Pre-Print of a paper published in Linguistics of the Tibeto-Burman Area

The sounds of Tawrã (Digaru-Mishmi), a Tibeto-Burman language

Jonathan P. Evans and Johakso Manyu
Institute of Linguistics, Academia Sinica

The present study is a phonological analysis of the segments and tones of the Tawrã language (ISO 69-3: mhu; Glottolog: Diga1241), a Sino-Tibetan language spoken in Arunachal Pradesh, India, and in Tibet, China. This paper, the first collaboration between a Tawrã-speaking non-linguist and a non-Tawrã-speaking linguist, attempts to clear up some confusion in the existing literature. For example, previous studies did not note that stop codas /-p, -k/ are in free variation with glides [-w, -j, ɯ̯], and that the morpheme, rather than the syllable, is the tone-bearing unit. Acoustic analyses provide justification for the phonemic representation of the vowels and the tones. Finally, the paper is designed to introduce Tawrã speakers to the recently standardized (2020) orthography, and to show how the letters and letter combinations function together as a system.

Keywords: Taraon, Digaru, Mishmi, phonology, tone, Arunachal Pradesh

1. Introduction

Tawrã (ISO 69-3: mhu; Glottolog: Diga 1241) is a Sino-Tibetan language spoken in Arunachal Pradesh, India, as well as in Tibet, China, and possibly in Kachin State, Burma (Figure 1).

The language name Tawrã (also spelled Taraon (Luce 1944; Sastry 1984a, 1984b; Pulu 1991; Chakravarty 1963)) is the preferred autonym (Matisoff 1996). Non-Tawrã speakers often use the term Digaru/Digaro (Needham 1978, Konow 1902, Benedict 1972), a name coming from that of the Digaru River. This river roughly forms the western extent of the area traditionally inhabited by the Tawrã speaking population, and probably constituted the location of contact with the people of the plains, who then extended the name of the river to identify the people they met there (Chakravarty 1963: ii). In China the language name is transcribed Dáràng (达让), based on the autonym /tarɯaŋ/ (Jiang, Li & Sun 2013; Sun et al. 1980; Sun 1991).

https://doi.org/10.1075/ltba.20013.eva
Linguistics of the Tibeto-Burman Area 44:1 (2021), **1–27**
ISSN 0731-3500 / E-ISSN 2214-5907

Within Myanmar, the Karaung village (formerly called Arundam) of Northern Putao District, Kachin State is populated by a few hundred people who call themselves Taron (Mya Tu et al. 1966; Rabinowitz 2001).[1] Within Myanmar, the Taron people are considered to belong to the Rawang ethnicity (Nyunt 2004: 2, 2015: 1). Due to a lack of linguistic data, it is not known if the Taron language and/or ethnonym is the same as Tawrã.

The Tawrã language is closely related to Idu, with which it forms a cluster, sometimes called Tawrã-Idu or Digarish (Shafer 1974). Sun (1993) demonstrated a lexical relationship between Tawrã-Idu and the Tani sub-branch of Tibeto-Burman, a claim supported by more recent studies (Post & Modi 2011; Modi 2013). Within China, the larger group is known as Dēng (僜), hence the language is also called Dáràng Dēng (达让僜). The consensus of speakers from India surveyed for this study, as well as the opinion of the Cultural and Literary Society of Mishmi (p.c.), is that they prefer the name and spelling Tawrã.

For the broader ethnic group, the preferred name in India is Mishmi, a term that also includes Idu (ISO 69-3: clk) and Kman (also called Miju, Kamman, or Kaman; ISO 69-3: mxj).[2] Linguistically, there is no particular relationship between Kman and Tawrã-Idu. Tawrã people are often referred to as Digaru Mishmi.

Within Arunachal Pradesh, Tawrã is spoken in Lohit District (Teju and Sunpura Circles, and part of Wakro Circle) and Anjaw District (Chaglagam, Goiliang, and Hayuliang Circles).[3] Within China, the language is spoken between the Dulai River basin and the Zayu (Lohit) River basin, in Chayu (Zayü) County, Tibet Autonomous Region (Jiang et al. 2013: 1). Some Tawrã-speaking areas are claimed by both India and China.

Native speakers in India estimate a population of 15,000 to 20,000. The 2011 census data for India gives a combined population of 33,493 Idu and Digaru Mishmis. Within Zayü County Tibet, the estimated number of speakers ranges from fewer than 1000 speakers (Li, 2002: 1) to as many as 1,500 (Jiang, Li & Sun 2013: 6).

---

1. The authors wish to thank an anonymous reviewer for bringing this to our attention.

2. The Meyor or Zakhring language has been claimed to be closely related to Kman (van Driem 2007), although the people are not considered Mishmi (Aiyadurai 2011). A more recent multilingual comparison concludes "Meyor cannot clearly be assigned to membership of any established branch of Sino-Tibetan" (Blench 2015).

