Northern Sangtam phonetics, phonology and word list

Coupe, Alexander R.

2020

Coupe, A. R. (2020). Northern Sangtam phonetics, phonology and word list. *Linguistics of the Tibeto-Burman Area*, 43(1), 147-189. doi:10.1075/ltba.19014.cou

<https://hdl.handle.net/10356/146605>

<https://doi.org/10.1075/ltba.19014.cou>

© 2020 John Benjamins Publishing Company. All rights reserved. This paper was published in *Linguistics of the Tibeto-Burman Area* and is made available with permission of John Benjamins Publishing Company.

Downloaded on 26 Apr 2023 18:53:04 SGT

Northern Sangtam phonetics, phonology and word list

Alexander R. Coupe
Nanyang Technological University, Singapore

Abstract

This paper presents a comprehensive phonetic and phonological description of Northern Sangtam, an essentially undescribed Tibeto-Burman language of central Nagaland belonging to the Aoic subgroup. It is a noteworthy language from a number of phonological perspectives, not least because its phoneme inventory contains two of the world's rarest phonemes: a pre-stopped bilabial trill, and a doubly-articulated labial-coronal nasal. These unique segments are described in detail, and an attempt is made to determine how they might have developed their phonemic status. The tone system is also of interest, as it demonstrates evidence of debuccalization resulting in the development of a new high tone. Following a systematic description of the syllable and word structure, the tone system, and the segmental phonology, some observed age-related differences in the phoneme inventory are discussed. The paper is linked to an online repository containing the audio-visual data and transcribed word lists of approximately 900 items, based on the recorded utterances of eight speakers.

Key words: Tibeto-Burman, Indo-Burmic, Aoic, Sangtam, phonetics, phonology, tone, historical linguistics, language documentation, typology, Nagaland

1 Introduction

This article presents an analysis of the phonology of Northern Sangtam [saŋ¹¹taŋ³³] (ISO 639-3: nsa), also known as Upper or Western Sangtam, and by the Sumi (a.k.a. Sema) term 'Lophomi' in the colonial-era literature. Northern Sangtam is a Tibeto-Burman language spoken on five ranges in the Longkhim-Chare sub-division of Tuensang District in Nagaland, a hill state of Northeast India sharing an international border with Myanmar. It is presently used as a medium of instruction up to the fourth standard, and there are plans to teach it up to the eighth standard once language teachers are trained and appointed. Children still acquire it as their first language in the twenty-two Northern Sangtam-speaking villages of Nagaland. It is assumed that there are at least two major dialects of the language and quite probably a number of minor dialects restricted to particular villages or ranges, but confirmation of the extent of dialectal diversity must await future research opportunities. According to the 2011 Census of India there are 76,000 speakers;¹ this figure enumerates all Sangtam speakers, including those of Kiphire District, so it is not specifically known how many of that number exclusively speak the Northern Sangtam dialect.

The description is linked to an online repository containing transcribed word lists totalling approximately 900 lexical items and arranged according to semantic domain and/or word class, together with their sound files, video files and metadata. The word list used for the bulk of this phonological analysis conforms closely to those used to reconstruct Tibeto-Burman proto-forms in Benedict 1972 and Matisoff 2003, the motivation for this choice being to facilitate comparisons with existing work. The corpus has been extended by the inclusion of words that are areally relevant to the languages of Northeast India, and other words that were spontaneously offered by consultants to demonstrate the realizations of particular phonemes and their

¹ <http://www.censusindia.gov.in/2011Census/Language-2011/Statement-1.pdf>

phonological contrasts. A second shorter word list containing the data of six male speakers was initially collected specifically to explore the acoustic phonetic characteristics of some segmental phonemes, and to facilitate a forthcoming instrumental analysis of the tone system. As it turns out, this smaller corpus also proved valuable for revealing age-related differences in the phonological inventories of Northern Sangtam speakers. The data were elicited using English or Nagamese as the contact language, depending upon the familiarity of the consultant with each of these languages. Audio-visual files containing the recorded data, accompanying word list transcriptions and the metadata can be accessed in Nanyang Technological University's Institutional Research Data Repository.² An acoustic analysis of the language based on the same data is currently in preparation and will serve to corroborate and augment the auditory analysis presented in this paper.

The paper first provides a comprehensive overview of previous work on the language in §2 before turning to a discussion of the dialects of Sangtam and persisting issues with the genetic classification of the languages of central and southern Nagaland in §3. Sources of data and background information on native speaker consultants are described in §4, and limitations of the study are briefly addressed in §5. The description of the phonology begins in §6 with a discussion of syllable structure, followed by the findings of an auditory analysis of the suprasegmental phonology in §7. Consonant phonemes are described in §8 according to their place and manner of articulation, after which the vowel phoneme inventory is discussed in §9. The considerable age-related differences in consonant phoneme inventories are addressed in §10. The concluding section of §11 summarizes the most important findings and flags aspects of the analysis that could benefit from further investigation. Lastly, the online materials containing the above-mentioned word lists and audio-visual data completes this description of Northern Sangtam phonology.

2 Previous work

Northern Sangtam is a critically underdocumented language that has hitherto received precious little attention from scholars since colonial British explorers first encountered the language at the end of the 19th century. The following discussion reviews all known work from the earliest to the most recent.

The Linguistic Survey of India (Grierson [1903-1928] 1967, hereafter LSI) presents a word list provided by Captain A.E. Woods, an Indian Civil Service officer who recorded 116 individual words, phrases and short sentences in a language he referred to as 'Thukumi' in the winter of 1899. This language was spoken by a group residing in the Tita Valley (a.k.a. the Tizü Valley) east of the Dikhu River, and was observed at the time to be very similar to another language known as Sangtam. LSI also cites some observations by Mr Davis, the first Sub-Divisional Officer of Mokokchung District, who noted that 'Thukumi' was the Sema word for 'Sangtam', and that speakers 'inhabit the upper portion of the Tita valley, the whole of the valley of the Nāzār' stream and

² See <https://researchdata.ntu.edu.sg/dataverse/sangtam>

extend across the Tita-Dikhu watershed to just opposite the Āo village of Mokokchung' (LSI [1903-1928]1967: 290).

The first significant work on the language and its speakers was an ethnographic description published in German by Kauffmann (1939). The paper describes the material and social culture of the Northern Sangtam as it was encountered in the 1930s and includes a word list of approximately twelve pages, plus four pages of example phrases and sentences in an appendix. Kauffmann (1939: 228) states that two brothers from Chare village provided the linguistic data. Interestingly, a former chairman of the Sangtam Literature Society expressed concern when I informed him that I had initially recorded the data of a Chare speaker for the purpose of phonological analysis, as he considered the Sangtam variety spoken in Chare village to be too influenced by the neighbouring Chungli dialect of Ao and thus not truly representative of Northern Sangtam. Kauffmann (1939: 208) reported at the time that most Sangtam speakers understood Chungli Ao. He also mentioned that one *khel* ('administrative ward' < Assamese) of Chare village was Ao, and that some Chang speakers also lived in the village (ibid.: 220).

Similarly, the Indian Civil Service administrator Hutton reported in his tour diary of 1923 that Chare village had an Ao *khel*. He noted that some Sangtam villages (Chatongre, Chimongre and Chongtore) followed certain Chang customs and contained a prominent admixture of Chang blood, and he predicted that they were very likely to turn into Chang villages eventually (Hutton [1929]1986: 44–46). The extent to which Ao, Chang and possibly other languages of central Nagaland may have influenced Northern Sangtam is beyond the immediate scope of this paper, but certainly remains an attractive topic for further investigation.

Marrison's (1967) dissertation includes a substantial number of lexical items recorded in the Lophomi (Northern) dialect, which he elicited in 1963 from a native speaker in Guwahati, Assam. Tone is not transcribed. He also reports the existence of a central Thukumi dialect and a southern Pochuri dialect spoken in the eastern Chakhesang sub-district of Kohima district. Bouchery & Sangtam (2012: 9) mention in a footnote that Southern Sangtam speakers residing in Kohima District chose to merge with the Eastern Rengma and initially the Eastern Angami in the 1960's to form a new amalgamation named 'Chakhesang',³ which was later revised to 'Pochury' after the Angami withdrew from this new grouping. This represented merely a political alliance, and their respective languages remained distinct. It is likely that Marrison's "Pochuri and Thukumi dialects" can both be subsumed under the southern dialect of Sangtam spoken in present-day Kiphire District. The extent to which the northern and southern dialects of Sangtam diverge is yet to be determined.

Kumar 1973 is a small dictionary of Sangtam published by the Nagaland Bhasha Parishad (Nagaland Language Council), designed primarily for use by Indian government administrators from outside the state. It was not consulted in the preparation of this article.

³ The neologism 'Chakhesang' represents an amalgamation of the first syllables of Chokri, Khezha and Sangtam. Chokri and Khezha are two languages predominantly spoken in Phek district of southern Nagaland and neighbouring districts of northern Manipur.

Weidert (1987) presents a small amount of Sangtam data in his investigation of tonology in Tibeto-Burman, but does not provide a phonological analysis of any of the words or document the name(s) of the village(s) from which they were collected. According to Bruhn (2014: 9), Weidert's book contains a total of only ~60 transcribed Sangtam forms.

Bouchery & Sangtam (2012) describe the kinship terminology of Sangtam. Their paper also provides a description of the distribution of Sangtam clans and dialect names, prior to providing a detailed description of the kinship system. The authors report on some unique characteristics for the region, such as the merging of the names for paternal aunt and mother, and some unusual cousin terminologies (see the discussion of pp. 20–21). As one of the authors is a native speaker, the description of dialects and their locations (p. 11) is particularly valuable. In addition to the principal Longkhim (Northern) and Kiphire (Southern) dialects, the paper mentions that there are several variants recognized by speakers, including Sangphure, Hurong, Phelungre, and Alisopur. The total number of Sangtam dialects is unknown, but these dialect names may coincide with lects confined to particular villages or ranges. From their description, we may also infer that the Tizü-Dikhu watershed could form the geographical border between the main Longkhim-Chare (Northern) and Kiphire (Southern) dialects.

Bruhn 2014 is a relatively recent thesis that uses the above-mentioned sources and secondary materials available on other languages to reconstruct the intermediate proto-language of Ao dialects and the Aoic group of central Nagaland (which he calls “Proto-Central Naga”). Some of Bruhn's findings are drawn upon to support the phonological analysis of the primary data at hand.

Imchen 2014 is an eleven-page sketch of Sangtam ethnography and aspects of grammar. As the sources of the data were not revealed and much of the phonological description and analysis diverges markedly from the Northern Sangtam dialect described here, it was not consulted.

Coupe 2015 is an acoustic description focusing specifically on the typologically rare pre-stopped bilabial trills of Northern Sangtam. The paper also briefly discusses other aspects of the segmental phonology. With the benefit of an expanded corpus and further insights, the present article revises and updates the phonological overview and corrects some aspects of that preliminary analysis.

Lastly, *Yusip: Bilingual Dictionary of Sangtam-Naga* is a dictionary of 681 pages published by the Sangtam Literature Board in 2016. It is a welcome addition to the very few materials currently available on the Northern dialect, but unfortunately suffers from an orthographic representation that is not based on phonemic principles. Nevertheless, it is commendable that unlike most locally-produced dictionaries, it makes an attempt to represent tone, although a check of its accuracy with my main consultant suggests that the tonal representation is not always reliable; nor is the phonemic representation. The largely avoidable shortcomings of this dictionary demonstrate just how important it is for the minority linguistic communities of Northeast India to be advised by professionally trained linguists.

3 Dialects and genetic classification

Sangtam is a language of the Aoic group of central Nagaland, which traditionally includes Lotha (a.k.a. Lhota), the Ao dialects Chungli, Mongsen, Changki and Yacham-Tengsa, and Yimkhiung⁴ (a.k.a. Yimchungrü, Yimchunger), but is likely to include other languages such as Tikhir, Chirr, Longpfür and Makhuri that have hitherto been assumed to be dialects of the Yimkhiungrü community, in addition to their prestige Langa dialect. Sangtam is not mutually intelligible with any other language at this lowest level of subgrouping.

As noted in the introductory section, Northern Sangtam speakers live in twenty-two villages of the Longkhim-Chare sub-division. This sub-division is located in the western part of Tuensang District and borders the range known to the Ao as the Lambangkong, the main Chungli Ao-speaking area of Mokokchung District immediately west of the Dikhu River. Southern Sangtam speakers are found in villages of Kiphire District, which was partitioned from the southern region of Tuensang District in 2004. Kiphire District is also home to villages speaking Sumi and Yimkhiung. According to Hutton (1921: 10, 360, 364), at the time of his writing the Sema (i.e. Sumi) were expanding their territory in a north-easterly direction at the expense of the Yimkhiungrü and Sangtam tribes. This intrusion could account for the subsequent development of Northern and Southern dialects, as the Sumi expansion cleaved the traditional territory of the Sangtam tribe in two. To the best of my knowledge, no previous work has been done on the Southern Sangtam dialect, so it is not possible to say precisely how it differs from the Northern dialect described in this paper.

⁴ The current name of this tribe and language is reportedly based on the Kuthur Village pronunciation, which demonstrates a regular /c^(h)/ > /tʃ^(h)/ sound change that is not shared by most if not all other speakers of the Langa dialect. In a dictionary workshop recently convened in Shamator town, Tuensang, representatives of the literature committee expressed an intention to correct the spelling of their language's name to Yimkhiung, and the tribe's official name to Yimkhiungrü (-rü is the agentive nominalizing suffix common to languages of the Aoic group). I follow their preference here.

