) and the LSI (1903-1928). And like its predecessors, these served to identify the genetic affiliation of the Tibeto-Burman languages of north-east India to the Sino-Tibetan super stock, but failed to establish convincing criteria for the sub-grouping of these languages within Tibeto-Burman.

2.3 *Sino-Tibetan: A Conspectus* (Benedict 1972)

The classification of Tibeto-Burman languages by Benedict (1972) in *Sino-Tibetan: A Conspectus* (hereafter STC) built on the work of Shafer and used the same linguistic sources. But rather than attempting to account for genetic affiliations within a traditional tree model, Benedict instead posited Kachin as a hub, both linguistically and geographically, to which were linked satellite nuclei representing seven major divisions of Tibeto-Burman, viz. Tibetan-Kanauri, Bahing-Vayu, Abor-Miri-Dafla, Kachin, Burmese-Lolo, Bodo-Garo and Kuki-Naga.

The classification was conservative in only positing nuclei that were confidently differentiated. For example, Konyak was not given the status of a division, but was nevertheless acknowledged as being profitably compared to Bodo-Garo. An analogous relationship was suspected to hold between the Kuki-Naga division and the Mikir, Meithei and Mru languages.

Shafer's vast Kukish section of his Burmic division was broken up into a core of Kuki languages subsumed by four lower level groupings: Central Kuki, Northern Kuki, Old Kuki and Southern Kuki. This revision was noted to correlate with the original LSI classification. Benedict was unable to find any clear evidence that would justify maintaining the distinction between Kuki and Naga, thus these two groups were united under a Kuki-Naga nucleus. Shafer's Northern Naga Branch of Kukish (corresponding precisely with the LSI's Central Naga sub-group) was retained as Northern Naga, and Shafer's Eastern Branch of Kukish (corresponding with the LSI's Western sub-group) was renamed Southern Naga.

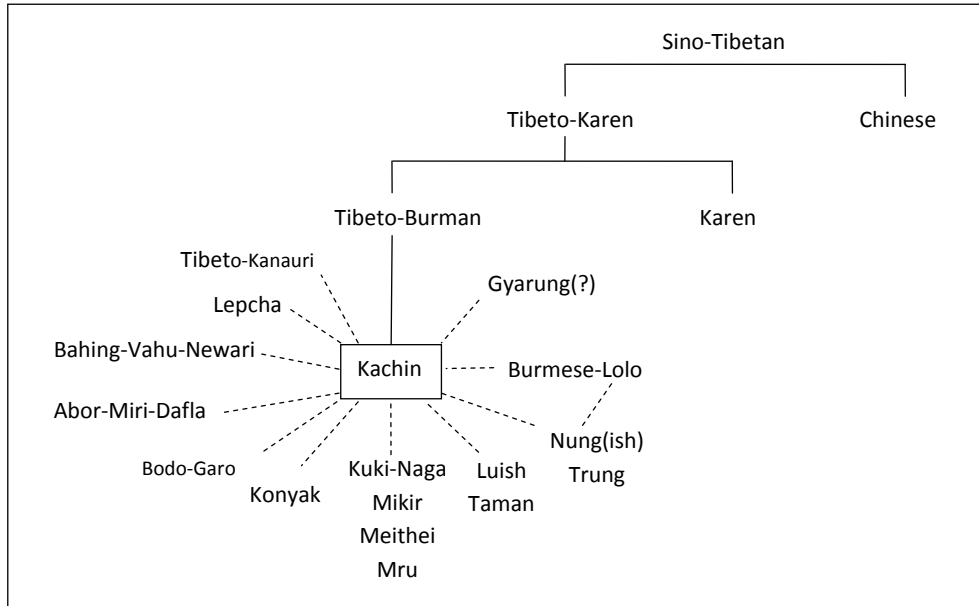


Figure 3: Schematic representation of Sino-Tibetan languages (Benedict 1972:6)

STC appears to have built upon an observation by Konow in the LSI (Grierson 1903, 1.1:68) that Kachin and the Eastern Naga languages (i.e. the Konyak group, or Shafer's Nagish section) share an affinity, as demonstrated by their distinctive roots for 'sun' and 'fire' (Benedict 1972:6-7). This was further investigated by Burling (1983), who proposed a new branch of Tibeto-Burman called the 'Sal' languages, based on shared lexical innovations between these two groups and the Bodo-Garo languages. We consider the evidence for this branch in the following sub-section.

2.4 The Sal languages (Burling 1983), and Classification of the North-eastern languages of India (Burling 2003)

Burling (1983) proposes to isolate a 'Sal' group from other Tibeto-Burman languages of north-east India, based on shared lexical innovations. The recognition of this new branch has generally been well-received, and data recently gathered from Konyak languages of the Indo-Burmese border area lend further support to the Sal hypothesis, although it also raises a number of questions concerning the use of lexical innovation as a sub-grouping tool.

Table 1 below partially reproduces a list of 'most likely Sal innovations' proposed in Burling (1983). These word lists have been augmented with hitherto unpublished Chang and Khamniungan data from my fieldwork notes, plus other work that has

appeared since Burling's (1983) article was published. A comparison of the columns of data reveals that of the forms provided in six Konyak languages — viz. Chang, Khamniungan, Konyak, Nocte, Phom and Wancho — only the correspondences of 'drink', 'face', 'forehead', 'finger', 'fire', 'foot/leg', 'mother' and 'sun' appear to constitute *bona fide* lexical innovations. The remaining lexical items turn out to be related phonologically to the Proto-Tibeto-Burman forms reconstructed by STC and subsequently by Matisoff (2003) in the *Handbook of Proto-Tibeto-Burman* (henceforth HBPTB), or else have other uncertainties associated with them. For example, according to Benedict (1972:19, fn.69), the reconstructed form **trap/drap* 'ashes' is an Austro-Tai loanword, while the forms given for 'burn' appear to be related not to PTB #330 **kaŋ*, which is in fact more accurately glossed as 'roast', but to **m-tsik* 'burn/angry', as reconstructed in HBPTB (Matisoff 2003:344-345), thus they cannot be considered innovative. Furthermore, if it is assumed that a lexical innovation occurring in the intermediate proto-language must necessarily be inherited by all the daughter languages for it to be convincingly diagnostic of a genetic relationship, then the existence of any word that demonstrates a phonological correspondence with a semantically equivalent reconstructed Proto-Tibeto-Burman form may cast suspicion over that whole set as constituting conclusive evidence for a sub-grouping. Such is found to be the case with the forms for 'dry', which is reconstructed as PTB **tan* by Matisoff (2003:258). The existence of the phonologically similar Phom form *ɜæn*³³ 'dry' challenges the premise that this could be an inherited lexical innovation. Of course, there is a chance that Phom did indeed inherit an innovated lexical morpheme from the intermediate proto-language and then borrowed or innovated a newer form that just happens to be phonologically reconstructable to PTB **tan*, but the evidence to prove this assumption is lacking.