3. Ethnologue (Lewis et al. 2015) also mentions speakers in Dibang Valley District of Arunachal Pradesh, but the speakers consulted for this study claimed that Tawrã-speaking communities there are not indigenous to those locations.

Tawrã lacks a traditional orthography. Development of the orthography documented here involved the assistance of those named in the acknowledgments, and was formally approved by the Cultural and Literary Society of Mishmi on June 28, 2016.

The Tawrã language community is highly multilingual. Speakers surveyed for this study estimate that about 90% of the population are bilingual or multilingual in languages such as Hindi, Nepali, English, and Assamese. Non-Tribal Indians in the area who run businesses, teach, or work in government positions typically speak Hindi. Due to the presence of television sets in homes, children learn Hindi before they begin school, where they are taught in Hindi. Many Tawrã youth typically speak Hindi, even when they are in their home villages. In Tezu and Sunpara, which are near Assam, Tawrã speakers have been bilingual in Assamese for many years. Older speakers tend to have greater command of animal and plant names and traditional stories than do younger speakers.

Figure 1. Locations of Tawrã speaking communities in India and China location of Taron in Myanmar is also indicated

The language used in this study is primarily that of the second author (a male born in 1978, and a lifelong resident of Tezu Circle, Lohit District), along with consultation with Mr. Jabralum Chaitom (born in 1969). Mr. Chaitom is from Hayuliang Circle, near the border with Tibet, and has lived in Tezu for about 15 years. The speakers have given verbal assent to the use of their names, voices, and language data for this study.

In the study which follows, we present the consonants (§ 2), vowels (§ 3), syllable structure (§ 4), and tones (§ 5) of Tawrã. Along the way, we review differences and similarities with previous works; e.g., Chakravarty (1963); Sastry (1984a, 1984b); Jiang, Li and Sun (2013); and Blench (2017, n.d.).

In line with the concerns raised in Rice (2011), an additional purpose of this paper is to present the orthography of Tawrã, which has been refined and standardized in recent years in conjunction with the Cultural and Literary Society of Mishmi (CALSOM), along with other stakeholders. There is an increasing drive among speakers of Mishmi languages to have their writing systems standardized, and to have their languages and writing taught in schools. By writing the forms in italics in Tawrã orthography alongside the IPA transcriptions, the explanation of the phonological structure of the language is made more available to its speakers. In addition, the last part of the paper (§ 6) explains why certain orthographic decisions have been taken.

2. Consonants

We find twenty-six consonant phonemes in Tawrã. All of them occur in word-initial position. Table 1 presents the consonant phonemes, while Table 2 demonstrates all of the consonants (both phonemically and orthographically) in word-initial position before /a/. An exception to this distribution is /ʔ/. The /ʔ/ occurs phonemically in word-initial clusters: /ʔjãᴴɹɑw/ *qyãraw* 'young woman', /jɑ/ *ya* 'night'. The phone [ʔ] is inserted predictably before word-initial vowels: /ɑŋᴸ/ [ʔaŋ²¹] *ang* 'house'. Glottal stop also occurs intervocalically between identical vowels: /gɑᴴʔɑ/ [gɑ˥ʔɑ˧] *gaqa* 'nearby', /tɑᴴᴸʔã/ [tɑʔ˥ ã˩] *taqã* 'light (weight)'. The progressive aspect marker is also /-ʔ/ (see § 2.4).

Table 1. Consonant phonemes

	Bilabial	**Alveolar**	**Palatal**	**Velar**	**Glottal**
	p pʰ b	t tʰ d		k kʰ ɡ	ʔ
Affricate		ts tsʰ dz	tɕ tɕʰ dʑ		
Fricative		s	ɕ		
Nasal	m	n		ŋ	
Approximant	w	ɹ	j		h
Lateral Approximant		l			

2.1 Initial single consonants

Table 2 gives instances of initial single consonants, occurring with /ɑ/.

Table 2. Examples of word-initial consonants

p	pɑ^{L}	pa	axe
pʰ	pʰɑ	pha	bunch of (CL)
b	bɑ^{L}	ba	father
t	tɑpo^{HL}	tapo	banana flower
tʰ	tʰɑ	tha	residue
d	dɑ^{L}	da	communication
k	kɑ^{L}	ka	crocodile
kʰ	kʰɑlɑŋ	khalang	door
g	gɑ^{L}	ga	self
ʔ	ʔwiŋ^{HL}	qwing	old (thing)
ts	tsɑk	tsak ~ tsaü	soak
tsʰ	tsʰaŋ	tshang	rotten
dz	dzɑ^{H}	za	junction
tɕ	tɕɑ	cha	s/he
tɕʰ	tɕʰɑ	chha	Assamese person
dʑ	dʑɑ	ja	what
s	sɑ^{HL}	sa	nerves
ɕ	ɕɑ^{H}	sha	fishing net
m	mɑ^{HL}	ma	mother
n	nɑ^{L}	na	leaf
ŋ	ŋɑ^{HL}	nga	fall down
w	wɑ^{HL}	wa	wound
ɹ	ɹɑ	ra	sharp
j	jɑ	ya	night
h	hɑ^{H}	ha	thigh
l	lɑ	la	tell

The glide [ɯ] *ü* occurs both as an allophone of /g, k/ in word-final position (e.g., /kwɑg/ ([kwɑɯɟ] *kwaü* 'dog'; Table 3), and as an allophone of /w/ after /p, b/ (/pwi/ [pɯi-ɟ] *püi* 'bear child', /bwi^{L}/ [bɯi˩] *büi* 'dance'; Table 3).