Figure 1. Districts of Nagaland, with the expanded inset demonstrating the location of Northern Sangtam villages in the Longkhim-Chare Sub-Division of Tuensang District (upper map sourced from: [Nagalandmap.png](#) by [Wikigringo](#), CC BY-SA 4.0, with modifications).

My primary consultant reports that a substantial number of Sangtam villages have populations that are bilingual in neighbouring languages: some located adjacent to Zunheboto District speak Sumi,⁵ some in Kiphire District speak Yimkhiung, and some

⁵ In 1923, Hutton ([1929] 1986) observed that Yazuthu village had a mixed population. Two-thirds of the village were Sangtam, and one third was Sema and lived in a separate ward of the village. He did not discuss their languages, but noted that it was a partly Sema village that followed Sangtam customs. This

villages of the Longkhim-Chare sub-division speak Chang or Chungli Ao, as already mentioned. According to anecdotal reports, famines, outbreaks of disease and intra-village feuds between clans sometimes motivated an entire clan to migrate to another village for succour. This might explain, for example, why there is a ward named Sangtemla in present-day Mokokchung town, which was previously a Mongsen Ao village known as Mokongtsü prior to the British annexation of the Ao territory in 1889. Kidnapping of women or the annexation of neighbouring tribes' lands and their populations also appear to have contributed to multilingual villages becoming established (e.g. see Coupe [2011: 23–24] for recorded textual evidence and related discussion, also Reid [1942: 116] for a report on Chang incursions into Ao tribal lands in the late 19th century). Multilingualism in Tibeto-Burman languages may be much more widespread in the villages of Nagaland than is currently appreciated, but is yet to be systematically investigated and documented.⁶ Such a sociolinguistic study would make a highly appropriate and valuable topic for a doctoral dissertation.

Sangtam is currently classified by Ethnologue as a Tibeto-Burman language belonging to a “Central Naga” branch of Tibeto-Burman (Eberhard, Simons & Fennig 2019). This is some improvement on Ethnologue's erroneous classification of Sangtam as a Bodo-Konyak-Jinghpaw (i.e. Sal) language in the preceding edition, but still confusing for anyone who is not a specialist in the field and unaware that “Naga” happens to be a fictitious branch of Sino-Tibetan.

The term “Naga” is now only appropriately used to refer to a loose ethno-political conglomeration of tribal people of the eastern border area who have patrilineal clan systems, previously followed animist belief systems but now overwhelmingly identify with a syncretic form of Christianity, practice subsistence swidden agriculture, eat a predominantly rice-based diet supplemented by hunting and gathering in the forests, and who support to a greater or lesser extent an ethno-nationalist political movement aiming to establish either outright secession from the Republic of India or a greater degree of autonomy for their perceived homeland. These people generally reside in villages located in the highlands of the Indo-Burmese Arc bounded by the Mishmi Hills of Arunachal Pradesh to the north, the Imphal Plain of Manipur to the south, and the Plain of Assam to the west. Recently this Naga cultural and political realm has been extended eastwards across the international border towards the Chindwin River valley and now incorporates other Tibeto-Burman-speaking communities of the Sagaing Region of Myanmar who share the same lifestyle, culture, and possibly political aspirations. Regrettably, Ethnologue mistakenly classifies these languages of the Sagaing Region as belonging to a “Naga” branch as well, and the classification of Glottolog naïvely follows suit, even labelling languages with colonial-era terms such

village is located in present-day Zunheboto District, a primarily Sumi-speaking region of south-central Nagaland.

⁶ Many people of Nagaland also speak Nagamese, a creole-like language with an Indic base that serves as a lingua franca for speakers of mutually unintelligible languages. Nagamese is a regional variety of Assamese that was initially used in trade exchanges between the hills and the plains before eventually spreading to the interior of the erstwhile Naga Hills District and beyond (Coupe 2007b). It is now well established as the first language of a sizeable community of Kachari speakers residing in Dimapur, with some families having adopted it as their native tongue for three generations. Many children of mixed marriages in cities such Dimapur and Kohima also acquire Nagamese as their mother tongue.

as “Ao Naga”, “Lotha Naga”, “Khezha Naga” and the like.⁷ This ignores the considerable progress that Tibeto-Burmanists have made in historical-comparative research over the last century and only serves to compound the current confusion.

From a linguistic perspective, “Naga” lacks credibility as a classificatory linguistic label because it clumps together groups who speak Tibeto-Burman languages that have long been known to belong to at least two and possibly even more branches of Sino-Tibetan (Burling 2003, Coupe 2012, Coupe 2017). The continued use of “Naga” as an anachronistic legacy of the colonial 19th century only serves to promote the historical fallacy that all the Tibeto-Burman speakers residing in the above-mentioned region speak a language belonging to a single branch of Sino-Tibetan. This is justification for replacing “Naga” with a more neutral geographic term, one that permits the field to refer to the non-Sal languages of central and southern Nagaland in a way that does not perpetuate unsubstantiated guesses about phylogenetic relationships based on cultural similarities. ‘Indo-Burmic’ has been proposed, adapted from the geographical name of the mountain ranges in which these languages are spoken (Coupe 2017). This label is more granular than ‘Kamarupan’ (Matisoff 1991) and hopefully will also prove to be a less caustic term for the speech communities it classifies, as well as for the academic linguistic community (see Burling 1999 and Matisoff 1999 for opposing thoughts on nomenclature).

As frustrating as the delay may be, it is still premature to propose internal branch affiliations for these Indo-Burmic languages of the northeastern border region; this is because the field does not yet have sufficiently comprehensive and reliable data to make credible proposals established by isoglosses, phonological reconstructions, or evidence of shared innovations. Credible subgrouping criteria will only be established with some degree of confidence after more work based on 21st century methods of linguistic description is done. Together with phonological reconstructions, shared sound changes and evidence of innovations, gradually scholars will identify a valid set of criteria by which genetically-related subgroups within Tibeto-Burman might be further established, if this cannot be achieved by lexical and phonological comparisons alone.

Like other languages of the Aoic and Angami-Pochuri sub-groups of central and southern Nagaland (e.g. Coupe 2012), Sangtam once had an innovative overcounting numeral system. An overcounting system expresses a value in relation to a higher parameter known as the augend, then calculates the number of units counted towards the augend. As in the Ao dialects,⁸ the cardinal numerals of Sangtam once followed a decimal system from *khātù* ONE to *t^hàrè màṇà* FIFTEEN, but switched to overcounting from SIXTEEN to NINETEEN. According to my 66-year-old Sangtam consultant, the system reverted to a decimal pattern of counting from TWENTY to TWENTY-FIVE, after which the overcounting pattern resumed again for the sixth to the ninth unit. This alternating system of decimal and overcounting patterns continued into the higher

⁷ Hammarström, Harald & Forkel, Robert & Haspelmath, Martin. 2019. *Glottolog 3.4*. Jena: Max Planck Institute for the Science of Human History. (Available online at <http://glottolog.org>, Accessed on 2019-06-13).

⁸ The only exception is the Ao dialect spoken in Yacham-Tengsa village, which developed a vigesimal system under intensive contact with Phom (Coupe 2012: 208)

decades *ad infinitum*, suggesting an alignment with other languages of central and southern Nagaland.

The following examples supplement the data previously published in Coupe 2012. The pattern of [AUGEND ‘NEG-complete’ UNIT TOWARDS AUGEND] demonstrated in (1) below is common to a number of languages of Nagaland that have documented evidence of overcounting numeral systems. Significantly for subgrouping purposes, the Northern Sangtam pattern is cognate with the overcounting numeral system previously documented in the Ao dialects, both in terms of the morphology used and in the overcounting pattern only applying from the sixth to the ninth cardinal numeral in each decade after TEN. This suggests that the innovative overcounting pattern was inherited from a common proto-language, and thus provides the most robust evidence to date of a genetic relationship between Northern Sangtam and the Ao dialects.

- (1) *m̀̀cà* *m̀̀pí* *tā́ók*
 twenty NEG-complete six⁹
 ‘Sixteen’ (literally ‘twenty not complete, six’, which should be understood as
 expressing ‘the sixth unit towards twenty’)¹⁰
m̀̀cà m̀̀pí t̃́ā́nék ‘17’
m̀̀cà m̀̀pí kék ‘18’
m̀̀cà m̀̀pí tā́kū ‘19’
m̀̀cà ‘20’

Another unusual feature of Sangtam cardinal numeral terminology is that the obsolete form for EIGHTY is *j̀̀l̀̀è̀̀è-à̀̀n̄* (‘forty-two’, understood literally as ‘forty twice’). This form might be profitably compared with the obsolete Longchang Mongsen form *lir̃̃ anekhi*, also understood literally as ‘forty twice’), as recorded by Mills (1926: 343), and the corresponding Chungli form *lir̃̃ anasy*, as documented by Clark ([1893] 2002: 45). It is noteworthy that the forms for EIGHTY in all three languages are semantically and structurally identical, involving a collocation of the numerals FORTY and TWO. Forms for EIGHTY based on the ‘forty twice’ pattern were not recorded in Lotha by Witter (1888) or Mills (1922), and no historical records are available to check for the same pattern in Yimkhiong, so the existence of similar forms for EIGHTY in other Aaic languages can now only be investigated by future enquiry with very elderly speakers.

The obsolete Sangtam expression for FORTY is also deserving of passing comment; this had the form *m̀̀cà-cà à̀̀n̄* and a structure suggestive of a reduplication (twenty-RDP-two). If so, then the abovementioned form for EIGHTY also appears to involve a

⁹ Abbreviations used in glosses conform to the Leipzig Glossing rules, with the following exceptions: AGT agentive case; ANMLZ agentive nominalizer; LNMLZ locative nominalizer; NMLZ1 nominalizing suffix form 1; NMLZ2 nominalizing suffix form 2; PFX nominal prefix; Available online at <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>, accessed 2019-12-24.

¹⁰ For the purpose of comparison, SIXTEEN was previously expressed by *m̃́kī m̃́-p̃n̄ t̃́ā́ūk* (twenty NEG-complete six) in Mongsen Ao. American Baptist missionaries replaced the overcounting pattern with a purely decimal system in the early 20th century to facilitate the teaching of arithmetic (see Coupe 2012 for discussion).

reduplication of the second syllable of *jɿɿè* FORTY, plus the numeral *əpɿ* ‘two’, representing multiplication by two.

4 Sources of data and presentation

The data upon which this phonological analysis is based were primarily provided by a native speaker (MS), who was sixty-six years of age at the time of recording. He was born and raised in Tronga [^to³³ŋa³³] village (a.k.a. Trongar, and accordingly pronounced by some as [^to³³ŋaɿ³³]), located in the Longkhim-Chare sub-division of western Tuensang District (see Figure 1). In his later life he worked in government service in various districts in Nagaland, including Kohima, Kiphire, Tuensang, Mon and Zunheboto. I would describe his idiolect as conservative, as it generally retains what appear to be archaic phonological features of Sangtam that are becoming lost from the phoneme inventories of speakers under fifty years of age. In addition to Sangtam he speaks Nagamese and English, he has some knowledge of the Thang dialect of Khiamniungan due to a posting in Noklak, (eastern Tuensang District), and he acquired good ability in Chungli Ao while attending college in Mokokchung District.

To supplement the data for an acoustic analysis of the tone system, I recorded six male speakers of Northern Sangtam in Tronga village, as noted in the introduction. These men were all lifelong residents of the village and were aged between twenty-eight and forty-seven years at the time of recording. A striking characteristic of their idiolects was the extent to which some of their phoneme inventories diverged from those of older speakers – see §10 for discussion and examples of the recorded differences.

In addition, the data of a 49-year-old male speaker of Sangtam from Alisopur village was recorded. This was solely for the purpose of investigating the phonetic features of the retroflex plosives and pre-stopped bilabial trills by means of static palatography, as presented and discussed in Coupe 2013 and 2015 respectively. Alisopur village is located on the same range as Tronga village, a little under four kilometres to the north-north-east as the crow flies. Based on the small amount of data of this single speaker, there appear to be minimal differences in the phonology of the Northern Sangtam varieties spoken in Alisopur and Tronga.

The data were recorded in Dimapur and Tronga village during three short field trips made between 2011 and 2013, closely following the methodology described in Coupe 2014. A reasonably well sound-attenuated recording studio was used for recording MS’s data in Dimapur, and a furnished room of a wooden house was the setting for recordings of the six speakers in Tronga village. The data were recorded using a Beyerdynamic TG H55c omnidirectional neck-mounted microphone coupled with a Fostex FR-2LE digital field recorder on the first visit in 2011, and the same microphone was used with a Marantz PMD661 Mk II digital field recorder in 2012 and 2013. The data were recorded at a sampling rate of 48kHz and a depth of 24 bits per sample. As I recorded hundreds of lexical items in these sessions and was concerned about boring my consultants to distraction, I limited the collection methodology to a single carrier sentence, which was usually preceded and followed by tokens of the target word:

- (2) _____ $\bar{i}$ $n\lambda$ _____ $t^{hl}ak-i\bar{e}$ _____
 1SG AGT write-PRS
 ‘I write _____’

The auditory tonal analysis was based on the tones realized on the sentence-internal token, as this was found to give a more reliable representation of pitch than tokens uttered in isolation preceding or following the carrier sentence, and the sentence-internal token was also less influenced by the potentially interfering effects of list intonation. Nevertheless, the words uttered in isolation before and after the carrier sentence were still useful for transcribing the segmental phonology, and for gaining an appreciation of intra-speaker variation in speech sounds.