Yet another concern is raised by the observation that some of these *a priori* innovated lexical items are not cognate across all the Konyak languages investigated, the forms for 'finger' and 'sun' being two such examples. Does the fact that a set of semantically equivalent morphemes are not phonologically similar refute the assumption that they could be genetically inherited innovations? Certainly the evidence for a genetic relationship and sub-grouping is much stronger if that set shares a phonological resemblance across all the languages, and also deviates phonologically from the reconstructed PTB form. It is also not possible to determine if any of the non-cognate morphemes in the syntagmatically arranged sets of data in Table 1 could have resulted historically from the replacement of a lexical item that was phonologically cognate with its corresponding reconstructed Proto-Tibeto-Burman form. If the latter were the case, then the number of *bona fide* lexical innovations and the evidence for the Sal sub-group might be diminished even further.

Table 1: Most likely Sal innovations, based on Burling 1983:19, Table 1a (ARC = my own data; GEM = Marrison 1967; B&P = Burling & Phom 1998; B&W = Burling & Wanchu 1998.) Transcriptions adjusted to accord with IPA conventions, apart from the reconstructions of Matisoff's (2003) HBPTB and Benedict's (1972) STC.

	HBPTB Matisoff 2003	STC Benedict 1972	Chang ARC	Khamtiungan ARC	Konyak GEM 1967	Nocte GEM 1967	Phom B&P 1998	Wancho B&W 1998
ashes	<i>*tal, *drap-trap</i>	<i>*tap</i> (18)	<i>*pla</i>	<i>low</i> ³³	<i>təpla</i>	<i>tap-la</i>	<i>ɔ¹¹ni²³³, dap³³æ³¹</i>	<i>βan⁴⁴di²⁴⁴, la⁴⁴bɔ⁴⁴</i>
burn	<i>*ka(ɕ)ŋ</i>	<i>*kaŋ</i> (330)	<i>tʃok¹¹, kaŋ⁵⁵</i> (roast)	<i>a¹¹-tsik⁵⁵</i>	<i>wok, aha</i>	<i>thak</i>	<i>tʃak³³</i>	<i>moŋ⁴⁴tsik⁵²</i>
cook	<i>*tsyow</i>	<i>*tsyow</i> (275)	<i>thuŋ¹¹</i>	<i>a³³-theŋ¹¹</i>	<i>dʒao, dʒaŋ</i>	<i>puon</i>	<i>tʃʌu³¹</i>	<i>phoŋ⁴⁴mok⁴⁴</i>
crow	<i>*ka-n</i>	<i>*ka</i>	<i>a⁵⁵khu⁵⁵mak¹¹</i>	<i>mə³³kho⁵⁵</i>	<i>aoʃa</i>	<i>vakha</i>	<i>a³³ʃæ⁵⁵</i>	<i>o⁴⁴kha⁴⁴</i>
drink	<i>*ʔam</i>	<i>*am = am</i> (481)	<i>joŋ¹¹</i>	<i>a¹¹-ʃu³¹</i>	<i>jiŋ</i>	<i>tʃok</i>	<i>juŋ⁵⁵</i>	<i>liŋ⁵²</i>
dry	<i>*tan</i>	<i>u⁵⁵</i>	<i>tha⁵⁵jek⁵⁵</i>	<i>a⁵⁵-waj¹¹</i>	<i>uwənpu, lipu</i>	<i>tʃhuoklam</i>	<i>ʒæŋ³³</i>	<i>lom⁵²e⁴⁴a⁴⁴zwa⁴⁴</i>
face	<i>*mazy, *mel</i>	<i>*mur</i> (366)	—	<i>kha¹¹</i>	<i>ʃita</i>	<i>than</i>	<i>the³¹</i>	<i>than⁵²</i>
forehead	<i>*d-pral</i>	—	—	—	<i>ʃakenj</i>	—	<i>ʃaŋ⁵⁵bæ³¹</i>	<i>khaŋ⁴⁴poŋ⁴⁴</i>
far	<i>*dzvaɪ, wəy</i>	<i>*dzvaɪ</i> (229)	<i>sa⁵⁵</i>	<i>a⁵⁵lay¹¹</i>	<i>dʒaipu</i>	<i>dʒamlap, halo</i>	<i>luanpi, taipü</i>	<i>tsaɪ⁴⁴lo⁴⁴</i>
father	<i>*pu, p^wa, pa</i>	<i>*pa = pwa</i> (24)	<i>pou⁵⁵</i>	<i>pow¹¹</i>	<i>apa</i>	<i>va, iva, apa</i>	<i>bæ³³</i>	<i>a⁵²pa⁴⁴</i>
finger	<i>*m-yuŋ</i>	<i>*m-yuŋ</i> (355)	<i>ik⁵⁵</i>	<i>pha⁵⁵kha¹¹</i>	<i>jaklan</i>	<i>daksu</i>	<i>lak³³tʃæŋ⁵⁵</i>	<i>tʃak⁵²khi⁴⁴</i>
fire	<i>*b^war, hwal</i>	<i>*bar, par</i> (220)	<i>wan⁵⁵</i>	<i>in¹¹</i>	<i>ka, a</i>	<i>van</i>	<i>ɔ³³</i>	<i>βan⁴⁴</i>
insect/worm	<i>*dyung</i>	—	<i>jaŋ⁵⁵</i>	<i>suŋ¹¹</i>	—	—	<i>loŋ³³</i>	<i>tʃoŋ⁴⁴</i>
leg/foot	<i>*kan, ken, pey</i>	<i>*kriy = kray</i> (38)	<i>jou⁵⁵</i>	<i>hok³³</i>	<i>ja</i>	<i>da</i>	<i>læ³³</i>	<i>tʃja⁴⁴ (= leg)</i>
live/green	<i>*rang, ring</i>	<i>*s-rin</i> (404)	<i>saŋ¹¹tuŋ⁵⁵</i>	<i>a³³-saŋ⁵⁵pa³¹</i>	<i>əhənpu</i>	<i>ahin</i>	<i>thuŋ⁵⁵læ⁵⁵</i>	<i>ho⁴⁴hiŋ⁵²</i>
long	<i>*low</i>	<i>*low</i> (279)	<i>lou⁵⁵</i>	<i>a³³-lay³¹</i>	<i>laopu</i>	<i>alo</i>	<i>moŋ³¹</i>	<i>lo⁴⁴</i>
mother	<i>*m-na, n(y)u</i>	<i>*ma</i> (487)	<i>pu⁵⁵</i>	<i>pu¹¹</i>	<i>əpu</i>	<i>injoŋ</i>	<i>a⁵²nu⁴⁴, ho⁵²nu⁴⁴</i>	<i>a⁵²nu⁴⁴</i>
pestle	—	—	—	—	—	—	<i>ma⁵⁵</i>	<i>man⁵²</i>
uncooked rice	—	—	<i>aŋ¹¹</i>	—	<i>tak, woŋ</i>	—	<i>ɔŋ⁵⁵</i>	<i>βoŋ⁴⁴</i>
salt	<i>*gyum, tsa</i>	<i>*g-ryum</i> (245)	<i>tʃəm⁵⁵</i>	<i>tʃem⁵⁵</i>	<i>həm</i>	<i>sum</i>	<i>hum³³</i>	<i>hum⁴⁴</i>
shoulder	—	—	<i>phak⁵⁵</i>	<i>saŋ³³thuŋ³¹</i>	<i>phaktan</i>	<i>tʃhuakho</i>	<i>phak³³tʃoŋ³³</i>	<i>swak⁵²khaŋ⁴⁴/to⁴⁴</i>
sky	<i>*m-ka-n, r-məw</i>	<i>*m-ka-n, r-məw</i> (488)	<i>mau⁵¹</i>	<i>maw¹¹ow¹¹</i>	<i>wəŋ</i>	<i>raŋ</i>	<i>phom⁵⁵jaŋ³³</i>	<i>zan⁴⁴kho⁵²</i>
sun	<i>*ka, tsyar</i>	<i>*tsyar</i> (187)	<i>tʃa¹¹ju⁵⁵</i>	<i>a⁵⁵tʃe³³</i>	<i>wəphi</i>	<i>san</i>	<i>ʒaŋ⁵⁵he⁵⁵</i>	<i>zan⁴⁴vat⁴⁴</i>
wing	<i>*duŋ</i>	—	—	—	<i>jaŋ</i>	<i>araŋ</i>	<i>yaŋ⁵⁵</i>	<i>o⁴⁴zaŋ⁴⁴</i>