We note some differences between this analysis and those found in the literature. The analysis of consonants found in Chakravarty (1963) is almost identical to that of Table 1, with the exception that the alveolar and post-alveolar affricate series are conflated into one: {c, ch, j}. Table 2 shows that, for the speakers consulted in the study, these series of sounds are distinct.

The primary difference between our analysis and that of Sastry (1984a, 1984b) is the lack of phonemic /ʔ/ in his analysis. He only describes [ʔ] as occurring

allophonically in word-initial position before vowels. However, for our consultants, /ʔ/ occurs in clusters, and even as a verbal affix /-ʔ/ to mark progressive aspect, and a nominal affix to mark definiteness. He distinguishes /z/ vs /dz/, which our consultants do not, which could be due to variation between speakers. Finally, Sastry appears to have a strategy of positing as few indivisible consonant phonemes as possible, and then building up the remaining segments via compounding. Thus, {c, j}, which stand for our /ts, dz/, are built up via compounding to form {ch} /tsʰ/, {cy} /tɕ/, etc. However, we treat all affricates as unitary phonemes /ts, tsh, dz, tɕ, tɕh, dʑ/.

The main difference between our consonantal analysis and that of Jiang et al. (2013) is that some sounds that we treat as consonant clusters, are analyzed in that source as single consonants. For example, sounds that we treat as clusters Cw and Cj are considered by Jiang et al. to consist of labialization and palatalization of initial consonants (C^{w}, C^{j}). We prefer the cluster analysis rather than adding seventeen labialized and palatalized consonants to the inventory (Section 2.2, Table 3).

Jiang et al. record the presence of retroflex initials /tʂ, tʂʰ/, but these seem to only occur in the borrowings /tʂoŋ^{55}tʂoŋ^{55}/ 'grey goose' (2013: 26, 282; cf. Tibetan *khrung khrung* 'crane'), /go^{51}tʂʰi^{55}/ 'leader' (p 26; cf. *mgo khrid* 'leader').[4]

Blench (n.d.) proposes a similar set of sounds to ours, but proposes that [ts, dz] "are probably in free variation with" [tɕ, dʑ], while speakers consulted for this study contrast affricates at two places of articulation (Table 2). He identifies a labio-dental approximant /ʋ/, for which we do not find evidence; it might indicate individual variation in the production of /w/. He also writes that "voiceless /h̥/" is not phonemic, but all other sources find reason to posit /h/; e.g., /hɑlo/ 'moon' and forms in Table 2.

2.2 Consonant clusters

Tawrã initial consonant clusters have the form CG, where C is either an oral stop (including /ʔ/ but not /tʰ/), or one of /m, w, l, h/ *m, w, l, h* (Table 3). The affricate /tsʰ/ *tsh* can also serve as an initial consonant, where it seems to take the place of /tʰ/ *th*. The second element in clusters is a (G)lide consonant, including /ɹ l w j/ *r, l, w, y*. The glide /j/ has the widest distribution, occurring with all initials except /tsʰ w h/. The glides /l, w/ have similar distributions, occurring primarily after labial, velar and glottal initials. Many possible combinations are not attested.

The sounds /hl, hɹ, hw/ are treated as clusters rather than as unitary phonemes for phonological reasons. These sounds do not occur word finally, unlike /-r -l -w/. They also do not occur as initial consonants in a cluster, unlike /w- l-/. Furthermore,

4. Jackson T.-S. Sun, p.c.

it is more economical to add another type of consonant cluster, those starting with /h-/, than to add three typologically unusual consonant phonemes /ɻ̥ l̥ w̥/.

Although glottal stop [ʔ] occurs predictably in the context [#_V], it occurs phonemically word-initially before glides; e.g., [ʔwiŋ˩] 'old (thing)' in Table 2. Although no exact minimal pairs have been found contrasting [#ʔG] and [#ØG], the near minimal pair in Figure 2, along with discussions with speakers, suggests that pre-glide occurrence of [ʔ] is contrastive. The phonemic status of /ʔ/ is explored further in the discussions of consonant codas.