The phonetic and phonemic representations of the data align closely but not perfectly with IPA transcription conventions (e.g. the subscript bridge _h distinguishing a dental place of articulation is dispensed with for typing convenience). Tones are represented by Chao tone numerals in phonetic transcriptions (Chao 1930), and by acute, macron and grave diacritics respectively for high, mid and low tonemes in the phonemic transcriptions. The same tone-marking conventions are used for the marking of tone in the data of other Aoic languages presented in this paper, most of which similarly have systems of three level tonemes according to previous work (e.g. see Coupe 2003 and Tamsunungsang 2009 for Ao dialects), as well as my own auditory impressions. The only exception is the Langa dialect of Yimkhiung, which appears to have just two contrastive tones in preliminary investigations.

5 Limitations of this study

Due to the logistical problems of arranging female speakers on my short visits to Nagaland, and the less-than-satisfactory quality of the data collected from some female speakers that I managed to record, the analysis is restricted to the data of male speakers only. Part of the difficulty in collecting representative data from speakers of both genders is cultural: tribal society in Nagaland is strongly patriarchal and paternalistic, and men accordingly dominate in areas of administration and decision-making. This paternalism extends to the membership of literature committees, and especially to considerations regarding who might be considered the best-qualified orators to share their language with a linguist. Despite the fact that none of the presented data come from the vocal tracts of women, my impression is that there are unlikely to be substantial differences between male and female speakers. Naturally this will require further research to verify.

Lastly, I have not attempted to deal with the tone sandhi that sometimes affects polysyllabic words, as this task is beyond the scope of the paper and requires additional investigation. Tonal transcription consequently represents the output tones.

6 Phonotactics

6.1 Syllable structure

The minimal syllable consists of a simple nucleus filled by one of six vowel phonemes, plus an associated obligatory tone. The onset C_1 is always simple and may be filled without restriction by any of the thirty-one consonant phonemes listed in the first column of Table 1 below.

Table 1. Northern Sangtam phonotactic distribution of phonemes

C_1	V	C_2	Tone
p p ^h t t ^h t ^l t ^{hl} c c ^h k k ^h	i u	p k	ó (High)
ts ts ^h tʃ tʃ ^h	e o	m ŋ	ō (Mid)
tʂ tʂ ^h	ʌ	(ɿ)	ò (Low)
m n ɲ ŋ	a		
nm			
f v s z ʃ x h			
l ɿ j			

(3) Syllable and word structure:

- σ = (C_1) V (C_2) + Tone
 C_1 = any C
V = any vowel
 C_2 = /p, k, m, ŋ (ɿ)/
Word = σ (σ)*

The coda position (C_2) is restricted to the plosives /p/ and /k/ and the nasals /m/ and /ŋ/, presenting a symmetry based on the place of articulation and captured as a natural class by the acoustic feature [+grave] (Jakobson & Halle 1956: 35).¹¹ Of these four possible coda consonants, the velar nasal overwhelmingly has the greatest frequency of occurrence. In the corpus of ca. 900 words, /ŋ/ accounted for approximately 250 instances of coda consonants (28% of total), followed by the glottal stop allophone of the velar plosive /k/ with approximately 150 instances (16%). The status of the velar nasal as a permissible coda consonant appears to be under threat, as all Northern Sangtam speakers demonstrate free variation between the overt realization of a velar nasal syllable-finally and zero realization with a nasalized vowel nucleus, e.g. [a³³sãŋ³³] ~ [a³³sã³³] ‘three’. This evolving sound change is the harbinger of nasalized vowels developing a phonemic contrast in the language. At present the velar nasal apocope appears to be limited almost exclusively to environments after the low central vowel /a/ for my main consultant, but high vowels are additionally nasalized in the data of younger speakers (see §9.2 and §10 for further discussion).

¹¹ There are only two phonetic realizations of a dental nasal coda in the entire corpus. One is [a¹¹xin¹tʂA³³] ‘bride’, which could possibly be segmented as *à-xìn-tʂA* PFX-be.fresh/raw-F. But since the citation form for ‘fresh’ is [a¹¹xin¹¹], the realization of the dental nasal coda is clearly a case of assimilation to the following dental place of articulation. The only other occurrence of a dental nasal coda is in the noun [san⁵⁵tʃi⁵⁵] ‘colour’, which is similarly accounted for by dental assimilation.

Table 2. Examples of Northern Sangtam words demonstrating syllable and word structures. Syllable boundaries in the third column are indicated by a period.

WORD	GLOSS	SYLLABLE STRUCTURE
ĩ	1SG	V
ʃλ	‘thread’	CV
tḅāŋ	‘needle’	CVC
ā-káp	‘RL-skin’	V.CVC
mλzĩ	‘claw/nail’	CV.CV
náktse	‘eye’	CVC.CV
ʃítái	‘be real’	CV.CV.V
kàmtʃĩŋ	‘chin’	CVC.CVC
mλtʃλmλp ^h ĩ	‘spittle’	CV.CV.CV.CV
tsúŋnáktseúfĩ	‘in-law’	CV.CVC.CV.CV

The most restricted coda consonants are the bilabials /p/ and /m/. The voiceless unaspirated bilabial plosive occurs as a coda in just 15 lexical roots (approx. 2% of total), and the bilabial nasal occurs only a little more frequently as a coda, accounting for just two dozen or so instances in the corpus (approx. 3%). Overall, CVC syllable structures are less common than CV structures, an observation consistent with the rather tight restrictions on possible coda consonants.

The voiceless unaspirated velar plosive occurs relatively frequently as a syllable onset but is not attested as a coda consonant. It is observed to be in complementary distribution with the glottal stop, which is conversely restricted to the coda position. Accordingly, [k] and [ʔ] are analysed as allophones of a single phoneme /k/. This analysis is supported by cross-linguistic evidence from other languages of the Aoic group demonstrating that Northern Sangtam [ʔ] uniformly corresponds with /k/ in cognate Mongsen Ao, Yimkhiung and Lotha words. Examples of the correspondences are presented below in Table 3. Note that the velar-to-glottal sound change occurs with a variety of vowel nuclei. This suggests that it has applied globally to every instance of a velar plosive coda, rather than to a particular rhyme of the syllable.¹²

¹² This description revises and updates a preliminary analysis of the glottal stop as having phonemic status in Coupe 2015.

Table 3. Correspondences with Northern Sangtam [ʔ] in selected Aoic languages.¹³ Nominal affixes are segmented from their roots, where applicable.

	N. Sangtam	Mongsen Ao	Yimkhiong	Lotha	PTB
‘black’	à-ná[ʔ]	tā-nāk	ā-māiāk	ē-jēkōʔ	*nak
‘breath’	sá[ʔ]	tā-sākā	ʃāk	ē-t ^h āk	*sak
‘dao’ ¹⁴	nó[ʔ]	ā-nūk	nūk	lipōk	---
‘eye’	ná[ʔ]tsē	tā-nik	māk	ð-jīk	*s-mik ≠ *s-myak
‘knee’	mā-k ^h ú[ʔ]	tā-mākūk	māk ^h ūk	ṇk ^h ōk	*(m-)ku-k
‘louse’	ā-xá[ʔ]	ā-ts ^h āk	kūjāk	ð-jōk	*sar~śar, s-rik
‘six’	t ^h āró[ʔ]	tā-ūk	t ^h ērōk	tirōk	*ruk

Borrowed words also appear to be affected by the prohibition on a velar plosive occurring as a coda consonant, as the Hindi loan word बटर [bətək^h] ‘duck’ is pronounced as [pa¹¹taʔ¹¹] in Northern Sangtam. That it is a relatively recent loan word perhaps also accounts for the observation that its final syllable carries an anomalous low tone. It is otherwise the case that Northern Sangtam syllables terminating in a glottal stop coda overwhelmingly carry a high tone, and such syllables also frequently manifest phonetic creaky voice phonation on the preceding vowel (see §7 for related discussion of the tone system). There are very few exceptions of glottal syllables not carrying a high tone, the only other examples in the corpus being [ɹe³³juŋ⁵⁵ɹeʔ¹¹] ‘lizard’, [mi⁵⁵tʃaʔ¹¹siŋ³³] ‘matchbox pine’ [a¹¹zi⁵⁵caʔ¹¹] ‘be small’ and [ʃi³³taʔ¹¹] ‘be correct’. The last example is possibly a borrowing from Ao [ʃi³³tak³³], and the realization of a low tone on glottal syllables in the trisyllabic words may be due to tone sandhi changes.

The central approximant /ɹ/ is a highly marginal coda consonant that does not occur in the C₂ position in any recorded native lexical roots. It is only observed to function as a coda in genitival [N-GEN N]_{NP} constructions. Under these circumstances, /ɹ/ is syllabified as a coda, e.g. /V.CVC CVC.CV/ in àzā-ɹ pām-tʃi (‘child-GEN sit-LNMZ’, literally ‘child’s sitting place’) ‘womb’.¹⁵ It also occurs infrequently as a coda in some loan words, e.g. hātʃāi ‘thousand’ (most probably < Hindi हज़ार [həza:ɾ] ‘thousand’). For reasons presently not understood, there is a dispreference for coronal phonemes of any manner of articulation to fill the coda slot. This restriction is partially reflected in

¹³ The glottal stop coda of the Northern Sangtam data is presented in phonetic transcription to demonstrate its realization as an allophone of /k/ in Table 3. All other data are presented in a phonemic transcription. The Yimkhiong (Langa dialect) and Lotha (Elumyo village variety) data presented in the tables of this section come from recordings and field notes currently undergoing preparation as phonological descriptions. The source of the Mongsen Ao data is Coupe 2007, and the reconstructed PTB forms are from Benedict 1972 and Matisoff 2003.

¹⁴ This is the Nagamese word for a type of hatchet with its tang enclosed in a handle, traditionally carried and used by men of the region as a cutting tool and weapon.

¹⁵ The rhotic may also occur as a coda when a related morpheme is used as a converb suffix or an agentive nominalizing suffix on a verb stem.

the closely related Lotha language, which similarly bans dental plosive codas, but permits a dental nasal to fill this slot.

An alternative approach to the analysis of the palatal plosives /c c^h/ and the palatal nasal /ɲ/ could identify these segments as complex onsets consisting of a sequence of a plosive plus a palatal glide /j/. This is less convincing, however, because if putative /kj, k^hj/ and /ɲj/ are permissible onset clusters, then one might reasonably expect to find palatalized /pj, p^hj/ and /tj, t^hj/ functioning symmetrically as complex onsets as well. However, such sequences are unattested in the sizeable corpus of ca. 900 words. Also, we find comparative evidence of a plausible sound change triggered by an adjacent high front vowel in at least some words containing a palatal plosive. A convincing case for this can be made by comparing Northern Sangtam *m̄l̄cà* ‘twenty’ with reflexes in Mongsen Ao and Lotha – *m̄k̄ī* and *m̄kw̄ī* respectively – both of which contain a syllable with a velar plosive onset and a high front vowel nucleus. Similarly, the roots of Northern Sangtam *cá* ‘plait’, *cá* ‘shoot’ and *c^hà* ‘wear’ can be compared with the roots of Yimkhiung *k̄j̄àk* ‘plait’, *k̄īp* ‘shoot’ and *k^hīm̄ī?* ‘wear’.

It is therefore feasible that the palatal stop in the Northern Sangtam reflexes resulted from a transphonologization of the palatal feature of a high front vowel to an adjacent velar segment, motivating the fronting of /k/ to /c/ in some words.¹⁶ This does not explain every occurrence of a palatal plosive though, as both the aspirated and unaspirated forms also occur in words whose cognates in other Aoic languages contain affricates, rather than velar plosives (e.g. see Table 4). But these too may have resulted from older cycles of sound changes that will require more comparative reconstruction to understand fully.

Table 4. Correspondences with Northern Sangtam /c, c^h/ in selected Aoic languages (nominalizing and tense marking affixes are segmented from their roots, where applicable)

	N. Sangtam	Mongsen Ao	Yimkhiung	Lotha	PTB
‘plait’ (v.)	cá-pā	m̄ts̄s̄	k̄j̄àk	p̄it̄f̄ō?	---
‘twenty’	m̄l̄cà	m̄k̄ī	m̄ūkū	m̄kw̄ī	*(m-)kul
‘shoot’	cá-pà	k̄əp	ā-k̄īp-k ^h ī	k̄ūwā?	*ga·p
‘wear’	c ^h à-rē	t̄ ^h ə̄m, ts̄ ^h ə̄ŋ	k ^h īm̄ī?	h̄à̀nà?	*gwa ⋈ *kwa

In the same vein, an alternative synchronic analysis of syllable structure might propose that the segments /t̄, t̄^h/ be interpreted as complex CC onsets consisting of a retroflex plosive and a central approximant, viz. /ɭ, t̄^hɭ/. Again, this is less convincing, because there are no examples of /pɭ, p^hɭ/ or /kɭ, k^hɭ/ complex onsets in the corpus, which would otherwise be suspicious omissions. Nonetheless, it seems plausible that the initial plosive of an older proto-cluster (probably *kɭ) underwent an anticipatory assimilation to the coronal place of articulation of the rhotic, and the consonant cluster subsequently phonemicized as a single segment. This could logically account for why /t̄, t̄^h/ are

¹⁶ Matisoff (1973: 79) describes a similar sound change affecting the PTB rhyme *-ik that resulted in Burmese developing a palatal fricative coda, triggered by the palatal (i.e. [+high]) feature transferring from the vowel to the consonant.

consistently pronounced with a rhoticized release. Such an explanation is also consistent with Matisoff’s (2003: 21) proposal that retroflexes in Tibeto-Burman languages are secondarily derived from proto-clusters with medial liquids. He notes that in some Chin languages, the *labial/*velar contrast is neutralized before a medial *-*r-* in favour of dental + rhotic clusters, thus yielding /tr, tʰr/, and that these are sometimes represented as retroflex stops /ɽ, ɽʰ/ in transcriptions (2003: 74–75).