On the other hand, it could be argued that in the absence of evidence to the contrary, what is of paramount importance is that each language of the proposed sub-group demonstrates a significant number of lexical items that is not cognate with the reconstructed PTB forms expressing equivalent semantic content. The lack of correspondences for some etyma may simply be a corollary of the subsequent replacement of innovated forms by yet newer lexical innovations.⁴

In spite of the reservations prompted by the new data of Table 1, it is still arguably the case that eight lexical innovations constitute quite robust support for recognition of the Sal languages as a distinct branch of Tibeto-Burman, although the evidence for this must now be considered a little less compelling than was originally assumed when Burling's 1983 article first appeared.

Burling (2003) is the most recent classification of the Tibeto-Burman languages of north-east India. This divides the north-east geographically into a northern area containing the languages of Arunachal Pradesh north of the Miri Hills (these will not be discussed, as their place in Tibeto-Burman is peripheral to the focus of this paper), a central area containing a Bodo-Konyak-Jinghpaw branch (the erstwhile Sal grouping) and an eastern border area (corresponding fairly closely with central and southern Nagaland and Manipur) containing STC's Northern Naga and Southern Naga groups. Burling tentatively proposes Ao, Angami-Pochuri, Tangkhul and Zeme sub-groups, but remains uncommitted with respect to the possibility of these four clusters converging to form a single higher-level node or two or more independent branches of Tibeto-Burman. The precise evidence for proposing these sub-groups is not presented or discussed in the chapter, but is stated to be based upon lexical comparisons.

Given the inconclusive results from using lexical correspondences alone for sub-grouping in the Tibeto-Burman languages of north-east India, it is relevant to consider if other kinds of evidence might now shed a brighter light on these hitherto murky genetic affiliations. Unfortunately the field is still limited by a paucity of reliable data, but one such diagnostic is suggested by a rare kind of cardinal numeral system previously reported in a handful of languages spoken in central and southern Nagaland.

3. Innovative overcounting cardinal numeral patterns

Historical sources of data demonstrate that in the late 19th and early 20th centuries, typologically rare overcounting cardinal numeral systems existed in a number of languages spoken in central and southern Nagaland, but were unreported in any of the

⁴ I am grateful to Laurent Sagart for offering this perspective on the problem of non-cognate lexical items sharing equivalent semantic content.

Konyak languages, which have historically employed vigesimal or decimal numeral systems. An overcounting system was also absent in the grammar of Tangkhul (Pettigrew 1918) and Karbi (Stack & Lyall 1908), and was similarly absent in the 19th century sketches of Kachcha Naga (now known as Zeme) (Soppit 1885), Manipuri (Primrose 1888) and Lushai (now known as Mizo) (Lorrain & Savidge 1898). In all of these languages, enumeration was done using decimal numeral systems at the times they were documented.

The term ‘overcounting’ was coined by Menninger (1969) and was subsequently adopted by Hurford (1975) and others in studies of numeral systems. In an overcounting system, a value is expressed in relation to a higher parameter known as the augend; overcounting literally expresses the number of units counted towards the augend. Overcounting systems are extremely rare typologically and have only been reported in a handful of languages spoken in a few areas of the world: in Mayan languages spoken in Mexico (Aulie 1957), in the isolate Ainu spoken on Sakhalin Island in north-east Asia, and in the pre-Christian Germanic north of Europe (Menninger 1969:76). More recently, Mazaudon (2009) has reported a limited overcounting pattern in the Dzongkha language of Bhutan and traces of such a system in other Bodish languages; this is almost certain to have arisen independently of the development of overcounting numeral systems in the languages of Nagaland, given our knowledge of the likely migration routes of Bodish speakers across the Tibetan plateau and into the high valleys of the main Himalayan range.

Table 2: Decimal cardinal numerals of Ao dialects, numerals ONE to FIFTEEN

	Mongsen (Mills 1926)	Chungli (Mills 1926)	Changki (Mills 1926)	Longla (Mills 1926)	Yacham (Mills 1926)	Mongsen (Coupe 2007)
1	ākḥā	ākḥā	ākḥāt	khě	khāt	a-khət
2	ānět	ānā	ānět	āně	ānět	a-nət
3	āsām	āshām	āsām	āsām	āsām	a-səm
4	pūli	phüzü	pūli	phüzü	phüle	phəli
5	phāngă	pōngö	phāngă	pōngö	phōngö	phaŋa
6	tērök	tērök	tērök	tērök	tülök	tə-ruk
7	tēni	tēnět	tēni	tüně	tēnyět	thəni
8	tsīt	tī	tēzět	tüzět	tēsět	tshət
9	tükū	tökū	tükū	tīkū	tükhū	thuku
10	teră	tür	teră	tüwō	tülo	thəra
11	teră ākhāt	türü kha	terări ākhāt	tüwō khě	tüloli khāt	thəra-ɾə à-khət
12	teră ānět	tür ānā	terări ānět	tüwō āně	tüloli ānět	thəra-ɾə à-nèt
13	teră āsām	tür āshām	terări āsām	tüwō āsām	tüloli āsām	thəra-ɾə à-səm
14	teră pūli	türü phüzü	terări pūli	tüwō phüzü	tüloli phüle	thəra-ɾə phəli
15	teră phāngă	türü pōngö	terări phāngă	tüwō pōngö	tüloli phōngö	thəra-ɾə pəŋə

Table 2 above lists the cardinal numerals from ONE to FIFTEEN in five dialects of Ao. The data comes from Mills (1926), supplemented by recently collected data from Coupe (2007). The system is consistently decimal across all the varieties, making it rather unremarkable. However, the forms also recorded by Mills for the numerals SIXTEEN to NINETEEN in Table 3 below clearly demonstrate that all Ao dialects except Yacham additionally used overcounting patterns in their cardinal numeral systems, at least in the higher teen numerals.