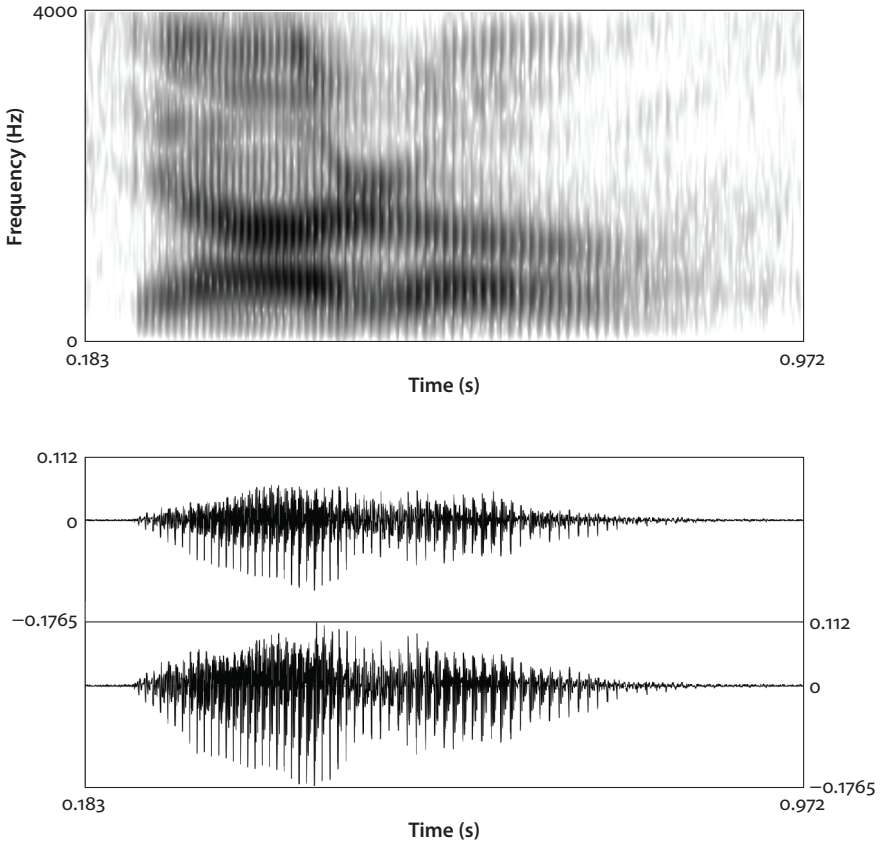

Figure 2a. Spectrogram and waveform of [ʔjã˥ɾɑw˧] qyãraw 'young woman'. Abrupt onset of spectrogram reflects glottal stop release

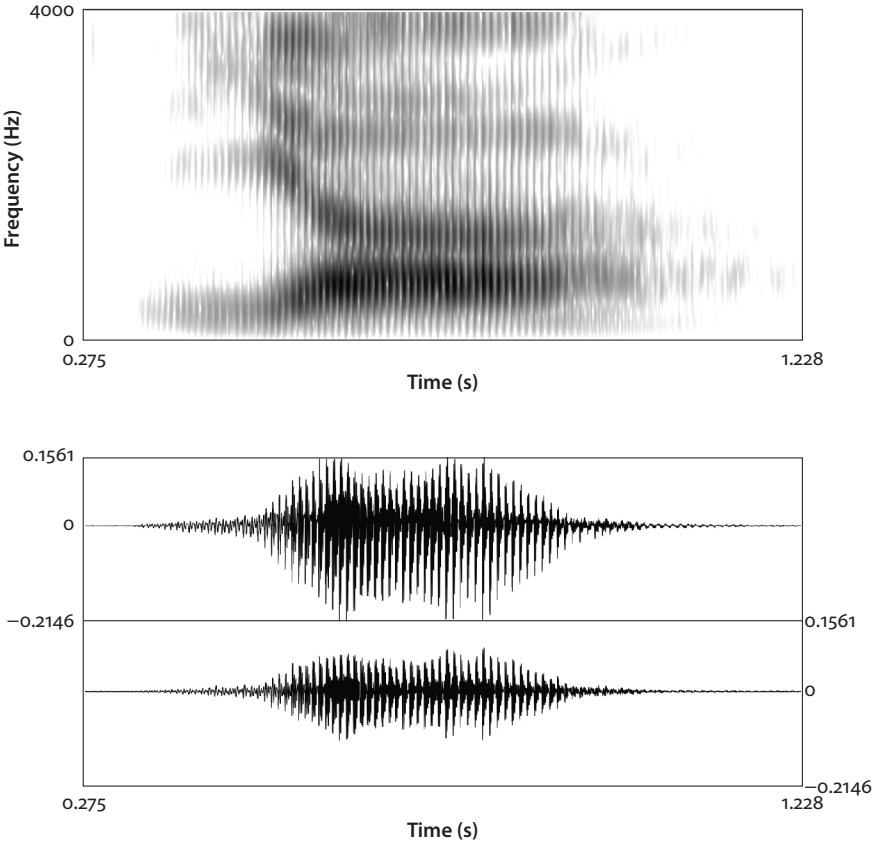

Figure 2b. Spectrogram and waveform of [ja˧] ya 'night'. Gradual darkening of voicing bar at bottom of spectrogram reflects gradual onset of voicing

In [ʔjã˥ɾaw˧] *qyãraw* 'young woman' (Figure 2b), the onset of voicing is more abrupt than it is for [ja˧] *ya* 'night' (Figure 2a), as indicated by the oval in Figure 2b.