Table 5. Correspondences with Northern Sangtam /tʰ/ in selected Aoic languages (nominalizing affixes and tense-marking suffixes are segmented, where applicable)

	N. Sangtam	Mongsen Ao	Yimkhiung	Lotha	PTB
‘weep’	tʰák-tʰō	tʰəp	tʰip	tīaʔ-tʰjōʔ	*krap
‘fear’	tʰà-pà	tsəpʰā	ā-tʰē	ē-kʰjān	*grāk~krāk; grok~ krok *kri(y)
‘rust’	tʰòŋtʰáíó-pà	ā-tshəŋ	tʰáŋpín-kʰì	sāsə	---

In light of the Chin evidence, and judging from the handful of words that appear to have cognate forms across the Aoic languages (see Table 5), it is likely that the rhotacized retroflex plosive did indeed originate from such a cluster in the intermediate proto-language, as the retroflex feature has been retained in Yimkhiung as well.

Lastly, the doubly articulated pre-stopped bilabial trills are a third set of segments that could possibly be treated as clusters in a synchronic analysis. Again, this is less convincing, because bilabial trills do not occur independently as simple onsets. It would also be extremely unusual from a cross-linguistic perspective for such a manner of articulation to occur as a C₂ constituent in a cluster. Moreover, an acoustic study concluded that the pre-stopped bilabial trills are articulated simultaneously, rather than as independent segments in a C₁C₂ cluster (Coupe 2015). They also demonstrate an aspiration contrast, as do all the simple plosives, and they occur before a variety of vowels in onsets both word-initially and word-medially, so there is no reason to treat them differently from other non-continuant phonemes that occur as simple onsets in the C₁ position.

Unlike the Chapakuran languages Wari’ (Everett & Kern 1997) and Oro Win (Ladefoged & Maddieson 1996) – the only other languages ever reported to have these typologically unique consonants – the pre-stopped bilabial trills of Sangtam are not constrained to only being realized before a high front rounded or back rounded vowel, and they are not analyzable as allophones of another phoneme.¹⁷ This gives Northern Sangtam the status of possessing the rarest consonant phoneme in the world, and it also underlines the importance of documenting minority languages in order to gain an

¹⁷ Confusingly, Everett & Kern (1997: 384) list $\widehat{t\beta}$ as a ‘distinctive segment’ of Wari’, but then describe it as an allophone of /t/ on the following page. They note that it sometimes occurs in variation with [t]. Other publications that discuss this consonant accord it phonemic status (e.g. Everett & Ladefoged 1996; Maceachern, Kern & Ladefoged 1997), but this is not overly convincing if it only occurs in the environment before back rounded vowels and is limited to just twenty-five words of an entire corpus. Everett & Kern (1997: 385) report that $\widehat{t\beta}$ is almost exclusively found in the phonological inventories of older speakers, and that it is evolving into [t]. A noteworthy correlation is that the same fate is currently befalling the pre-stopped bilabial trill of Northern Sangtam – see §8.3 and §10 for further discussion.

informed appreciation of the extent of the planet’s phonological diversity. The unique articulatory characteristics of this consonant are discussed in greater detail in §8.3.

6.2 Vowel-vowel sequences

There are just three lexical roots containing segmental sequences of what could be analyzed as Vowel plus Glide, viz. [moj⁵⁵lɛm⁵⁵] ‘hawk’, [ʃi⁵⁵taj¹¹] ‘be real’, and [hoʔ⁵³jaj¹¹] ‘hang’. Alternatively, these could be treated as Vowel-Vowel sequences, e.g. [mo⁵⁵i⁵⁵lɛm⁵⁵], [ʃi⁵⁵ta¹¹i¹¹], and [hoʔ⁵³ja¹¹i¹¹]. It is not economically justifiable to accommodate a post-nuclear glide segment in the syllable canon on the basis of just these three words, therefore it is expedient to treat the glides phonemically as constituting the nuclei of separate syllables. These are accordingly represented phonemically as *móílɛm*, *ʃítàì* and *hókjàì* respectively. The validity of analyzing VV sequences as constituting the nuclei of separate syllables is most convincingly demonstrated when each vowel of the vocalic sequence carries a different tone, e.g. as in *àmλsāú kànpà* ‘grow worse’, *à-ū* ‘PFX-father’, and *múí* ‘drug’.

6.3 Rhyme patterns

Table 6 revises and updates the description of Sangtam rhymes presented in Bruhn (2014: 165), which was based on secondary data sourced from Marrison 1967, Kumar 1973 and Weidert 1987.

Table 6. Northern Sangtam rhymes

coda	-Ø	-p	-m	-ŋ	-k (= [ʔ])
nucleus					
i	✓	–	–	✓	✓ (rare)
e	✓	✓ (rare)	–	✓ (rare)	✓
u	✓	–	–	✓	✓
o	✓	–	–	✓	✓
ʌ	✓	✓	✓	✓	✓ (rare)
a	✓	✓ (rare)	–	✓	✓

As noted in §6.1, the bilabials /p, m/ have rigid constraints on their realization as coda consonants. The additional data provided by the Northern Sangtam word list confirm Bruhn’s observation (2014: 164) that bilabial codas can only occur with a central vowel nucleus.¹⁸ The sources of the Sangtam data he relies upon represent that central vowel as either schwa (*ü* in Marrison 1967 and Kumar 1973) or a high back unrounded vowel (*u* in Weidert 1987). My own auditory impressions, corroborated by a preliminary calculation of formant values, suggest that the actual articulatory position of this vowel is a little lower and further back in the vowel space than schwa [ə], although schwa does indeed occur marginally as an allophone in free variation in the environment after a palatal consonant, which exerts a slight centralizing and raising effect. In phonologically neutral environments – e.g. word-finally after a bilabial – its articulatory position is closer to Cardinal Vowel 14, viz. [ʌ]. As [ʌ] has the widest distribution and

¹⁸ The single exception to this is *jép*, the root of ‘sleep’. The vowel of this syllable may be subject to fronting and raising in the environment after a palatal approximant.

a default realization that is not conditioned by any particular phonetic environment, the symbol /ʌ/ is accordingly selected to represent the central-back vowel phoneme.

6.4 Word structure

Northern Sangtam lexical roots range in structure from single syllables, e.g. *xù* ‘bamboo’, *ŋū* ‘fish’, to disyllables, e.g. *ɬòmī* ‘face’ and *ʃòŋp^hók* ‘mushroom’. The majority of lexical roots are disyllabic.

A key feature of word structure that characterizes all Indo-Burmic languages is the use of relational and non-relational nominal prefixes on many semantic classes of bound nominal roots (e.g. see Marrison 1967 Vol. I: 110). These terms were first coined to account for the distribution of two types of nominal prefix in Mongsen Ao: the relational form *tə-*, occurring on the bound roots of kinship and body part terms, and the non-relational form *a-*, occurring on the bound roots of terms for cultural artefacts and objects of the biosphere (Coupe 2007: 246–248). Other scholars working on Indo-Burmic languages (e.g. Temsunungsang 2009, Bouchery & Sangtam 2012, Teo 2014, Bruhn 2014) have since adopted these terms to describe nominal morphology with a similar distribution and function, but non-cognate forms.

Matisoff (1989) proposed the term ‘sesquisyllabic’ to account for the syllable-and-a-half structure that he encountered in many languages of Southeast Asia. While the (C)ə.CV(C) structure is certainly a common pattern, it turns out that the minor syllable with its often phonologically-reduced central vowel is by no means restricted to monosyllabic roots in most of the Indo-Burmic languages (e.g. for Mongsen Ao examples and discussion, see Coupe 2003: 21–24), and such prefixes are more generally found on many disyllabic roots in these languages as well. However, many of these disyllabic roots are also likely to have been created via the accretions of older prefixes, which in turn have their ancient source in the (C)ə.CV(C) structure recognized by Matisoff. It is thus possible to discern diachronic layers in the formation of words in these languages.

Wolfenden (1929: 12) was the first to observe that a related erstwhile prefix occurs in some Ao and Lotha nominal stems, the latter language manifesting a dental nasal variant ‘before dentals, gutturals and palatals, retaining m- only before labials’, e.g. Lotha *ñ-lā* ‘navel’, *ñ-k^hók* ‘knee’, *ñ-tʃō* ‘penis’ and *mē-mfā* ‘lip’. In Mongsen Ao cognates, a related form has been reinterpreted as part of the root and is now preceded by a newer relational prefix with the form *tə-*, e.g. *tə-məzəŋ* ‘RL-claw/nail’, *tə-məts^hə* ‘RL-bud’. Preliminary work on Yimkhiung suggests that a prefix with a labial nasal onset is still preserved as such in some relational nouns, e.g. *mə-k^həʔ* ‘chin’, *mə-zān* ‘claw/nail’ and *mə-lūŋ* ‘heart’.

Synchronically, Northern Sangtam has two extant nominal prefixes: *a-* and *mʌ-*. The *a-* form is more common and is found with either a low or mid tone on some but not all noun roots denoting body parts and related entities, e.g. *ā-sá* ‘bile’, *à-k^hālīŋ* ‘kidney’, *à-là* ‘root’, somewhat more consistently on the roots of kinship terms and relational nouns, e.g. *à-kù* ‘mother’s brother/father-in-law’, *ā-fā* ‘older sister’, *à-lūŋlā* ‘centre’ and *à-pókki* ‘below’, and a few numerals, e.g. *ā-nā* ‘two’. The *mʌ-* form also occurs on some body part terms and words expressing part-whole relationships, and on

the roots of just two numerals – *mλ-jλ* ‘four’ and *mλ-cà* ‘twenty’. It variously carries a high, mid or low tone that does not appear to harmonize obligatorily with the tone of an adjacent root syllable, e.g. *mλ-lē* ‘navel’, *mλ-p^hā* ‘calf of leg’, *mλ-zī* ‘claw/nail’ and *mλ-lōŋtʃ^hī* ‘bud’. These are the only semantic classes of nominal roots that *mλ*-prefixes. It is almost certainly related to Wolfenden’s nasal prefix discussed in the preceding paragraph, as a segmentally similar syllable occurs in many cognate words in the Aoic languages.

The neat partition of semantic classes of bound nouns that occur with relational and non-relational prefixes in Mongsen Ao is not reflected in Northern Sangtam. For example, the nouns *k^hék* ‘hand’, *t^hák* ‘head’ and *fā* ‘body (among many others)’ were freely elicited without a prefix being uttered. This suggests that not all body part terms are bound roots, or that the prefix is an optional affix in some words, but this will necessarily require future investigation of the morphology to confirm the status of body-part terms as bound or free morphemes.

There is also a separate nominalizing prefix *à-*, which derives verbal nouns from verb stems, e.g. *à-xλm* ‘odour’ (< *xλm-pà* smell-NMLZ1 ‘to smell’). As the nominal relational prefix *a-* and the nominalizing prefix *a-* have different distributions and functions, they must necessarily be recognized as distinct prefixes in a synchronic description. The same observation applies to the paired nominal relational prefix *tə-* of Mongsen Ao and the nominalizing prefix *tə-* (e.g. see Coupe 2007: 247–248), which correlate in respectively having identical functions to those of Northern Sangtam, respectively sharing the same segmental form, and also demonstrating diversity in their tonal realizations. The observed isomorphism of relational and nominalizing prefixes in a number of Indo-Burmic languages is an interesting correlation worthy of further investigation. It seems likely that these pairs of prefixes respectively share the same historical origin in each of these languages, although they are usually not cognate across the languages of this grouping. Possibly drift/parallel grammaticalization accounts for their similarity in function, but cross-linguistic differences in form (e.g. Sapir 1921, LaPolla 1994).

7 Suprasegmental phonology

Sangtam has three contrastive level tones occurring on open (unchecked) syllables, and a single high tone on checked (glottal) syllables. The following sub-minimal sets of words demonstrate the phonemic contrasts on open-syllable tonemes. These maintain a relatively stable pitch height over their rhymes when uttered within carrier sentences and also share a similar duration, suggesting that the pitch level at the beginning of the rhyme is the primary cue for their differentiation.

(4) <i>làn</i>	‘fathom’	(5) <i>jà</i>	‘sun’
<i>lān</i>	‘road’	<i>ā-jā</i>	‘PFX-two’
<i>lāntít^hùŋ-lā</i>	‘clan name-ANMLZ’	<i>jà</i>	‘snot’

In addition, a high tone is very consistently found on syllables that terminate in the glottal stop allophone of /k/, e.g. *lák-tʃ^hò* [lɑʔ⁵³tʃ^ho¹¹] ‘beat-NMLZ2 and *t^hók* [t^hoʔ⁵³]

‘brain’, as discussed in §6.1. The high tone of such checked syllables has an almost identical pitch height to the high tone of open syllables at the onset of the rhyme and shares a similar duration, but falls away slightly towards the end of its trajectory. Because of a corresponding drop in intensity, however, the fall in pitch is audibly so slight that it is almost imperceptible, so that it gives the impression of being a high-level checked tone. A comparison of the tones of cognates in related languages demonstrates that the high checked tone of Sangtam corresponds with a mid-tone in other Aoic languages (e.g. see Table 3), but this preliminary finding requires further investigation.