Table 3: Historical overcounting numeral systems in four of five Ao dialects (with overcounting patterns in bold face).

	Mongsen (Mills 1926)	Chungli (Mills 1926)	Changki (Mills 1926)	Longla (Mills 1926)	Yacham (Mills 1926)	Mongsen (Coupe 2007)
16	mükyi müpen tērök	metsü maven tērök	müki müpen tērök	metsü maven tērök	tüloli tülök	thəɬa-ɬə tə-ruk
17	mükyi müpen tēni	metsü maven tēnēt	müki müpen tēni	metsü maven tünē	tüloli tēnyēt	thəɬa-ɬə thəni
18	mükyi müpen tsīt	metsü maven ti	müki müpen tēzēt	metsü maven tiüzēt	tüloli tēsēt	thəɬa-ɬə tshət
19	mükyi müpen tükū	metsü maven tökū	müki müpen tökū	metsü maven tikū	tüloli tükhū	thəɬa-ɬə thuku
20	mükyi	metsü	müki	metsü	tāmōŋ khāt Lit. ‘body one’	məki

According to my Mongsen Ao consultants, the expression *mükyi müpen tērök* in the top left-hand cell of Table 3 is understood literally as ‘(the) twenty not completed, (the) six’, i.e. the sixth unit towards twenty, meaning ‘sixteen’. This interpretation holds for the identical pattern demonstrated by the constructions expressing SEVENTEEN, EIGHTEEN and NINETEEN. Upon reaching TWENTY, the system temporarily reverts to the decimal pattern seen for numerals ONE to FIFTEEN in Table 2, until TWENTY-five is reached. At TWENTY-SIX the overcounting pattern is resumed until THIRTY is reached, when the decimal system is again adopted. Thereafter the cardinal numeral system switches between decimal and overcounting patterns according to the alternating mode described above, *ad infinitum*. One elderly Mongsen Ao speaker, Mr. Imti Luin, recalls how the American Baptist missionaries replaced the overcounting numeral system with an entirely decimal system after they set up their school in the Mongsen-speaking village of Mangmetong in 1911. This suggests that overcounting must have eventually become defunct in that dialect some time after it was documented by Mills in 1926, because overcounting has not survived in present-day Mongsen. The verb root *pən* seen in the overcounting construction is rarely encountered and currently appears to be

restricted to the semantic field of temporal quantification, as in the elicited example of (1).⁵

- (1) Mongsen Ao (Coupe 2007:118)
pa a-kəm phəli pən-ukù
 3SG NRL-year four complete-ANT
 ‘He has reached four years of age.’

The overcounting system was not confined to the teen numerals or just some dialects of the same language as assumed by Matisoff (1997:33), who referred to such systems in the Tibeto-Burman languages of Nagaland as ‘subtractive’, although he did observe that the formations were different from subtractive patterns found in Latin. Matisoff based his analysis solely on Marrison (1967), the incomplete data of which showed an unrecognized overcounting system for Mongsen Ao and Khonoma Angami, but a decimal system for Chungli Ao and Kohima Angami. Presumably this is what led Matisoff to conclude that there was no genetic basis for their numeral systems. Because Marrison only included the higher teen numerals in his word lists, Matisoff also concluded that this unusual system of counting was confined just to numerals SIXTEEN to NINETEEN. However, Clark’s (1893) *Ao Naga Grammar* provides independent evidence that the overcounting numeral system also existed in Chungli Ao at the end of 19th Century, that it extended into all of the higher compound numerals, and that its character was actually not subtractive. Her valuable discussion of the Chungli cardinal numeral system warrants citing *verbatim*:

As will be observed from the above [i.e. a table of numerals], the Aos have distinct names for the digits, and the compounds are regularly formed up to sixteen, as, ten and one are eleven ‘teri ka’, ten and two are twelve ‘teri ana’, &c.; also from twenty to twenty-six, twenty and one ‘metsyri ka, [sic]’ twenty and two ‘metsyri ana’, &c. The same with thirty, forty, &c. But when the six is reached in the compounds, the succeeding ten seems to be anticipated, and we have for sixteen ‘metsy maben trok’ twenty not brought six, equivalent to the sixteen before twenty; ‘metsy, [sic] maben tenet, the seventeen before twenty, &c. In the same manner from twenty-six to thirty, on reaching the six, thirty is anticipated, thus ‘semyr maben trok’ the six before thirty, twenty-six, ‘lir maben trok’ the six before forty, thirty-six, &c. Clark (1893[2002:45]).

⁵ Abbreviations used in examples: ANT anterior tense/aspect; NRL non-relational nominal prefix; SG singular number; RL relational nominal prefix.

In a table of cardinal numerals, Clark gives Chungli Ao examples for the numerals THIRTY-SIX, FORTY-SIX and FIFTY-SIX that are consistent with the pattern described in the cited extract and Mills' examples listed in Table 3 above. This demonstrates that the overcounting system was much more extensive than assumed by Matisoff (1997). Furthermore, if overcounting extended into the higher compound numerals in Chungli Ao, then by analogy it must have extended into the higher numerals in all the Ao dialects. It is a mystery how Marrison failed to notice this, as Clark (1893) was cited as one of the sources of his Chungli Ao data.

Yacham's being recorded as having a divergent numeral system is attributable to its location in the Phom-speaking area. Yacham shares with Phom a vigesimal numeral system common to the Konyak and Chang languages of northern Nagaland. The other Konyak group languages of the eastern border region—Tangsa, Nocte and Wancho—are spoken in the Changlang and Tirap districts of Arunachal Pradesh. Of these three, only Wancho had a vigesimal numeral system, the others demonstrating a decimal system. The overcounting system common to all other Ao dialects was almost certainly replaced by the vigesimal system of Phom through language contact. The Yacham term *tāmōŋ khāt* for TWENTY in Table 3 is literally 'one body' which, as Mills (1926:343) points out, possesses ten fingers and ten toes. The initial constituent of this numeral compound is cognate with the bound root *tə-maŋ* 'RL-body' of Mongsen Ao. It is therefore likely that the word for TWENTY in Yacham is a calque of the Phom vigesimal term. In an annotated footnote in Mills (1926:342), Hutton adds a comparative note describing how he had heard a Phom speaker use the expression 'a whole man' when asked how many people were present, meaning 'at least twenty'.