Table 3. Word-initial consonant clusters

	-ɹ-	-l-	-w-	-j-
p-	/pɹɑ/	/plɑ^{L}/	/pwi/ [puɥi˧]	/pjɑ^{HL}/
	pra	*pla*	*püi*	*pya*
	'fine'	'salt'	'bear child'	'bird'
pʰ-	/pʰɹɑ̃/	/pʰlɑ̃^{L}/		/pʰjɑk/
	phrã	*phlã*		*phyak ~ phyaü*
	'sap'	'stone'		'pointed'

Table 3. (*continued*)

	-ɹ-	-l-	-w-	-j-
	/bɹɑ/ *bra* 'father-in-law'	/blɑj/ *blay* 'tears'	/bwi^{L}/ [bɯiɻ] *büi* 'dance'	/bjɑ/ *bya* 'flattening of crops by storm'
t-	/tɹu^{HL}/ *tru* 'meet'			/tjɑ/ *tya* 'right now'; 'grandfather'
d-	/dɹɑk^{HL}/ *drak ~ draü* 'pineapple'			/djɑ/ *dya* 'far'
tsʰ-	/pjɑ^{HL} tsʰɹi/ *pya tshri* 'small bird (sp.)'			
k-	/kɹɑ̃/ *krã* 'hollow'	/klɑj^{HL}/ *klay* 'excreta'	/kwɑg^{HL}/ *kwak ~ kwaü* 'dog'	/kjɑ^{HL}/ *kya* 'see omen'
kʰ-	/kʰɹɑ̃^{L}/ *khrã* 'spacious'			/tɑ khjã/ *ta khyã* 'place that collects water'
g-	/gɹɑ^{L}/ *Gra* 'call'	/glɑj/ *glay* 'break (vi) of branch'	/gwɑk^{H}/ *gwak ~ gwaü* 'priest'	/tɑ^{L}gjã/ *ta gyã* 'spark'
ʔ-	/ʔɹɨ^{L}/ *qrü* 'pry'		/ʔwiŋ^{L}/ *Qwing* 'old (of thing)'	/ʔjã^{H}ɹɑw/ *qyãraw* 'young woman'
m-	/mɹɨm^{L} ~ mɹum^{L}/ *mrüm ~ mrum* 'about to storm'	/mlãʔ^{H}/ *mlã* 'hum'		/mjɑ^{L}/ *mya* 'hand-width used to measure pigs'
w-	/wɹɑ̃^{L}/ *wrã* 'boat'			
l-				/ljaŋ^{HL}/ *lyang* 'evening'
h-	/hɹɨ^{HL}/ *hrü* 'become sour'	/hliŋ^{HL}/ *hling* 'nail'	/hwɑ^{L}/ *hwa* 'doubt'	

The phoneme /ɹ/ occurs as a glide initially, as in [ɹɑ˧] *ra* 'sharp'. However, in obstruent-initial clusters, it is pronounced as a tap:/dɹɑkᴴᴸ/ [dɾʌkˋ] *drak* 'pineapple', /pjɑᴴᴸ tsʰɹi˧/ [pjɑˋtsʰɾi˧] *pya tshri* 'small bird (sp.)'. Intervocalically, /ɹ/ can occur as either an approximant or a tap: /tɑɹɑᴴ/ [tɑ˧ɹɑ˥] ~ [tɑ˧ɾɑ˥] *tara* 'long knife'.

With regard to the analysis of consonant clusters in previous publications, Chakravarty (1963) and Blench (n.d.) do not explicitly discuss consonant clusters. Sastry (1984a, 1984b) finds many of the same clusters that we do, although since he treats aspirated stops as clusters, he misses combinations like /pʰl-/, /kʰɹ-/, etc. Jiang et al. (2013) record similar clusters as we have found, except that clusters with /-w-, -j-/ are recorded as labialized and palatalized consonants, respectively. They also do not record clusters with initial /t, w, ʔ/.

2.3 Geminate consonants

Blench (n.d.) seems to be the first author to note geminate consonants in Tawrã. He claims tautomorphemic geminate /lː, nː/, both word-initially and word-medially. In borrowed words with geminates (e.g., /kɑtsɑbᴴbɑ/ 'tortoise') it is not clear whether they are phonologically geminated, or if they consist of codas followed by identical initials; however, we are only concerned with native vocabulary in the present study. In our investigation, we find [lː] as an allophone of /l/ after /i/ (Table 4).