The high checked toneme presents a convincing case of debuccalization, as first described by Haudricourt (1954) in his brilliant paper on tonogenesis in Vietnamese. This represents the outcome of an oral phoneme /k/ evolving a laryngeal allophone [ʔ] when realized as a coda consonant. The closure of the glottis involves considerably increased tension on the vocal folds as a result of the vocalis, cricothyroid and lateral cricoarytenoid muscles contracting, and this tension surely contributes to raising the pitch of a speaker’s voice just prior to full closure of the glottis. Such an explanation conceivably accounts for the consistent realization of the [53] pitch associated with glottalized syllables (as well as its slight terminal fall as the laryngeal musculature relaxes), and is consistent with Matisoff’s (1973) observations of tonogenesis in Southeast Asian languages. It also suggests that Northern Sangtam’s high stopped allotone has independently developed language-internally from segmental features, rather than via inheritance from an intermediate proto-language or through language contact and diffusion. The phonotactic behaviour of the glottal stop of Sangtam diverges significantly from the Ao dialects and Lotha, since in these related languages it functions synchronically as a word prosody and is obligatorily deleted before a morpheme boundary. This restricts the glottal stop in some Aoic languages to the word-final position (e.g. see Coupe 2003: 24–27 for discussion of Mongsen Ao).

A related phenomenon of relevance is that glottalized syllables are often pronounced with some degree of creaky voice phonation. In view of this, conceivably we could analyze creaky voice as an additional allophonic realization of /k/ in this environment and consider it to be an associated consequence of the debuccalization process. Word-internal glottal stops that are represented in the current Sangtam orthography by <h> are often not pronounced; e.g. my main consultant writes *lásá-tʰò* ‘defy-NMLZ2’ as <lahsahcho>, but spectrograms of this and other words with orthographic <h> as a coda often reveal insignificant jitter and no evidence of glottal closure at all, so Northern Sangtam also appears to be shifting towards syncopating its glottal stop allophones in the environment before a syllable boundary. The end result of this process is that erstwhile high-toned checked syllables are evolving to high-toned open syllables when occurring word-internally.

As noted in §6.1, dental plosives do not occur as codas, and the rarity of rhymes with a bilabial stop coda renders it difficult to propose the possible range of tonal contrasts that can occur on other closed (i.e. checked) syllables that do not terminate in a glottal stop. The very few examples found in the data have either high tone, e.g. *máp-ɛ̃* ‘beat/strike with weapon-PRS’, or mid tone, e.g. *pāp* ‘cage’, *kālāp-tʰò* ‘bake-NMLZ2’. At present it is not possible to say if the lack of examples of low-toned

rhymes with bilabial plosive codas is attributable to gaps in the data, or due to a phonotactic constraint on the tonal system.

Three-register toneme systems appear to be common to most Aoic languages and are reported for Mongsen Ao (Coupe 2003) and Chungli Ao (Temsunungsang 2009). Preliminary auditory impressions suggest that Lotha has low, mid and high terrace-like tonemes, but the Langa dialect of Yimkhiung diverges in only appearing to have two contrastive level tones. Boundary-marking intonation patterns at the phrasal and clausal level are prominent in narrative texts and can override the lexical tones of individual syllables in all of these languages, just as they do in Mongsen Ao (Coupe 2007: 73–77). The most common presentation is a rising intonation coinciding with non-finite clause margins, but a non-final rising intonation also occurs at the boundaries of finite clauses when a speaker wishes to indicate the continuity of a topic via prosodic means. This use of non-final intonation at the boundaries of clauses appears to be a shared areal characteristic of all Aoic and Sal languages of central Nagaland.

8 Consonant phoneme inventory

The following subsections describe in detail the segmental phonemes of the consonant inventory according to their various manners of articulation and provide examples of their realizations in words. From a typological perspective (e.g. Maddieson 2013), the consonant phoneme inventory of Northern Sangtam is moderately large in possessing thirty-one members; it also has some unexpected segments, such as the retroflex and palatal stops, and some extremely rare doubly-articulated segments. The language is additionally distinguished both from an areal and global perspective in having a surprisingly large number of fricative phonemes.

The consonant phonemes of Northern Sangtam are listed according to their place and manner of articulation in Table 7 below.

Table 7. Northern Sangtam consonant phoneme inventory

	Bilabial/ Lab. dent	Dental	Retroflex	Palatal/ Pal. alv.	Velar	Glottal
Plosive	p	t	tʰ	c	k	
	pʰ	tʰ	tʰʰ	cʰ	kʰ	
Affricate		ts		tʃ		
		tsʰ		tʃʰ		
Doubly-articulated pre-stopped bilabial trill	tʰb					
	tʰbʰ					
Doubly-articulated nasal	nm					
Plain nasal	m	n		ɲ	ŋ	
Fricative	(f)	s		ʃ	x	h
	v	z				
Approximant		l	ɭ	j		

The consonant phoneme inventory demonstrates six distinct places of articulation and seven manners. A basic opposition of voiceless aspirated versus voiceless unaspirated segments characterizes the voice onset time (VOT) of oral plosives, affricates and doubly articulated pre-stopped bilabial trills, and a voiced~voiceless opposition is exhibited by the labiodental and dental fricatives. A two-way VOT opposition in plosives is a uniformly widespread feature of the Aoic languages, as well as the Sal languages of central Nagaland (viz. Chang, Khiamniungan, and Phom). The only exception is the Chungli dialect of Ao, which lacks the common two-way VOT contrast in syllable-initial plosives and affricates.¹⁹

8.1 Plosives

The plosive series of Sangtam diverges from the better-known Ao dialects in having pairs of voiceless aspirated and unaspirated retroflex plosives /tʰ, tʰʰ/, as well as a pair of palatal plosives /c, cʰ/, in addition to the expected bilabial, dental and velar phonemes of this series.

The voiceless unaspirated bilabial plosive /p/ is potentially realized as a voiced allophone [b] when flanked by voiced segments word-medially, e.g. *nápōŋ* [na⁵⁵boŋ³³]

¹⁹ A two-way voice onset time contrast extends to the sonorant series of Lotha, and also characterizes some but not all village varieties of Mongsen Ao (Coupe 2007: 29. While a two-way contrast in plosives is consistently found in the Sal languages of central Nagaland, not all demonstrate a VOT contrast for their affricates (e.g. Chang, Phom). The affricate series of the Patsho and Thang dialects of Khiamniungan diverge from other members of the Konyak group by sharing the same VOT contrasts as their plosive series.

‘nose’ and *jópāpā* [jo⁵⁵ba³³ba¹¹] ‘put-NMLZ1’. The voiceless unaspirated bilabial plosive is the only member of this series to have a voiced allophone, plausibly due to aerodynamic factors: the anatomical composition of the oral cavity permits simultaneous glottal pulsation and accommodation of air behind the bilabial occlusion to a considerably greater extent than is possible for more posterior places of articulation. Voiceless unaspirated bilabial plosives occur as syllable onsets word-initially and word-medially, and rather rarely as codas after a central vowel only, e.g. *pék-tf^hō* ‘cast.away-NMLZ2’, *mípùŋ* ‘abdomen’, *kālāp-tf^hō* ‘bake-NMLZ2’, and *k^hénāp* ‘cubit’. The single exception to this generalization occurs in *jép-pā* ‘sleep-NMLZ1’.

Like the other aspirated plosive segments, the voiceless aspirated bilabial plosive /p^h/ is restricted to the onset position, e.g. *p^hòkp^hú* ‘owl’ and *láp^hā* ‘blemish’. This is a widely observed phonotactic constraint on aspirated obstruents in the languages of mainland Southeast Asia. Northern Sangtam appears to be at the end-stage of developing a voiceless labial-dental allophone [f], which is only realized in the environment before a mid-central vowel [ʌ]. However, [p^h] continues to occur as an onset before [ʌ] in a very limited number of words – the *Yusüp: Bilingual Dictionary of Sangtam-Naga* has just four examples of <phü> (equivalent to [p^hʌ]) occurring in lexical roots – and the only examples in the corpus are *íp^hā-pā* ‘pace-NMLZ1’ and the village name of *p^hālōŋlā*. The dearth of examples thus warrants according the voiceless labial-dental fricative only marginal status as a phoneme.

The dental plosives /t, t^h/ are articulated with the tongue tip touching the upper teeth, an observation confirmed by static palatography (Coupe 2015). Both dental plosives occur as onsets word-initially and more rarely word-medially, e.g. *tākū* ‘nine’, *jītā-tf^hō* ‘refuse-NMLZ2’, *t^hàlè* ‘ten’ and *māl-t^hé-pā* ‘send-NMLZ1’, but never as codas.

The retroflex plosives /ṭ, ṭ^h/ have a characteristically rhotacized release. In the articulation of these consonants, the sub-apical aspect of the retracted tongue tip and the sides of the tongue form a complete occlusion at the postalveolar place of articulation and the gingival margins of the molar teeth. The retroflexed configuration of the tongue is maintained immediately following the release of the oral occlusion, and this is responsible for the rhotic quality that accompanies the release. The retroflex plosives occur as onsets word-initially and word-medially only, e.g. *ṭ^hḷsú-tf^hō* ‘be frightened-NMLZ2’, *áṭ^hōŋ* ‘many, much’, *ṭ^hḷŋxāŋ* ‘rib’, and *māl-ṭ^hāŋ* ‘fog/cloud’.

As mentioned in §6.1, the palatal plosives /c, c^h/ are likely to have resulted from the transphonologization of a [+high] feature of an adjacent segment to an erstwhile velar plosive. They pattern like the other plosives in having a two-way VOT contrast and word-initial and word medial distributions, e.g. *cā-tf^hō* ‘cut-NMLZ2’, *à-cā* ‘PFX-egg’, *c^hā-pā* ‘lend-NMLZ1’, *āts^hāk pìc^hāk* ‘cold season’.²⁰ The aspirated palatal stop is a rare word-medial onset in roots.

²⁰ Northern Sangtam is a typical Tibeto-Burman language of Northeast India in having a suite of nominalizing morphemes, one of which has been reanalyzed as a past tense marker. To illustrate, the suffix *-tf^hō* functions synchronically as a nominalizing morpheme in the citation forms of verbs, e.g. *lāk-tf^hō* ‘bind-NMLZ2’, as a relativizing/nominalizing suffix in participial-type relative clause constructions, e.g. *[[ī-súŋ-ō-tf^hō]_{REL} kāk]_{NP}* (1SG:POSS-bear-NMLZ2 year) ‘the year I was born/my birth year’, and as a past tense marker, e.g.

Both the voiceless aspirated and voiceless unaspirated velar plosives /k, k^h/ occur frequently as word-initial and word-medial syllable onsets, e.g. *kī* ‘carbuncle, boil’, *ā-kāp* ‘PFX-skin’, *k^hīŋ* ‘frog’ and *jōŋk^hā* ‘rice beer’. As discussed in §6.1, the voiceless unaspirated velar plosive /k/ has two allophones in complementary distribution: the velar allophone [k] is realized in the onset position, and the glottal stop allophone [ʔ] is realized in the coda position, e.g. *kék* [keʔ⁵³] ‘eight’. In addition to producing a high checked tone, the shift to a laryngeal place of articulation is sometimes accompanied by creaky voice phonation on a preceding vowel, which is demonstrated by the presence of jitter in spectrograms. This is predictable, as it only seems to occur in the presence of [ʔ]. Unlike other Aoic languages, the glottal stop allophone is not always deleted before a syllable boundary, e.g. [vɔʔ⁵³pa¹¹] *vókpa* ‘sweep-NMLZ1’, [a¹¹pɔʔ⁵³ki³³] *à-pókki* ‘PFX-below’, but there is nevertheless an overwhelming preference for it to have a word-final realization. Word-medial glottal stops continue to be represented by <h> in the current Sangtam orthography yet are infrequently pronounced, so the only residual evidence of their former presence is a high tone.

8.2 Affricates

A pair of voiceless affricates occurs at the dental place of articulation and manifests a phonemic contrast based on the feature of aspiration. These segments are restricted to the onset position of the syllable, occurring word initially and word-medially, e.g. *tsú-īē* ‘eat-PRES’, *náktsē* ‘eye’, *ts^hé* ‘buttocks’, and *īts^hā-pā* ‘imitate-NMLZ1’.

The voiceless palato-alveolar affricates contrast with the dental affricates in place of articulation but share the same aspiration contrast, as previously noted. These also are restricted to the onset position of the syllable. They occur word-initially, e.g. *tʃā* ‘stick’, *tʃ^hā* ‘excrement’, and word-medially, e.g. *à-tʃā* ‘PFX-grandmother’, *ī-tʃ^hō* ‘buy-NMLZ2’.

Northern Sangtam, Yimkhiung and Lotha stoutly maintain a phonemic contrast between their dental and palato-alveolar affricates and occur before all vowels, whereas these phonemes appear to be undergoing a merger in their place of articulation in the Mongsen and Chungli dialects of Ao. Temsunungsang (2009) treats Chungli [ts] and [tʃ] as allophones of a single phoneme /tʃ/, while Coupe (2007: 31) reports that the only environment in which Mongsen /ts, ts^h/ can occur in contrast with /tʃ, tʃ^h/ is before schwa, and therefore concludes that, like Chungli, a merger in place of articulation is underway.

8.3 Pre-stopped bilabial trill

This typologically unique phoneme is a doubly-articulated segment that begins with the speaker first making an apico-dental occlusion and subsequently releasing it into a voiceless bilabial trill in a near-simultaneous gesture. A phonemic aspiration contrast distinguishes the voiceless unaspirated pre-stopped bilabial trill /tʙ̥/ from the aspirated

sàŋtām *cāmāḷā* *tsā* *thājān* *lānā* *pōmā* *īō-tʃ^hō*
 Sangtam ancestor DIST Thailand ABL Burma come-PST
 ‘The Sangtam ancestors came from Thailand to Burma.’