Since all other Ao dialects apart from Yacham have documented proof of an older overcounting numeral system that uses a negated verb form as a ligature, the weight of evidence supports the argument that this system is a genetic characteristic of Ao. If so, the overcounting numeral system must have existed in the Yacham dialect before it was replaced by the vigesimal system as a consequence of language contact. Mills (1926:2) reports that the inhabitants of Yacham spoke a dialect resembling Chungli Ao but followed Phom or Konyak customs. He was told by inhabitants of Yacham that they did not know whether they were Ao, Phom or Konyak, because they were not accepted by any of these tribes as kinsmen. Chungli Ao speakers report that they find the language spoken in the present-day Yong and Yacham villages largely unintelligible and attribute this to the influence of Phom.

Until the 1920's, a substantial number of languages spoken in the central and southern areas of the Naga Hills used overcounting patterns in their cardinal numeral systems to varying degrees. The data is patchy because our appreciation of such patterns is dependent upon the whim of authors to document numeral systems beyond ONE to

TEN and the higher round decads, but it is significant that the languages for which we have documented evidence of overcounting all belong to the LSI's Western and Central sub-groups, respectively correlating with Burling's Angami-Pochuri and Ao clusters, and with STC's Northern Naga and Southern Naga groups of the Kuki-Naga division.

Table 4: Examples of historical overcounting patterns in Angami, Sumi and Lotha (with overcounting expressions in bold face).

	Angami (McCabe 1887)	Sumi (Hutton 1921b)	Lotha (Witter 1888)
5	pangu	pongu	mūngo
6	suru	tsogha, soghoh	tirōk
7	thenā	tsini	ti-ing, tscang
8	thethā	tache, thache	tizā
9	tekwū	tuku	tōkū
10	kerr	chūghi	taro, tarā, tero
11	kerr o pokrō	chūghi-khaki	taro sū, or si ekhā
12	kerr o kennā	chūghi-kini	taro sū enni
13	kerr o sê	chūghi-kūthu	taro sū etham
14	kerr o dā	chūghi-bhidi, -bidi	taro sū mezū
15	kerr o pangu	chūghi-pongu	taro sū mūngo
16	kerr o suru	chūghi-tsoga, soghoh	taro sū tirōk, or mezūna mekwū m-pen
17	mekwū pemo thenā	muku-ma tsini (or chūghi-tsini)	taro sū tīng
18	mekwū pemo thethā	muku-ma thache (or chūghi-thache)	taro sū tizā
19	mekwū pemo pokrō	muku-ma tuku (or chūghi-tuku)	taro sū tōkū
20	mekwū	muku	mekwi, or mekwū
21	mekwū pokrō	muku-khaki	mekwū sū thamdrowe ekha
22	mekwū kennā, &c.	muku-(na-)khini	—
27	serr pemo thenā	segghi kumpa tsini (segghi kupvuma tsini)	—
28	serr pemo thethā	segghi kumpa thache etc.	—
29	serr pemo pokrō	—	—
30	serr	segghi	thamdro

When Angami was documented by McCabe in 1887,⁶ the language used a similar overcounting pattern to that found in Ao, except that it began at the seventh unit towards the augend. The Sumi pattern as documented by Hutton (1921b) was identical to Angami in beginning overcounting at the seventh unit in the teens and continuing the series until the ninth unit before the augend, then reverting to a decimal system at each round decad. It is also identical in using a verb stem negated by a suffix. Negative suffixes are a fairly unusual feature of languages of north-east India and thus may constitute evidence of a close genetic relationship between languages sharing this morphological feature.

The numeral system of Lotha was the most divergent because it appears to have used a decimal system as well as subtractive and overcounting patterns. According to the description of Mills (1922:211), who included a grammatical sketch of Lotha in his anthropological monograph, SIXTEEN could be expressed either decimally as *taro sü tirōk* ('ten and six'), or subtractively as *mezūna mekwü m-pen*, literally 'by four twenty not making'. Both of these patterns were also documented in the previous century by Witter (1888), as demonstrated by the data of Table 4 above. While there was no mention of overcounting in Mills' (1922) discussion of Lotha numerals, in his earlier description of the numeral system Witter (1888:27) specifically describes a method of counting by Lotha speakers that can only be interpreted as the overcounting pattern common to the languages of the LSI's Western and Central sub-groups. The last sentence in the quotation below attests to the possibility of using either overcounting or subtractive patterns in this language.

On reaching twenty-one they often call attention to the fact that thirty is the next ten, by saying mekwü sü thamdrowe ekhā = twenty and, toward thirty, one, after which they count on to thirty calling the digits in regular order, thus, mekü sü enni = twenty and two, &c. Then again in the same way. They also sometimes proceed throughout by mentioning each time how many have been counted in the direction of the next ten, or how many their calculation falls short of it.

Languages of the region that have documented overcounting patterns often use a negated verb stem as the linking word between the augend and the numeral denoting the progress made towards the augend. This negated verb usually expresses an insufficient amount, e.g. 'not complete' in Mongsen Ao (Coupe 2007:118), 'not brought' in Chungli Ao (Clark 1893[2002:45]), 'not entire' in Sumi (Hutton 1921b:272, 420), and 'falling

⁶ According to Hutton (1921a:294), McCabe's grammar is based on 'a sort of amalgamation of the dialects of the Khonoma group of which Jotsoma, Khonoma and Mozema are the principal villages.'

short of’ in Angami (Hutton 1921a:314). The Angami form actually contains a negative suffix *-mo* that is not represented by Hutton’s translation.⁷ A partial exception to the use of a negated verb stem is demonstrated by Lotha, which appears to use the negated verb for the subtractive pattern but not for the overcounting pattern, at least on the basis of the very slim amount of data provided by Witter (1888). Two dialects of Rengma (to be discussed below) also appear to use something other than a negated verb in their overcounting expressions.

Hutton (1921b:272) comments in a footnote that the overcounting pattern was rapidly becoming obsolete in Sumi, and that the younger generation was moving towards the exclusive use of the parallel decimal pattern. He makes the observation that a decimal system was similarly supplanting overcounting in Angami (1921a:314). The replacement of overcounting patterns in these languages can probably be attributed to the language engineering efforts of the missionaries, who were encouraging the use of decimal systems in their schools at the time in order to teach children arithmetic more effectively. This policy was stated by the American Baptist missionary Mary Mead Clark (1893 [2002:45]) in her *Ao Grammar*:

This method of counting is very objectionable to children learning the use of figures, as in adding up a column if the amount is seventeen—‘metsy maben tenet’—the mind catches the twenty, and *two* is very likely carried instead of *one* to the next column. In the schools an effort is being made to discard the above irregularities, and count regularly thus, ‘teri trok’ sixteen, ‘metsyri trok’ twenty-six, &c.