Table 4. Geminate [lː] after /i/

[lː]		[l]	
[hɨl˧lɑˋ]	'leaves'	[tɑ˧lɑˋ]	'musk deer'
[hɨl˧lɑ˥]	'underneath'	[tɑ˧lɑ˥]	'length'
[hɨl˧lɑ˩]	'misguide, instigate'	[tɑ˧lɑ˩]	'male bird'
[bɨl˧li]	'clan title'		
[bɨl˧liˋ]	'excessive'		

Other than [lː], consultation with native speakers does not yield evidence for tautomorphemic geminate consonants, whether word-medial or -initial.

2.4 Codas

Phonemically, Tawrã codas consist of a subset of stops and nasals, subject to various constraints. First, there are no alveolar codas; e.g., *-t, *-d, *-l, *-n *-ɹ. Second, /ʔ/ only occurs as a coda after an intervening morpheme break: /suᴴᴸ/ *su* 'boil', /su-ʔᴴᴸ/ *suq* 'boiling' (progressive aspect). Third, obstruent codas are both devoiced and unreleased. The underlying voicing status of final consonants can be ascertained by examining suffixed forms, such as (1), (2) from Sastry (1984b: 92):

(1) /p/ [kalja̱p] ‘field’ [kalja̱p-gũ][5] ‘from the field’

(2) /b/ [tedʑa̱p] ‘Tezu’ [tedʑa̱b-gũ] ‘from Tezu’ (city in Assam)

Instances of glides [-j, -w, -ɯ] occur as free variants of /-p, -k/, as can be seen in Table 5. Labial stops /p, b/ surface as [w] (‘elbow’, ‘Tezu’), while final /k, g/ surface as [j] after front vowels /i, e/ (‘son’s child’, ‘buffalo’), and surface as [ɯ] after /a/ (‘soak’, ‘dog’).

Table 5. Codas in underlying and surface forms

	Underlying	Variant 1	Variant 2	Orthography	Gloss
/-ap/	/laᴴkɹap/	[la˥kɹap˦]	[la˥kɹaw˦]	*lakrap ~ lakraw*	‘elbow’
/-ab/ [p]	/tedʑab/	[te˦dʑap˦]	[te˦dʑaw˦]	*Tezu*	‘Tezu’ (city)
/-ik/	/aᴴjik/	[a˥jik˦]	[a˥jej˦]	*ayik ~ ayey*	‘son’s child’
/-ek/	/madʑekᴴ/	[ma˦dʑek˥]	[ma˦dʑej˥]	*majik ~ majei*	‘buffalo’
/-ak/	/tsakᴴ/	[tsak˥]	[tsaɯ˥]	*tsak ~ tsaü*	‘soak’
/-ag/	/kwag/	[kwak˦]	[kwaɯ˦]	*kwak ~ kwaü*	‘dog’
/-m/	/amᴴ/	[am˥]	--	*Am*	‘cloud’
/-ŋ/	/aŋᴸ/	[aŋ˩]	--	*Ang*	‘house’

Previous works do not seem to identify the free variant relationship between final stops and glides. For example, Jiang et al. (2013: 32, 33) has /gʷak[35]/ ‘priest’ and /kʷaɯ[53]/ ‘dog’, where we find /gwakᴴ/ [gwak˥ ~ gwaɯ˥] and /kwag/ [kwak˦ ~ kwaɯ˦].

Otherwise, our analysis of codas is essentially the same as that found in Sastry (1984a, 1984b) and Jiang et al. (2013). Chakravarty (1963) and Blench (n.d.) do not explicitly enumerate possible codas. Jiang et al. (2013) give an example of tautomorphemic coda /ʔ/ *q*, which we do not find among the speakers consulted for this study, although we do find a definite marker and a continuous aspect marker that are both pronounced /-ʔ/ *-q*.

As of this time, we do not have an explanation for the lack of alveolar codas. The constraint is especially striking, given that coronal consonants are extremely common in the world’s languages (Lindblom & Maddieson 1988; Mielke 2009). As observed by an anonymous reviewer, the Aoic language Northern Sangtam, spoken in Central Nagaland shares this constraint (Coupe 2020).

5. Forms transcribed according to our phonologization

3. Vowels

Tawrã speakers contrast six vowels, contrasting front, central, and back places of articulation and two height distinctions: /i, e, ɨ̣, ɑ, u, o/. The vowel transcribed /ɨ̣/ often sounds like the high back unrounded vowel [ɯ]. The transcription /ɨ̣/ preserves phonological symmetry; for example, nasalization occurs on one vowel at each degree of backness (§ 3.1). It seems to perform in some ways like the featureless vowel /ə/; e.g., emerging between the first and second consonants in a sesquisyllabic word (see § 4). Nevertheless, an articulatory target like [ɨ] or [ɯ] seems to be present (Figure 3).

The vowel phonemicization in this study is shared by Chakravarty (1963), Sastry (1984a, 1984b), and Jiang et al. (2013), with only transcriptional differences. Blench (n.d.) proposes some additional vowels, but they do not appear to be distinctive (minimal pairs not given). Table 6 exemplifies the six vowels after /b/, /s/.