See Coupe (2008, 2011, 2017) for further examples of the reanalysis of nominalizers and their extended functions in the Tibeto-Burman languages of central Nagaland and beyond.

counterpart /t̪ʰ/, and both are found word-initially and word-medially in the onset position of syllables, e.g. *t̪ʰà* ‘fence’, *tsút̪ʰū-ɪ̃* ‘relative-ANMLZ’, *lékt̪ʰh̃ŋ* ‘bat’, and *t̪ʰék* ‘cut by placing a blade on flesh and making a long straight incision’. In distinction to an almost identical sound described in two Chapakuran languages by Ladefoged & Everett (1996), the Northern Sangtam pre-stopped bilabial trill is by no means a rare phoneme in the language – over two dozen examples were elicited in the ca. 900-item word list, and at least 150 headwords containing this consonant are listed in the *Yusü: Bilingual Dictionary of Sangtam-Naga*. It occurs before all vowels.

Speakers precede the release of the dental occlusion with a characteristic slight compression and protrusion of the lips (see Figure 2 below). This lip attitude results in an increased area of surface contact in the sagittal plane, resulting in a particular configuration that is aerodynamically conducive to producing a brief bilabial trill when the segment is followed by a high back rounded vowel.

Figure 2. Lateral and frontal views capturing a Northern Sangtam speaker's articulation of [t̪ʰeʔ⁵³] ‘cut by placing a blade on flesh and making a long straight incision’, demonstrating the characteristic slightly protruded and compressed lip attitude immediately preceding the release.

The strength, duration and nature of the release is determined by the vocalic environment. It may extend to three successive oscillations of the lips if the vowel is [+hi, +back]; in the environment before other vowels, there may be just a single tap if the segment is unaspirated, or the dental occlusion may be released into a voiceless bilabial affricate [t̪ʰɸ], especially if the segment is aspirated. An allophone [t̪ʰ] with a voiced bilabial component is likely to be realized when the unaspirated segment occurs word-medially before a vowel, due to assimilation to voicing, e.g. [kʰi⁵⁵t̪ʰa³³] ‘hip joint’. When occurring word-initially as an onset, it is articulated with a near-coincident VOT.

In their identification of doubly-articulated labial-dorsal stops and nasals, Ladefoged & Maddieson (1996: 333) propose that such segments should have a near simultaneous cooccurrence of two articulations and durations comparable to segments produced with a single articulation. These criteria plausibly apply to the doubly-articulated pre-stopped bilabial trills of Northern Sangtam. VOT measurements of the unaspirated pre-stopped bilabial trill averaged 9 ms for nine citation forms uttered in

isolation, while those of the aspirated bilabial trill averaged 58 ms for thirty tokens. This compares favourably to measurements for plain unaspirated and aspirated plosives (averaging 9 ms and 61 ms respectively for six randomly selected tokens of each segment), suggesting that the pre-stopped bilabial trills qualify as unitary segments in respect to duration.

According to Ladefoged & Maddieson (1996: 130), ‘all bilabial trills historically developed from a sequence of a prenasalized bilabial stop followed by a relatively high back rounded vowel, i.e. a sequence such as **mbu**’. Given the cooccurrence of a nasal with a plosive in some cognate words of the Aoic group, as well as such sequences in a few of Matisoff’s (2003) reconstructed proto-forms (see Table 8 below), Ladefoged & Maddieson’s proposal possibly accounts for the historical origin of Northern Sangtam pre-stopped bilabial trills. However, such consonants perhaps do not strictly require a following high back rounded vowel to develop the trilled release, as none of the reconstructed proto-forms of Table 8 present such sequences involving high back rounded vowels. Furthermore, very few of the Proto-Tibeto-Burman cognates reconstruct with sequences of nasal+stop that might have developed into bilabial trills as proposed, so these peculiar segments will require additional investigation to determine how they might have evolved.

Table 8. Correspondences with Northern Sangtam /t̪ʰ, t̪ʰ/ in the Aoic languages

	N. Sangtam	Mongsen Ao	Yimkhiong	Lotha ²¹	PTB
‘arrow’	lit̪ʰ ²²	lit̪ʰāk-t̪ʰān ²³	sāŋlùŋ?	lōts ^h ə	*b/m-la, *m-da
‘deaf’	nāŋt̪ʰān	t̪ʰ-nāiūŋ t̪ʰ-t̪ʰāk ²⁴	nākēn pīyāŋ	nəp̪t̪ʰŋ?	*m-baŋ
‘drum’	t̪ʰān	āphēn	sāŋtān	ə-p ^h iaŋ	---
‘grey’	kūt̪ʰūk	t̪ʰpū	pījūk	máŋzə	*pwəy
‘fence’	t̪ʰā	ikhū	ʃāʔhīpēn/ʃāʔōn	ip̪i	*hwaŋ, *kram
‘groin’	k ^h it̪ʰāk	kājā-t̪ʰān	māʔ ^h i	ə-zó?	*kap
‘needle’	t̪ʰān	impēn	tīpān	ə-cēm	*ka:p~ʔap
‘plank’ (wood-flat)	sīŋ-t̪ʰāk	sāŋ-pāk	ʃəpsāŋ	tsāŋ-p ^h āk	*pləŋ

An age difference appears to account for the presence or absence of pre-stopped bilabial trills in Northern Sangtam speakers’ phoneme inventories.²⁵ Some younger speakers lack these unusual phonemes entirely, instead articulating the words of Table 8 with plain dental plosives [t], [t^h]. An interesting correlation can be drawn with the Chapakuran languages, in which the pre-stopped bilabial trill was also reported to be

²¹ The Lotha words were provided by a speaker originating from Okotso village, northern Wokha District.

²² The correct Northern Sangtam gloss for this may be ‘arrow fletching’, rather than just ‘arrow’.

²³ Literally ‘catapult/bow + seed’.

²⁴ Obviously non-cognate; literally ‘ear + break’.

²⁵ When I elicited words containing the pre-stopped bilabial trill from my male speakers aged between thirty-two and forty-seven years in Tronga Village, older speakers observing the proceedings criticized the younger speakers who substituted this sound with a plain plosive and expressed the view that they were not speaking the language properly.

endangered and undergoing replacement with plain dental plosives at the time of their documentation (Everett & Kern 1997: 385). See §10 for further discussion.

8.4 *Nasals*

Four nasal phonemes match the plosives at each place of articulation, viz. bilabial, dental, palatal and velar. As noted in §6.1, the nasal phonemes demonstrate some constraints on their phonotactic distribution. The bilabial nasal may occur as an onset word-initially and word-medially e.g. *mīpùŋ* ‘abdomen’, *ʃām̄l̄* ‘bear’ (n.), and more rarely as a coda, e.g. *jíŋ.ɿām-tʃʰò* ‘become dark-NMLZ2’. There is a strong preference in the corpus for the bilabial nasal to occur as the onset of a syllable with a central vowel nucleus.

Like the dental plosives, the dental nasal is constrained to occurring as an onset word-initially and word-medially, e.g. *nàŋkù* ‘ear’, *nā* 2SG, *tʃʰónù* ‘moon’, and *tʃlànì-ɿl̄* ‘cousin-ANMLZ’. Coronal consonants never occur as syllable codas.

The palatal nasal occurs word-initially and word-medially before all vowels, but is restricted to the onset position. Since this does not apply to the bilabial and velar nasal phonemes, it might be considered evidence for assuming that the palatal plosives and nasals are relatively recent phonological developments, or it could be attributable to the coronal coda constraint. Some examples in words are *āŋáŋl̄* ‘morning’, *ɲl̄l̄* ‘buffalo’, *tʰāŋéŋ* ‘seven’ and *ɲĩtsāɿl̄* ‘human being’.

The velar nasal occurs rarely as a word-initial or word-medial onset, e.g. *ŋũtʃʰi* ‘work (n.)’ and *à-ŋl̄ŋók* ‘PFX-sinew’, but more frequently as a coda both word-medially and word-finally, e.g. *kāŋzáŋ* ‘insect’, *hòŋ* ‘wind’. There is a good deal of free variation in the articulation of velar nasal codas. In the speech of many consultants, the nasal feature is becoming transphonologized to a preceding vowel. This is suggested by both the considerable [Vŋ] ~ [Ṽ] free variation noted in elicited words, and the observation that a velar nasal segment represented orthographically as <ng> in the currently used writing system is often not actually pronounced as [ŋ] at all. Free variation produced by all consultants in numerous elicited examples demonstrate that the [+nasal] feature is undergoing anticipatory assimilation to a preceding vowel, thereby permitting syncope or apocope of the velar nasal without loss of meaning, particularly before a low central vowel, e.g. [a¹¹tsiŋ³³lã³³] ~ [a¹¹tsiŋ³³laŋ³³] ‘lung’, [a³³saŋ³³] ~ [a³³sã³³] ‘three’, and [a¹¹ŋã³³] ~ [a¹¹ŋaŋ³³] ‘price. The velar nasal also undergoes a neutralization of contrast in some environments; e.g. MS writes *ʃàŋ-pà* ‘run-NMLZ1’ as <shangba> in the current Sangtam orthography, but pronounces this word as [ʃãmbà], thus demonstrating the loss of a phonemic opposition between /ŋ/ and /m/ here.

8.5 *Doubly-articulated labial-coronal nasal*

The doubly articulated labial-coronal nasal /*nm̄*/ is tentatively accorded phonemic status in this description. This typologically unusual segment is represented by <nb> or <np> in the *Yusüp: Bilingual Dictionary of Sangtam-Naga*. One of the two Northern Sangtam speakers whose examples of /*nm̄*/ I recorded demonstrated free variation between articulating a doubly articulated labial-coronal nasal [nm̄] and a prenasalized stop [nb], with the latter realization occurring much less frequently, and

only in the word-medial environment in the recorded examples. Further investigation with other speakers is required to determine the extent of $[\widehat{nm}] \sim [\widehat{nb}]$ free variation in the speech community. This should focus on the speech of speakers over 50 years of age, as younger speakers are likely to merge this sound with a plain bilabial nasal, in parallel with their merger of the pre-stopped bilabial trills with plain dental plosives, or replace it with a prenasalized dental plosive, depending upon its position of occurrence in the word (see §10 for further discussion).

The initiation of this sound is somewhat similar to the pre-stopped bilabial trill described in §8.3. Speakers first make an apical-dental occlusion while simultaneously compressing their lips to create a second stricture, resulting in a doubly articulated nasal involving two overlapping articulatory gestures. It is noteworthy that the slightly compressed and protruded lip attitude previously demonstrated by the speaker in Figure 2 is also consistent with the articulation of $\widehat{nm}/$, which raises suspicions of its relationship to $\widehat{tb}/$, or perhaps historically, to a prenasalized stop.²⁶

Examples are rare in the *Yusüp: Bilingual Dictionary of Sangtam-Naga*, which has approximately two dozen headwords containing $\langle nb \rangle$ or $\langle np \rangle$, with the majority of examples falling under the semantic domains of spirituality and types of body hair (and thus making rather strange bedfellows). It occurs both word-initially and word-medially in the recorded corpus, e.g. $\widehat{nm}\widehat{a}t\widehat{f}\widehat{\lambda}-p\widehat{a}$ ‘forget-NMLZ1’ and $\widehat{a}-\widehat{nm}\widehat{a}$ ‘body hair’, but is likewise restricted to just a handful of words. Note that $\widehat{tb}/$ occurs as an onset in identical vocalic environments, which argues against treating $\widehat{nm}/$ as an allophone of the pre-stopped bilabial trill, but it is nevertheless suspiciously limited to occurring only before the central vowels $[\Lambda, a]$ in the corpus. The phonotactic prohibition on dental codas rules out treating the dental nasal component as the coda of a preceding syllable in words such as $\widehat{a}-\widehat{nm}\widehat{\lambda}$ PFX-spirit/soul and $\widehat{\lambda}-\widehat{nm}\widehat{\lambda}$ ‘my spirit/soul’, and the possibility of it occurring word-initially in the onset slot in $\widehat{nm}\widehat{a}t\widehat{f}\widehat{\lambda}-p\widehat{a}$ ‘forget-NMLZ1’ confirms its status as a complex segment involving multiple gestures.

The doubly-articulated labial-coronal nasal is an extremely rare speech sound from a typological perspective; the only other language of the world in which a similar labial-coronal nasal has been reported is Yeletnye, an isolate spoken on Rossel Island, Papua New Guinea (Ladefoged & Maddieson 1996: 344). It may well be significant to understanding its development that Yeletnye is also reported to have a doubly-articulated labial-alveolar plosive $\widehat{tp}/$ in addition to $\widehat{nm}/$, further raising the suspicion of some historical relationship holding between Northern Sangtam’s doubly-articulated segments $\widehat{tb}^{(h)}/$ and $\widehat{nm}/$.

8.6 Fricatives

Northern Sangtam has six fricative phonemes produced at four distinct places of articulation: labial-velar, dental, palatal, and velar. The labial-dental fricative is voiced, the dental fricatives demonstrate a voicing contrast, and the palatal and velar segments are voiceless. Also included here is the so-called voiceless glottal “fricative”

²⁶ Videos with simultaneous lateral and frontal views of two speakers producing various realizations of $\widehat{tb}/$, $\widehat{tb}^{(h)}/$ and $\widehat{nm}/$ are available in the research data repository – see fn. 1.

/h/, which is best viewed as being unspecified for its place of articulation, and thus not a true fricative.

The labial-dental fricative /v/ occurs word-initially and word-medially but is not a common segment in the recorded corpus, the recorded examples being *kùvā* ‘hair (of the head)’, *āvék* ‘land leech’, *tāvī* ‘ashes’, *à-vā* ‘PFX-leaf’ *mivā-pà* ‘diarrhoea-NMLZ1’ (lit. ‘stomach-go’), *jùvā* ‘voice’, *ívā-pà* ‘follow-NMLZ1’, *vā-ŋ* ‘go-IMP’ *vókpa* ‘sweep’, and *kívī-pà* ‘swim-NMLZ1’ (lit. ‘water+swim’). Despite the relative infrequency of its occurrence in the data, over one-hundred headwords containing /v/ can be found in the *Yusüp: Bilingual Dictionary of Sangtam-Naga*, including those words erroneously listed under <W w>, as there is no true labial-velar approximant phoneme in this language.