Lastly, overcounting was used in the recorded dialects of the Eastern Rengma group, somewhat more marginally in the northern group of Western Rengma, and extensively in the southern group of Western Rengma, all of which belong to the LSI’s Western sub-group. The dialects of the Eastern Rengma group spoken in the villages of Meluri, Lephori and Sahunya had the most prolific overcounting system encountered in the historical sources. A table of numerals provided by Mills (1937:292-293) shows that the overcounting system operated from SIXTEEN to NINETEEN until TWENTY was reached. Mill’s description then suggests that in counting from TWENTY-ONE to TWENTY-NINE, THIRTY functioned as the augend, and overcounting proceeded in a similar fashion for the remaining higher compound numerals up to ONE HUNDRED. The numeral compounds

⁷ Mr. Visakuolie Vakha, a native speaker of Khonoma Angami, translates *pe* as ‘add’ and *-mo* as ‘not’, but feels that a more accurate rendering of this in English would be something approximating ‘not reaching’. Now in his early forties, he recalls the old folk of Khonoma village counting in this fashion in his youth.

given for TWENTY-ONE, THIRTY-ONE, FORTY-ONE and so on up to NINETY-ONE give support to this interpretation. Regrettably, these were the only higher compound numerals documented.

Table 5: Examples of historical overcounting patterns in three dialects of Rengma (with overcounting expressions in bold face), from Mills (1937:292-293).

	Sthn group of W. Rengma	Nthn group of W. Rengma	E. Rengma
1	me	kesü	ke, kesü
2	khohüŋg	keni	keni
3	khüŋgshäng or shang	keshäng	keche
4	pezi	mezü	mezu
5	pfü	manga	manga
6	tsaro	tüo	taro
7	tsanü	tüghü	terü
8	tütse	tüza	tüzu
9	tükhü	tükhu	tokhu
10	tsarü	taa	tera
11	tsarü me chü lit. ‘ten one added’	taake	terake
12	tsarü khohüŋg chü	taakeni	terakeni
13	tsarü khüŋgshäng chü	taakecham	terakecha
14	tsarü pezi chü	taamezü	teramezu
15	tsarü pfü chü	taamanga	teramanga
16	nki pamo tsaro (lit. ‘six towards twenty’)	kwü she tüo (kwü is a contraction of mëkwüŋg)	mükwe shun toro
17	nki pamo tsanü	kwü she tüghü	mükwe shun terü
18	nki pamo tütse	kwü she tüza	mükwe shun tüzu
19	nki pamo tükhü	kwü she түkhu	mükwe shun tokhu
20	nki	mëkwüŋg	mükwe
21	nki me chü, etc.	mëkwüŋg kesü, etc.	achera kesu, etc.
26	shenrü pamo tsaro, etc.	chaa she tüo, etc.	achera toro, etc.
30	shenrü	chaa	chera

The southern group of Western Rengma is rather like Ao, Angami and Sumi in beginning the overcounting pattern at SIXTEEN and using a negated verb stem. Unfortunately the data is too scant to determine the function and meaning of *she* and *shun* in the northern group of Western Rengma and Eastern Rengma respectively, but it suffices to say that collectively, the cardinal numeral systems of the Rengma dialects have much in common with other languages of the LSI’s Western and Central sub-groups.

4. Overcounting as a diagnostic tool for sub-grouping

At present, sub-groupings of the Tibeto-Burman languages of north-east India are based on lexical comparisons and geographic proximity. But as Thurgood (2003:5) points out, sub-groupings based on shared morphological innovations or full historical reconstructions are more convincingly established than are those that are based on lexical correspondences alone. Given their rarity in the world's languages, the overcounting cardinal numeral systems that once existed in the languages of central and southern Nagaland constitute a hitherto unexploited resource for establishing genetic relationships with greater confidence than that afforded exclusively by lexical correspondences augmented by geographical considerations, especially since evidence of the overcounting pattern is not found in any languages external to the LSI's Western and Central sub-groups.

Table 6: Tibeto-Burman languages of the eastern border region that historically lacked overcounting patterns

	Tangkhu (Pettigrew 1918)	Zeme (Soppit 1885)	Karbi (Stack & Lyall 1908)	Manipuri (Primrose 1888)	Lushai [Mizo] (Lorrain & Savidge 1898)
1	a-kha (khat-ka)	kât	ĩsĩ	amá	pa-khat
2	kha-ni	ganâ	hĩnĩ	ani	pa-hnih
3	ka-thum	gũjũm	kethòm	ahũm	pa-thum
4	ma-li	mâdai	philĩ	mari	pa-li
5	pha-nga	mingêo	phôngô	mangâ	pa-nga
6	tha-ruk	sũrũk	theròk	tarũk	pa-ruk
7	shi-ni	senâ	theròk-sĩ	tarêt	pa-sari
8	chi-shat	dasât	něrkèp	nipân	pa-riat
9	chi-ko	sũgũi	sirkèp	mâpan	pa-kua
10	thără	gârêo	kèp	tarâ	shom
11	thără-ta ăkha	‘There are no single words to express numbers between ten and twenty’ (Pettigrew 1918:8).	krē-ĩsĩ	tarâ-mâthoi	shom leh pa-khat
12	thără-ta khă-ni		krē-hĩnĩ	tarâ-nithoi	shom leh pa-hnih
13	thără-ta kă-thum		krē-kethòm etc.	tarâ-hũmthoi	shom leh pa-thum
14	thără-ta mă-ti			tarâ-mari	shom leh pa-li
15	thără-ta phă-ngá			tarâ-mangâ	shom leh pa-nga
16	thără-ta thă-ruk			tarâ-tarũk	shom leh pa-ruk
17	thără-ta shĩ-ni			tarâ-tarêt	shom leh pa-sari
18	thără-ta chi shat			tarâ-nipân	shom leh pa-riat
19	thără-ta chĩ-ko			tarâ-mâpan	shom leh pa-kua
20	maga	eĩkai	ingkoi	kul	shom hnih

Of course, our knowledge of historical cardinal numeral systems is entirely dependent upon the chance documentation of the Tibeto-Burman languages of north-east India and the fulsomeness with which they were recorded for posterity. Evidence from the historical records suggests that overcounting cardinal numeral systems were already endangered in the early 20th century, and it is possible that such patterns may have become decimalized well before any language that once had such a system could be documented, especially in areas where Christian missionaries were active. Be that as it may, it is reassuring to find that the overcounting pattern is not found outside of the LSI's Western and Central sub-groups in the historical literature at hand, as demonstrated by the data of Table 6 above. The grammars and anthropological sketches of these languages were all published between 1885 and 1918; therefore if they happened to have had overcounting patterns in their numeral systems at the time, it would be likely for these to have been captured in their descriptions, particularly because Christian missionaries tended to focus on writing grammars in order to develop materials for proselytising and educational purposes. The fact that the cardinal numeral systems of the languages of Table 6 are all basically decimal suggests that overcounting was foreign to their grammars at the time of their documentation.