Table 6. Tawrã vowel contrasts

i	/bi/	*bi*	'elder sister'	/si/	*si*	'bird claw'
e	/ɑbe^{H}/	*abe*	'much later on'	/se/	*se*	'choose'
ɨ̣	/bɨ̣/	*bü*	'be blown'	/mɑ^{L} sɨ̣/	*masü*	'have a cold'
ɑ	/bɑ^{L}/	*ba*	'father'	/sɑ^{HL}/	*sa*	'nerve'
u	/bu/	*bu*	'pus'	/su^{HL}-jɑ/	*suya*	'to boil'
o	/we^{H}-bo^{HL}/	*webo*	'and then'	/ɑso^{H}/	*aso*	'fat'

For the purposes of acoustic analysis, the vowel contrast examples were each pronounced twice by the two speakers. Recordings were made on a Zoom H4n recorder at 44.1 kHz sampling rate and standard settings. Each vowel utterance was sampled across the stable portion of the spectrogram using standard settings in Praat, from which formant means were calculated. Vowels following /b/ and /s/ were acoustically similar in F1 × F2 vowel space, and the vowel spaces for the two speakers were also similar. Means across the vowel instances for both speakers are plotted in Figure 3a, b.

As shown in Figure 3, vowels tend to have their expected IPA values. However /ɑ/ raises to [ʌ] before final /k/ (/t^{h}ɑk^{HL}/ [t^{h}ʌkˀ] *thak* 'bite'). However, the presence of [w] before /ɑ/ blocks this raising (/kwɑk^{HL}/ [kwɑkˀ] *kwak* 'dog').

Tawrã vowels do not occur with identical lexical frequency and phonological distribution. The mid vowels /e o/ *e, o* occur with less lexical frequency than the other vowels, while /ɑ/ *a* is most common. The same phenomenon has been noted in the dialect spoken in China (Jiang et al. 2003: 29).

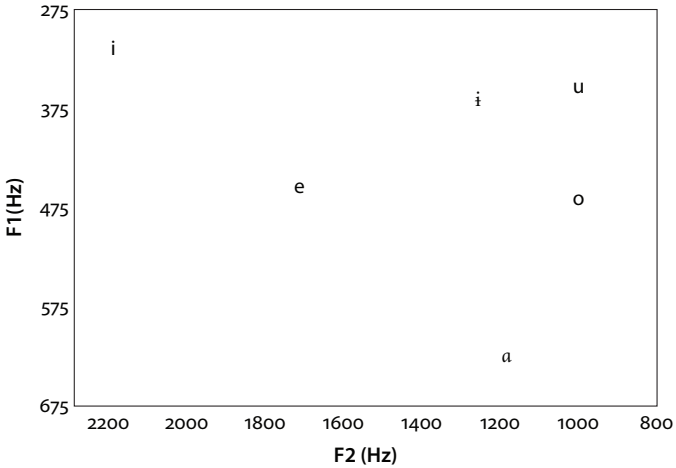

Figure 3a. Tawrã oral vowel formants (second author, JM)

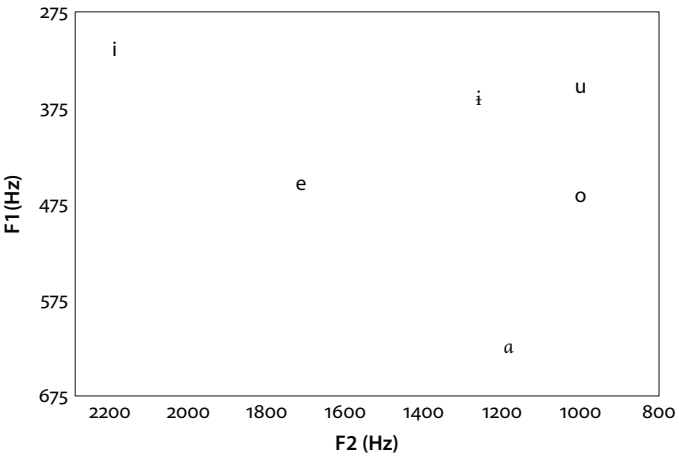

Figure 3b. Tawrã oral vowel formants (JM)

Tawrã has distinctive vowel length, as can be seen in Table 7.

Table 7. Lexically distinctive vowel length

/pu^{HL}/	[pu˥˩]	*pu*	'sour'
/puː/	[puu˧]	*puu*	'pound'
/kapiŋ^{HL}/	[ka˧piŋ^{H}˩]	*kapüng*	'bat'
/kaː^{L}piŋ/	[kaː˩piŋ˧]	*kaapüng*	'temple of head'

Sastry (1984) identifies allophonic vowel lengthening on single mora words, and no phonemic lengthening. Chakravarty (1963: iv) identifies the same phenomenon, but also lists exceptions. We find evidence that vowel length can be lexically specified (Table 7). On nouns, vowel lengthening with tone change relates to referential marking, as shown in (3), (4). The exact changes in tone caused by referential marking remain to be fully understood.