It was suggested in §8.1, that the rarely-occurring voiceless labio-dental fricative is most likely to be an evolving allophone of /p^h/ in the environment before a central vowel. Possibly the lenition associated with the central vowel affects the onset, resulting in an allophonic realization of [f] in this particular phonetic environment.²⁷ The few words in which it was recorded are *fā* ‘body’, *ā-fā* ‘PFX-older sibling’, *fāzā* ‘dog’ and *fātfāŋ* ‘hoe’. Considerably more examples can be found in the *Yusüp: Bilingual Dictionary of Sangtam-Naga*; these too overwhelmingly occur before the mid-central vowel (represented orthographically by <ü> in that publication, and the only two listed exceptions – <fifiu> ‘whistle’ and <fehfehle> ‘windy’ – can be accounted for by onomatopoeia.

The dental fricatives /z, s/ are more precisely described as having a denti-alveolar place of articulation and are unremarkable, occurring as syllable onsets word initially and word medially only, e.g. *zīŋ-pà* ‘pull-NMLZ1’, *‘za* ‘grass’, *kūzūŋ* ‘line, row’. Examples of the voiceless dental fricative in words are *sák* ‘breath’, *súló-tʰō* ‘bear-NMLZ2’, *ísà* 1PL.INC, and *ā-sá* ‘PFX-bile’.

The voiceless palatal fricative /ʃ/ occurs as an onset word-initially and word-medially before all vowels, e.g. *ʃiŋ* ‘ginger’, *ʃè* ‘thread’, *ʃú* ‘meat’, *ā-ʃá* ‘PFX-blood’, *ʃók* ‘sambar deer’ and *ʃámāʃá* ‘barking deer’. The phonemic status of a voiced palatal fricative [ʒ] is dubious. It occurs very rarely in the recorded corpus, but in the same environments as the voiceless palatal fricative, e.g. [a¹¹mā⁵⁵ʒiŋ⁵⁵] ‘PFX-tail’, [a¹¹ʒā³³ʒe⁵³] ‘PFX-grease’. Deeper investigation reveals that it is in free variation with the palatal approximate [j], and for this reason it is best analyzed as a variant realization of /j/. See §8.7 for further discussion and examples.

The voiceless velar fricative /x/ occurs in the onset slot word-initially and word-medially, e.g. *xá-pà* ‘reap-NMLZ1’, *xù* ‘bamboo’, *à-xā* ‘PFX-intestine’, and *à-xiŋ* ‘PFX-raw, fresh’ (meat etc.). This segment is palatalized before a high front vowel, resulting in a voiceless palatal fricative allophone in this environment; e.g. *xísek-tʰō* ‘strangle-NMLZ2’ is realized phonetically as [çi⁵⁵seʔ⁵³tʰo³³].

²⁷ Comparison might profitably be made to Chokri, in which [pf] is realized as an allophone of /p^h/ in the environment before a high central vowel [i] (Bielenberg & Nienu 2001: 90). This presumably also developed due to phonological weakening and centralizing of the vowel subsequently triggering lenition of the preceding onset.

The glottal fricative /h/ occurs word-initially and word-medially in the onset position before a variety of vowels, e.g. *hòṁā* ‘plantain’, *hápáp* ‘fireplace shelf’, and *tʃ^hĩghì* ‘star’, as well as in the demonstrative pronouns *hàtsā* ‘that’ and *hī* ‘this’.

8.7 Approximants

There are three voiced approximants: one lateral, one central retroflex, and one palatal. The lateral approximant /l/ occurs word-initially and word-medially, e.g. *lólā* ‘joint’, *lè* ‘leg’, and *p^hàlì-ā* ‘nephew-ANMLZ’ (of the distaff side). The central retroflex approximant is restricted to the onset position in lexical roots, occurring word-initially and word-medially, e.g. *ǰáŋ-pà* ‘seize-NMLZ1’, *ǰita-tfō* ‘argue-NMLZ2’, *ǰlíoŋ* ‘fox’ and *ǰāā* ‘paddy’. As noted in §6.1, the only situation in which this segment can fill the syllable coda is when the genitivizing/nominalizing suffix *-ǰ(ā)* is resyllabified as such in nominal derivations and converb constructions.

The voiced palatal approximant /j/ functions as a syllable onset word-initially and word-medially, e.g. *ǰùṁmāzā* ‘rat’,²⁸ *ǰètǰāŋ* ‘net’, *mā-jā* ‘PFX-four’, and *k^hǰǰák* ‘armpit’. This phoneme has two allophones [j] and [ʝ] occurring in free variation. Some examples of variant realizations in words are, [ʝe⁵⁵] ~ [je⁵⁵] ‘penis’, [ǰəm⁵⁵pa¹¹] ~ [ʝəm⁵⁵pa¹¹] ‘squeeze-NMLZ1’, and [ʝā¹¹ǰe¹¹ǰe¹¹ a¹¹ǰā³³] ~ [jā¹¹ǰe¹¹ǰe¹¹ a¹¹ǰā³³] ‘obsolete form for EIGHTY’. Bruhn (2014:162) reports variation between the transcription of [ʝ] and [j] in the published sources of data on Sangtam. A related example of this can be found in Kauffmann (1939: 230), who recorded free variation between [z] and [j] in the Chare village dialect of Northern Sangtam, judging from his listing of variant forms *ázong*, *áyong* for ‘Berg’ (mountain). Free variation between a voiced coronal fricative and a palatal approximant has also been documented in Mongsen Ao – see Coupe (2003: 47–50) for discussion – suggesting that this might be a shared phonological characteristic of the Aoic languages.

9 Vowel phoneme inventory

The Northern Sangtam vowel phoneme inventory consists of four peripheral vowels /i, e, o, a/, and one centralized mid-back vowel /ʌ/, presenting a quasi-symmetrical triangular system contrasting three degrees of height, three degrees of backness, and two degrees of rounding. There are no phonemic diphthongs, as sequences of vowels form the nuclei of separate syllables. The vowel phonemes are presented in Table 9 below.

Table 9. Sangtam vowel phoneme inventory

	Front	Central	Back
close	i		u
mid-close	e		o
mid-open		ʌ	
open		a	

²⁸ This means literally ‘drink-not-sharpen’, a name deriving from a folktale about how a duplicitous rat agreed to sharpen a dao (hatchet) in exchange for rice wine, then refused to sharpen the dao after drinking the wine.

9.1 Monophthongs

The high front vowel /i/ occurs in initial, medial and final positions of the word, e.g. *ī* 1SG, *k^hákíók* ‘scorpion’, and *ā-nì* ‘younger sibling’. The mid-close front vowel /e/ demonstrates a more limited occurrence, as there are no examples in the corpus of /e/ appearing in the word-initial position. Word-medial occurrences in lexical roots are also quite rare, some examples being *nèŋ* ‘name’, *k^hék* ‘arm’, *c^hékūlā* ‘woman’ and *ŋék^hī* ‘cloud’. It does appear very frequently in the word-final position, however, e.g. *t^hàlè* ‘ten’, *pè* ‘mouth’, and it forms the nucleus of the present tense inflection, e.g. *mālū-lē* ‘boil-PRS’, *lì-lē* ‘dwell-PRS’, a morpheme with a relatively high functional load.

The high back rounded vowel /u/ is frequent in lexical roots and occurs word-initially, word-medially and word-finally, e.g. *úpí* ‘peacock’, *jùk^hék* ‘tale’, and *kū* ‘house’. The mid-close back rounded vowel /o/ is much more frequent in the corpus than its front unrounded counterpart /e/. It occurs as what may be a particle in the interrogative pronouns *k^hájpa ò* ‘which one?’ and *tú ò* ‘what?’, and it occurs word-medially and word-finally in lexical roots, e.g. *jòŋp^hók* ‘mushroom’, *mājò* ‘palm of hand’ and *jìtā-t^hō* ‘argue-NMLZ2’.

Bruhn (2014: 163) notes that older sources (Marrison 1967, Kumar 1973, Weidert 1987) report separate vowel phonemes /o/ and /u/, yet he suspects that they represent just one back rounded vowel, because he finds transcription inconsistencies both within and between these publications. The reported inconsistencies may be indicative of a merger that is gradually erasing the phonemic contrast between the back rounded vowels in certain environments. For example, my sixty-six consultant MS accepted either [o] or [u] before a velar nasal in words such as [kaŋ¹¹lõŋ³³.l³³] ~ [kaŋ¹¹lũŋ³³.l³³] ‘animal’, [a³³le³³mõŋ¹¹] ~ [a³³le³³mũŋ¹¹] ‘Himalayan mole’, and [lõŋ³³] ~ [lũŋ³³] ‘stone’, and he demonstrated considerable free variation between these two articulatory positions in elicited tokens with velar nasal codas. Despite this, subsequent investigation confirmed a clear phonemic contrast in minimal pairs such as *xù* [xu¹¹] ‘bamboo’ and *xò* [xo¹¹] ‘yam’, therefore two back rounded vowel phonemes must still be posited in a synchronic description. The fact that either back rounded vowel is accepted in the environment before a velar nasal indicates that a phonemic merger may be evolving. Such a development would then affiliate Northern Sangtam to the Ao dialects, in which a single high back rounded vowel phoneme has [u] and [o] allophones in free variation, particularly before a velar consonant (e.g. see Coupe 2003: 42).

The vowel /ʌ/ demonstrates two allophonic realizations in near-complementary distribution. When preceded by [–high] consonants it is realized as [ʌ], a centralized allophone that is slightly lower and further back than schwa in the vowel space, e.g. *ts^hàzā* [ts^hʌ¹¹.l³³] ‘mithun’ (*Bos frontalis*), *kālālā* [kʌ¹¹la¹¹.l³³] ‘mongoose’ and *pāl* [pʌ³³] ‘axe’, but it may be realized as an allophone approaching [ə] when occurring in the environment after [+high] consonant onsets, e.g. *à/á/ék* [a¹¹ʃə⁵⁵ʃe⁷⁵³] ‘be slippery’, *mājāl* [mʌ¹¹jə³³] ‘poison’, *ŋāpōŋ* [ŋə³³pōŋ³³] ‘wind’. The [+high] feature of palatal and palato-alveolar consonants requires the front of the tongue to be raised and slightly advanced to put it in the proximity of the palatal place of articulation. This results in an enhanced centralization and raising of the following vowel, the assimilated position contributing to its realization as [ə]. Sangtam words with /ʌ/ frequently correspond with

cognates containing a schwa in Mongsen Ao, e.g. Sangtam $\bar{a}p\bar{a}$ ~ Mongsen Ao $\bar{a}n\bar{a}t$ ‘two’, Sangtam $m\bar{a}c\bar{a}$ ~ Mongsen Ao $m\bar{a}k\bar{i}$ ‘twenty’, Sangtam $m\bar{a}k^h\bar{u}k$ ~ Mongsen Ao $t\bar{a}-m\bar{a}k\bar{u}k$ ‘knee’, and Sangtam $m\bar{a}t\bar{f}\bar{a}m\bar{a}p^h\bar{i}$ ~ Mongsen Ao $m\bar{a}t\bar{s}\bar{a}$ ‘spittle’.

The low central vowel /a/ occurs word-initially, word-medially and word-finally in lexical roots and grammatical morphemes, e.g. $\bar{a}p\bar{i}$ 3SG, $m\bar{a}m\bar{a}$ ‘breast’, and $z\bar{a}$ ‘grass’. Its high frequency in the corpus can be attributed to the fact that one of the nominal prefixes of Northern Sangtam is *a-*, which is a requirement for word formation of numerous nouns, e.g. $\bar{a}-h\bar{a}$ PFX-beak’, $\bar{a}-l\bar{i}$ ‘PFX-ground’ and $\bar{a}-n\bar{i}$ ‘PFX-younger sibling’.

9.2 Transphonologized nasality

It is observed that Northern Sangtam speakers write many words with <ng> to represent a velar nasal coda in their current orthography, but often pronounce such words with a nasalized vowel and a syncopated or apocopated velar nasal coda, e.g.

(6) phonetic transcription	orthographic representation	gloss
$[k^h\bar{a}^{11}t^h\bar{u}^{55}ba^{11}]$	<khangthrua>	‘open-NMLZ1’
$[s\bar{a}^{33}je^{11}]$	<sangre>	‘thirty’
$[a^{11}\eta\bar{a}^{33}]$	<angang>	‘price’

This suggests that the low central vowel /a/ is developing a nasal allophone $[\bar{a}]$ via transphonologization of the nasal feature from the velar nasal coda. This is a well-known process that has resulted in the diachronic development of nasal vowels in French and other modern Romance languages (e.g. see Sampson 2011). At present, vowel nasalization and loss of the velar nasal coda appears to affect rhymes with a low central vowel nucleus almost exclusively in the speech of my sixty-six year-old consultant, but younger speakers are extending nasality to other vowels. It was observed that $lékt\bar{b}^h\bar{i}\eta$ ‘bat’ was pronounced variously as $[le\eta^{53}t^h\bar{i}^{33}]$ and $[le\eta^{53}\bar{t}\bar{b}^h\bar{i}\eta^{33}]$ by consultants in their thirties and forties, and similar variation was noted in their pronunciations of $\bar{a}t\bar{b}^h\bar{i}\eta$ ‘fin’ as $[a^{33}t^h\bar{i}^{33}]$ or $[a^{33}t^h\bar{i}\eta^{33}]$, and $h\bar{u}\eta$ ‘neck’ as $[h\bar{u}^{11}]$ or $[hu\eta^{11}]$. Examples of variation can be found in the archived word lists.