It is unlikely that the overcounting systems demonstrated in this paper are based on loan translations, as the negative morpheme in the linking verb is either a prefix or a suffix according to the structural characteristics of the specific language, while some languages lack a negated verb stem in their overcounting expressions entirely. Secondly, the extent of the overcounting system varies from language to language. Ao commences the overcounting pattern from the sixth numeral in each successive decad from the teen numerals onwards, Angami starts from the seventh numeral in the teen numerals onwards, while the expression for the numeral TWENTY-ONE of Eastern Rengma in Table 5 shows that overcounting begins at the first unit towards the augend of THIRTY, the pattern being repeated for each higher decad. Thirdly, if these overcounting patterns have been borrowed as a consequence of language contact, then it is difficult to explain why such a typological rarity has never been reported in languages of the Konyak group. Phom, Chang, Khamniungan and Konyak are all spoken in areas contiguous with those of the overcounting languages, yet none of these Konyak languages has had any semblance of an overcounting pattern reported in their numeral systems. We may therefore assume that a group of languages that evinces the same type of extremely rare cardinal numeral system demonstrates collective evidence of a single historical innovation that has been inherited from a common parent language.

We must also consider the very small likelihood of such a system being borrowed. Lehmann (2005:391-392) notes that while virtually anything can be borrowed, there is a continuum of borrowability. Independent lexical items are most easily borrowed, and

grammatical paradigms are the most resistant. The overcounting numeral systems formerly found in certain languages of Nagaland must be much closer to the grammatical paradigm pole of this cline, which would make them resistant to borrowing according to Lehmann's model. Furthermore, we know from the historical records that overcounting systems were much less robust than cardinal numeral systems, which explains their synchronic obsolescence. Their ephemeral nature also renders overcounting systems less likely to be borrowed as the result of language contact.

Figure 4 below presents a recent classification of Tibeto-Burman languages of the eastern border area of India by Burling (2003), which has been modified to take the findings of this paper into account. Burling's classification includes languages belonging to his Ao, Angami-Pochuri, Zeme and Tangkhul groups, plus a number of others of uncertain affiliation (represented here by broken lines). The heavy lines linking the Ao and Angami-Pochuri sub-groups to a new Angami-Ao node replace broken lines in Burling's original figure and represent this paper's contribution to the genetic classification of these languages, using overcounting as evidence of an inherited innovation. Data still needs to be collected on Yimchungrü and Sangtam in this sub-grouping (assuming that elderly speakers who can remember overcounting are still alive), but anecdotal evidence from Sangtam speakers suggests that overcounting once existed in their language as well. Taking into account the data of Table 6, we can see that the languages that lack evidence of overcounting patterns are not immediately related to those falling under the newly proposed Angami-Ao node.

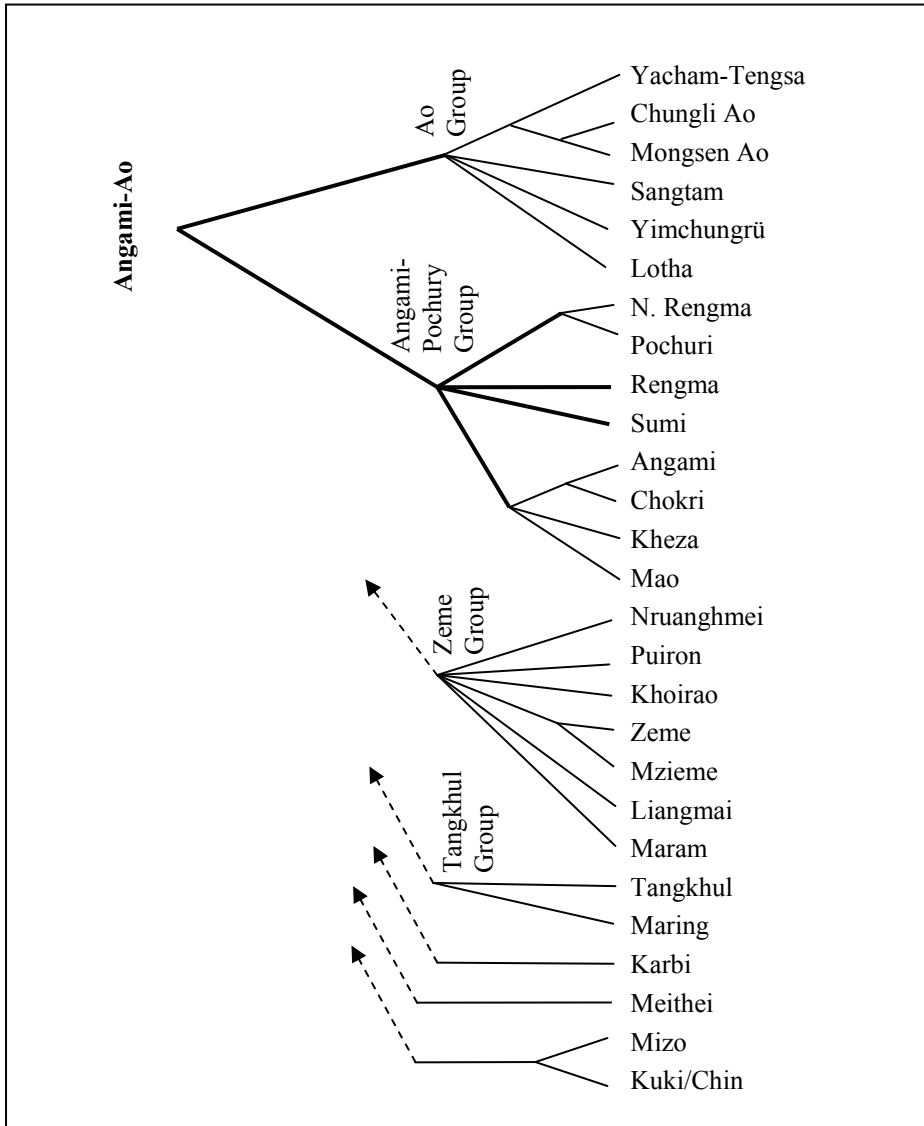


Figure 4: Relationships among the languages of the eastern border (Burling 2003:184, with modifications)

Given that we only have two comprehensive descriptions of these languages at present—viz. Meithei (Chelliah 1997) and Mongsen Ao (Coupe 2007)—it would be prudent to remain agnostic with respect to the positing of additional higher level nodes or branches until further evidence for genetic affiliations can be established.