10 Aged-related phonological differences in Northern Sangtam

As reported in Coupe (2015), there is an evolving shift away from marked sounds in the phonological inventories of younger speakers. This was first noticed in the speech of consultants aged less than 30 years who were living in Dimapur, the multilingual commercial centre of Nagaland. However, the same changes were observed in speakers aged up to 47 years who had lived in Tronga Village for their entire lives, so the observed sound changes cannot be attributed to the influences of language contact in Dimapur as initially suspected.

The following comments apply to the elicited data of six male speakers aged 28, 31, 32 (two individuals), 37 and 47 years of age. Firstly, there is a very strong tendency for all speakers in this age range to replace their prestopped bilabial trills $\bar{t}\bar{b}/$ and $\bar{t}\bar{b}^h/$

entirely with the plain voiceless dental stops /t/ and /tʰ/ respectively. At present it is not understood what could be motivating this sound change in the speech of younger speakers. I have observed an 18-month-old infant playfully making bilabial trills, and preceding such a manner of articulation with a apical-dental occlusion should not render such a sound overly difficult to articulate. Lack of perceptual salience between the pre-stopped bilabial trills and plain stops could possibly motivate the loss of a phonemic contrast, but it is not clear why the putative loss of perceptual salience should only be apply to younger generations of speakers, if this is a valid explanation. Perhaps the most plausible motivation could be social – that such a sound has come to be considered awkward by younger speakers or representative of an older generation – and for this reason it is avoided. None of these possible explanations seems entirely satisfactory, so I leave this as a topic in need of further investigation.

A second major change was noted in younger speakers' pronunciation of *xò* 'taro', which was uniformly articulated as [ho¹¹] and is consistent with the widely observed phenomenon of lenition. Two speakers also depalatalized postalveolar fricatives in some words, resulting in pronunciations of *ʃè* 'thread' as [se¹¹], and one of those speakers also depalatalized *tʃà* 'stick', rendering it as [tsa¹¹], but retained *tʃ* in other words.

The youngest (28 year old) speaker pronounced *zā* 'grass' as [ja³³], demonstrating the *z* ~ *j* variation noted by other authors and discussed above in §8.7. He also palatalized dental affricates in some words, but was the only speaker to do this, so that might have been merely ideolectal variation or performance error.

Five of the six speakers replaced the doubly articulated labial-dental nasal *nm̃* with a voiced prenasalised dental stop [nd] in the word *â-nm̃λ* 'PFX-soul, spirit', while the youngest speaker pronounced this word with a prenasalized bilabial stop [a³³mba¹¹], so it appears that this segment has also disappeared from the consonant inventories of younger Northern Sangtam speakers. It is unfortunate that word-initial examples of *nm̃* were not recorded, as the word-internal environment may have contributed to this variant realization. It was also inconsistent in the data of my sixty-six year-old consultant MS, who fluctuated between [a³³nba¹¹] and [a³³nm̃a¹¹] in his elicited tokens, but he asserted that [a³³nm̃a¹¹] was the correct pronunciation for this word when questioned about the variation in his pronunciation.

Lastly, all speakers apocopated velar nasal codas in many words and nasalized preceding vowels of differing qualities in compensation – see §9.2 for examples and discussion.

11 Concluding comments

One of the most intellectually stimulating and satisfying aspects of working on undescribed minority languages is that one can never predict just what kinds of unusual and important discoveries one's research might bring to light. The phonology of Northern Sangtam is a excellent example of this.

The most remarkable findings in this description are the language's doubly articulated pre-stopped bilabial trill and the doubly articulated labial-coronal nasal.

These consonant sounds represent extreme phonological rarities in the world's languages, and it is likely that Northern Sangtam has the further distinction of being the only language spoken on the planet in which pre-stopped bilabial trills are phonemic. The labial-coronal nasal is also exceptional, having only been previously reported in the isolate Yeletnye of Rossel Island, Papua New Guinea. Whereas labial-velar articulations are fairly common, particularly in West Africa, the combination of a labial + coronal articulation is so unusual that initially it was thought that it did not exist in any of the world's languages, and it was not until fieldwork was done on Yeletnye that this assumption was revised (Ladefoged & Maddieson 1996: 344).

Is it just a peculiar coincidence that both Yeletnye and Northern Sangtam have similarly rare segments in their phonological inventories? What might be the relationship, if any, holding between their doubly-articulated labial-coronal consonants? What factors could be motivating the replacement of the pre-stopped bilabial trills with plain dental stops in the speech of younger speakers? Such questions encourage further investigation of these segments' unique characteristics and diachronic development, given their typological importance for appreciating the extent of diversity in the phoneme inventories of the world's languages.

Another finding of significance that is particularly relevant to subgrouping is the discovery of an obsolete overcounting numeral system in Northern Sangtam. This turns out to be identical to the overcounting cardinal numeral patterns previously documented in the Ao dialects, both in the alternating overcounting/decimal pattern of counting observed in each decade after FIFTEEN, as well as in the use of cognate morphological forms and structure. This shared innovation presents the best evidence to date for establishing the genetic relationship of Northern Sangtam to Aoic.

Lastly, a classic case of debuccalization has resulted in the evolution of a new stopped tone in the language, thus providing a textbook demonstration of how sound change involving coda consonants can give birth to new tones.

Acknowledgements

I am pleased to acknowledge support from the Singapore Ministry of Education in the form of two Academic Research Fund grants that have facilitated background research on the languages of Nagaland: MOE2012-T2-1-100 'Exploring the crossroads of linguistic diversity: Language contact in Southeast Asia', and MOE2016-T1-001-220 'Archaeological linguistics and the prehistory of Northeast India: reconstructing the past through ancient technologies and practices, and correlating the results with migration histories'. An Alexander von Humboldt Fellowship for Experienced Researchers (2016-18) permitted me to work up the initial draft of this paper at the University of Cologne's Institute for Linguistics, and I am grateful to both of these German institutions for their generous support. The paper has benefited considerably from the comments and suggestions of two anonymous reviewers, the editor, and Scott Moisik, none of whom bears responsibility for any remaining errors of fact or deficiencies in the analysis. Finally, I express my sincere gratitude to MS and other members of the Sangtam community of Nagaland for contributing their time and data to this phonological investigation.

References

- Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press.
- Bielenberg, Brian & Zhalie Nienu. 2001. Chokri (Phek dialect): Phonetics, phonology. *Linguistics of the Tibeto-Burman Area* 24.2: 85–122.
- Bouchery, Pascal & Lemlila Sangtam. 2012. The Kinship Terminology of the Sangtam Nagas. *European Bulletin of Himalayan Research* 41. 9–29.
- Bruhn, Daniel Wayne. 2014. A phonological construction of Proto-Central Naga. Unpublished PhD dissertation, University of California, Berkeley.
- Burling, Robbins. 1999. On “Kamarupan”. *Linguistics of the Tibeto-Burman Area* 22.2: 169–171.
- Burling, Robbins. 2003. The Tibeto-Burman languages of northeastern India. In Thurgood, Graham and Randy J. LaPolla (eds.) *The Sino-Tibetan languages*. 169–191.
- Clark, Mary Mead. [1893] 2002. *The Ao-Naga grammar*. New Delhi: Mittal Publications.
- Chao, Y. 1930. A system of “tone-letters”. *Le Maître Phonétique* 30: 24–27.
- Coupe, Alexander R. 2003. *A phonetic and phonological description of Ao: a Tibeto-Burman language of Nagaland, north-east India*. Canberra: Pacific Linguistics.
- Coupe, Alexander R. 2007. *A grammar of Mongsen Ao*. Mouton Grammar Library 39. Berlin and New York: Mouton de Gruyter.
- Coupe, Alexander R. 2008 (ed). Special issue on nominalization in Tibeto-Burman, *Linguistics of the Tibeto-Burman Area* Vol. 31.2.
- Coupe, Alexander R. 2011. On core case marking patterns in two Tibeto-Burman languages of Nagaland. *Linguistics of the Tibeto-Burman Area* 34.2: 21–47.
- Coupe, Alexander R. 2012. Overcounting numeral systems and their relevance to sub-grouping in the Tibeto-Burman languages of Nagaland. *Language and Linguistics*, Vol. 13.1: 193–220.
- Coupe, Alexander R. 2013. Phonological peculiarities of Sangtam: A preliminary investigation. Paper read at the 19th Himalayan Languages Symposium, The Australian National University, Canberra 6-7 September.
- Coupe, Alexander R. 2014. Strategies for analyzing tone languages. *Language Documentation & Conservation* 8: 462–489. <http://hdl.handle.net/10125/2461>
- Coupe, Alexander R. 2015. Pre-stopped bilabial trills in Sangtam. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper no 0734.1-5.

- Coupe, Alexander R. 2017. Mongsen Ao. In Thurgood, Graham & Randy J. LaPolla (eds.) *The Sino-Tibetan languages*, 2nd Edn., 277–301. Oxford and New York: Routledge.
- Eberhard, David M., Gary F. Simons, & Charles D. Fennig (eds.) 2019. *Ethnologue: Languages of the World, Twenty-second edition*. Dallas, Texas: SIL International. Online edition: <http://www.ethnologue.com>. Accessed 13/06/2019.
- Everett, Daniel L. & Barbara Kern. 1997. *Wari'*. Routledge: London & New York.
- Grierson, George Abraham. (ed.). 1967 Reprint. *Linguistic survey of India*. Vol 3, Part 2. Delhi: Motilal Banarsidass. Series first published 1903–1928.
- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242: 69–82.
- Hutton, John Henry. 1921. *The Angami Nagas, with some notes on neighbouring tribes*. London: MacMillan and Co.
- Hutton, John Henry. 1986 Reprint. Diaries of two tours in the unadministered area, east of the Naga Hills. *Memoirs of the Asiatic Society of Bengal*, vol. 11, No. 1. Calcutta: Asiatic Society of Bengal. Delhi: Mittal Publications. First published 1929.
- Imchen. 2014. Linguistic ecology of Sangtam. In War, J. and S. K. Singh and S. A. Lyngdoh and B. Khyriem (eds.), *Tibeto-Burman Linguistics of North-East India*, 54–75. Guwahati: EBH Publishers.
- Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.
- Kauffmann, H.E. 1939. Kurze Ethnographie der nördlichen Sangtam-Naga (Lophomi), Assam. *Anthropos* 34: 207–245.
- Kumar, Braj Bihari. 1973. *Hindi-Sangtam dictionary*. Kohima: Nagaland Bhasha Parishad.
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Malden, MA: Blackwell.
- Ladefoged, Peter & Daniel Everett. 1996. The status of phonetic rarities. *Language* 72.4: 794–800.
- LaPolla, Randy J. 1994. Parallel grammaticalizations in the Tibeto-Burman languages: evidence of Sapir's 'drift. *Linguistics of the Tibeto-Burman Area* 17(1): 61–80.
- MacEachern, M. R., Kern, B., Ladefoged, P. 1997. Wari' phonetic structures. *Journal of Amazonian Languages*, vol. 1, n.1, 3–28.
- Ian Maddieson. 2013. Consonant Inventories. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/1>, Accessed on 2019-09-13.)

- Marrison, Geoffrey Edward. 1967. The classification of the Naga languages of north-east India, Volume 1 (Text). Unpublished PhD dissertation, Department of Phonetics and Linguistics, School of Oriental and African Studies, University of London.
- Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman (ed. *Consonant types and tone*. Southern California Occasional Papers in Linguistics No. 1. Los Angeles: Linguistics Program of the University of Southern California.
- Matisoff, James A. 1989. The bulging monosyllable, or the mora the merrier: Echo-vowel adverbialization in Lahu. In Davidson, J. (ed.) *South-East Asian Linguistics: Essays in Honour of Eugenie J. A. Henderson*, 163-97. London: School of Oriental and African Studies.
- Matisoff, James A. 1991. Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20: 469–504.
- Matisoff, James A. 1999. In defense of Kamarupan. *Linguistics of the Tibeto-Burman Area* 22(2): 173–182.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.
- Mills, John P. 1922. *The Lotha Nagas*. London: MacMillan & Co.
- Mills, John P. 1926. *The Ao Nagas*. London: MacMillan & Co.
- No author. 2016. *Yusüp: Bilingual dictionary of Sangtam Naga*. Dimapur, Nagaland: Sangtam Literature Board.
- Reid, Robert. 1942. *History of the frontier areas bordering on Assam, from 1883–1942*. Shillong: Assam Government Press.
- Sampson, Rodney 2011. *Nasal vowel evolution in Romance*. Oxford: Oxford University Press.
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. Orlando FL: Harcourt Brace Jovanovich.
- Temsunungsang, T. 2009. Aspects of the prosodic phonology of Ao: An inter-dialectal study. Unpublished PhD dissertation, The English and Foreign Languages University, Hyderabad.
- Teo, Amos. 2014. *A phonological and phonetic description of Sumi, a Tibeto-Burman language of Nagaland*. Canberra: Asia-Pacific Linguistics.
- Weidert, Alfons. 1987. *Tibeto-Burman tonology: A comparative account*. Amsterdam: Benjamins.
- Witter, W. E. 1888. *Outline grammar of the Lhōtā Nāgā language: With a vocabulary and illustrative sentences*. Calcutta: Supt. of Govt. Print.
- Wolfenden, Stuart Norris. 1929. *Outlines of Tibeto-Burman linguistic morphology*. London: Royal Asiatic Society.