References

- Aulie, Wilbur. 1957. High-layered numerals in Chol (Mayan). *International Journal of American Linguistics* 23.4:281-283.
- Bauman, James. 1974. Pronominal verb morphology in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 1.1:108-155.
- Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. Cambridge: Cambridge University Press.
- Burling, Robbins. 1983. The Sal languages. *Linguistics of the Tibeto-Burman Area* 7.2: 1-32.
- Burling, Robbins. 2003. The Tibeto-Burman languages of Northeastern India. *The Sino-Tibetan Languages*, ed. by Graham Thurgood & Randy J. LaPolla, 169-191. London: Routledge.
- Burling, Robbins, and L. Amon Phom. 1998. Phom phonology and word list. *Linguistics of the Tibeto-Burman Area* 21.2:13-42.
- Burling, Robbins, and Mankai Wanchu. 1998. Wanchu phonology and word list. *Linguistics of the Tibeto-Burman Area* 21.2:43-71.
- Chelliah, Shobhana Lakshmi. 1997. *A Grammar of Meithei*. Berlin & New York: Mouton de Gruyter.
- Clark, Mrs. Edward Winter (Clark, Mary M.). 1893[2002]. *The Ao Naga Grammar*. Delhi: Gian Publications.
- Coupe, Alexander R. 2004. Recounting the numeral systems of Nagaland: old evidence for a new sub-grouping? Paper presented at the 10th Himalayan Languages Symposium, December 1-3, 2004. Thimphu, Bhutan.
- Coupe, Alexander R. 2007. *A Grammar of Mongsen Ao*. Berlin & New York: Mouton de Gruyter.
- Damant, G. H. 1880. Notes on the locality and population of the tribes dwelling between the Brahmaputra and Ningthi Rivers. *Journal of the Royal Asiatic Society of Great Britain and Ireland*, New Series 12.2:228-258.
- DeLancey, Scott. 1989. Verb agreement in Proto-Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 52.2:315-333.
- Grierson, Sir George A., and Sten Konow. (eds.) 1903-1928. *Linguistic Survey of India*, Vol. 1 Part 1: Introductory, and Vol. 3 Part 2: Specimens of the Bodo, Nāgā and Kachin groups. Calcutta: Superintendent of Government Printing.
- Hurford, James R. 1975. *The Linguistic Theory of Numerals*. Cambridge & New York: Cambridge University Press.
- Hutton, John Henry. 1921a. *The Angami Nagas*. London: MacMillan.
- Hutton, John Henry. 1921b. *The Sema Nagas*. London: MacMillan.

- LaPolla, Randy J. 1992. On the dating and nature of verb agreement in Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 55.2:298-315.
- Lehmann, Christian. 2005. On the methodological bases of genetic language comparison. *Language Research* 41.2:379-404.
- Li, Fang-Kuei. 1977. *A Handbook of Comparative Tai*. Honolulu: University Press of Hawai'i.
- Lorrain, J. Herbert, and Fred W. Savidge. 1898. *A Grammar and Dictionary of the Lushai Language (Dulien Dialect)*. Shillong: Assam Secretariat Printing Office.
- Marrison, Geoffrey Edward. 1967. *The Classification of the Naga Languages of North-East India*, Vol. I (Texts) and Vol. II (Appendices). London: SOAS, University of London dissertation.
- Matisoff, James A. 1997. *Sino-Tibetan Numeral Systems: Prefixes, Protoforms and Problems*. Pacific Linguistics B-114. Canberra: The Australian National University.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley: University of California Press.
- Mazaudon, Martine. 2009. Number building in Tibeto-Burman languages. *North East Indian Linguistics*, Vol. 2, ed. by Stephen Morey & Mark W. Post, 117-148. New Delhi: Cambridge University Press India.
- McCabe, Robert Blair. 1887. *Outline Grammar of the Angami Naga Language*. Calcutta: Superintendent of Government Printing.
- Menninger, Karl. 1969. *Number Words and Number Symbols: A Cultural History of Numbers*, translated by Paul Broneer for the revised German edition. Cambridge: MIT Press.
- Mills, James Philip. 1922. *The Lotha Nagas*. London: MacMillan.
- Mills, James Philip. 1926. *The Ao Nagas*. London: MacMillan.
- Mills, James Philip. 1937. *The Rengma Nagas*. London: MacMillan.
- Ostapirat, Weera. 2000. Proto-Kra. *Linguistics of the Tibeto-Burman Area* 23.1:1-251.
- Pettigrew, William. 1918. *Tangkhul Naga Grammar and Dictionary (Ukhrul Dialect) with Illustrative Sentences*. Shillong: Assam Secretariat Printing Office.
- Primrose, Arthur John. 1888. *A Manipuri Grammar, Vocabulary, and Phrase Book, to which are added some Manipuri Proverbs, and Specimens of Manipuri Correspondence*. Shillong: Assam Secretariat Printing Office.
- Shafer, Robert. 1950. The Naga branches of Kukish: vocalism. *Rocznik Orientalistyczny* 16:467-530.
- Shafer, Robert. 1955. Classification of the Sino-Tibetan languages. *Word* 11.1:94-111.
- Soppit, C. A. 1885. *A Short Account of the Kachcha Naga (Empeo) Tribe in the North Cachar Hills, with an Outline Grammar, Vocabulary & Illustrative Sentences*. Shillong: Assam Secretariat Printing Office.

- Stack, Edward, and Sir Charles Lyall. (eds.) 1908. *The Mikirs, from the Papers of the Late Edward Stack*. London: David Nutt.
- Thurgood, Graham. 2003. A subgrouping of the Sino-Tibetan languages: the interaction between language contact, change, and inheritance. *The Sino-Tibetan Languages*, ed. by Graham Thurgood & Randy J. LaPolla, 3-21. London: Routledge.
- Thurgood, Graham, and Randy J. LaPolla. (eds.) 2003. *The Sino-Tibetan Languages*. London: Routledge.
- Witter, W. E. 1888. *Outline Grammar of the Lhōtā Nāgā Language: With Vocabulary and Illustrative Sentences*. Calcutta: Superintendent of Government Printing.
- Woodthorpe, R. G. 1882. Notes on the wild tribes inhabiting the so-called Naga Hills, on our north-east frontier of India, Part I. *The Journal of the Anthropological Institute of Great Britain and Ireland* 11:56-73.

[Received 20 September 2010; revised 4 February 2011; accepted 18 May 2011]

Division of Linguistics and Multilingual Studies
School of Humanities and Social Sciences
Nanyang Technological University
Singapore 637332
arcoupe@ntu.edu.sg

那加蘭邦藏緬語之「預估進位」數詞系統 及其對系譜分類的作用

Alexander R. Coupe

南洋理工大學

本文評估印度東北部那加蘭邦藏緬語族語言之間親屬關係最具影響力的某些論述，著重探討是否有足夠的證據將其建構為藏緬語族「那加語支」。調查結論認為：儘管音韻和詞彙的對應似乎足以說明那加蘭邦語言都屬於藏緬語族語言；討論當地語言內部的系譜分類時，那些現象卻作用有限。除了孔亞克諸語言基於詞彙創新，公認與景頗語及博多-加羅語同群；其餘那加蘭邦語言的下位分群，仍因缺乏有效的標準而未有定論。本文對此提供了新的證據。根據十九世紀末二十世紀初的文獻記錄，那加蘭邦中、南部的語言（即奧和昂那米-波丘語言）曾經有類型罕見的「預估進位」數詞系統。該複雜的數詞系統，經由歷史比較的分析，不像是彼此移借的結果，也不曾見於鄰近的孔亞克、卡爾比、澤梅、庫基及唐庫爾諸語言中。本文認為，那加蘭邦中、南部的藏緬語族語言「預估進位」數詞模式反映傳承的創新演變，顯示這些語言之間有密切的系譜關係。

關鍵詞：藏緬語，預估進位，數詞，那加語，系譜分類