(3) /**aː** ktɕʰaŋ/
[aː˧ ki˧tɕʰaŋ˧]
child:REF urine
'The child urinates'

(4) /**a**^{HL} ktɕʰaŋ/
[a˥˩ ki˧tɕʰaŋ˧]
child urine
'A child's urine'

On verbs, lengthening of the vowel of the root is part of negative marking (5). Examples (6), (7) show that the verb 'cook' has a short vowel by default.

(5) tɕa habɹa **pʰlãː**-jim.
3s millet cook-NEG
Cha habra phlããyüm
'S/he is not cooking millet.'

(6) tɕa habɹa **pʰlã**-ʔ
3s millet cook-PROG
Cha habra phlãq
'S/he is cooking millet.'

(7) tɕa habɹa **pʰlãᴴ**-ja la
3s millet cook-FUT say
Cha habra phlãya la
'S/he says s/he will cook millet.'

Blench (n.d.: 4) mentions vowel length, but does not give minimal pairs; Jiang et al. (2013) do not mention vowel length.

3.1 Vowels and nasality

A reduced set of vowels occurs before nasal finals; specifically, front /i, e/ and back /u, o/ vowels do not contrast height when followed by a nasal coda. Furthermore, /ɨ, u/ do not contrast before /m/, but only /ɨ/ precedes /ŋ/ (Table 8).

Table 8. Vowel contrasts before nasal codas

	/-m/	***-m***		**/ŋ/**	***-ng***	
i	*			/tapiŋ/	*taping*	'residue of grain'
e	/pem/	*pem*	'anyway'	*		
ɨ	/tapɨm^{H} ~ tapum^{H} /	*tapüm ~ tapum*	'insect'	/kapɨŋ^{HL}/	*kapüng*	'bat'
a	/pam^{H}/	*pam*	'wasp'	/paŋ^{H}ja^{HL}/	*pangya*	'finish'
u	/tamjum^{HL} ~ tamjɨm^{HL}/	*tamyum ~ tamyüm*	'monkey'	*		
o	*			/poŋ/	*pong*	'bomb'

Before /m/, nuclear vowel /e/ surfaces with a lower sounding vowel [pɛm] ~ [pæm] *pem* 'anyway'.

Of the six oral vowels, Tawrã contrasts one nasal vowel at each degree of backness: /ẽ ã ũ/ *ẽ, ã, ũ*; there is no height contrast in nasal vowels, just as there is a diminished height contrast among oral vowels followed by nasal consonant. The back nasal could be written /õ/ in order to preserve orthographic consistency in phonological height (as in Sastry 1984a, 1984b); however, because of auditory resemblance to a nasalized high back vowel, it is transcribed /ũ/ *ũ*. Table 9 shows the oral/nasal vowel contrasts:

Table 9. Contrast of oral and nasal vowels

/tatsʰa^{L}/	tatsha	'stinging nettle'	/ka^{L}se/	kase	'choose'	/tsʰu^{L}/	tshu	'add insult to injury'
/tatsã^{H}/	tatsã	'flesh between bone and skin'	/tasẽ^{H}/	tasẽ	'small trowel'	/tsʰũ^{L}/	tshũ	'become brighter'

Figure 4 plots the vowels in the above set, using the same method as described for oral vowels above, and plotted on the same scale as the oral vowels.

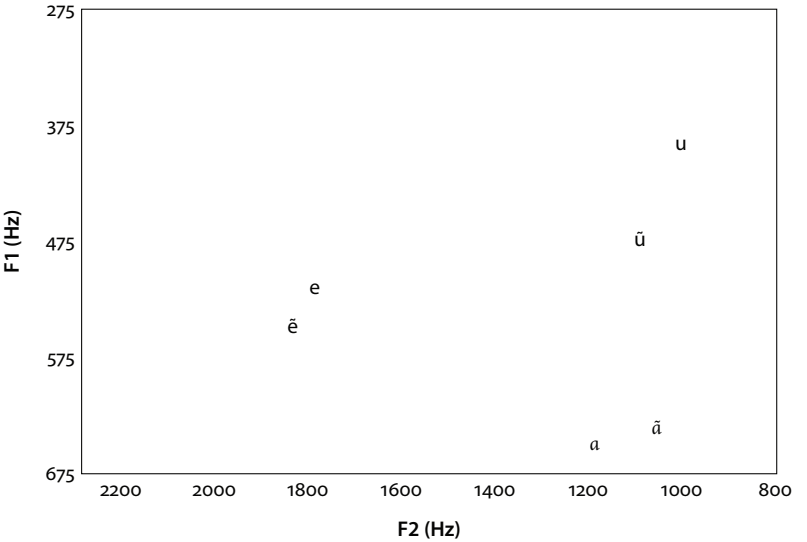

Figure 4a. Oral and nasal vowel formants (JM)

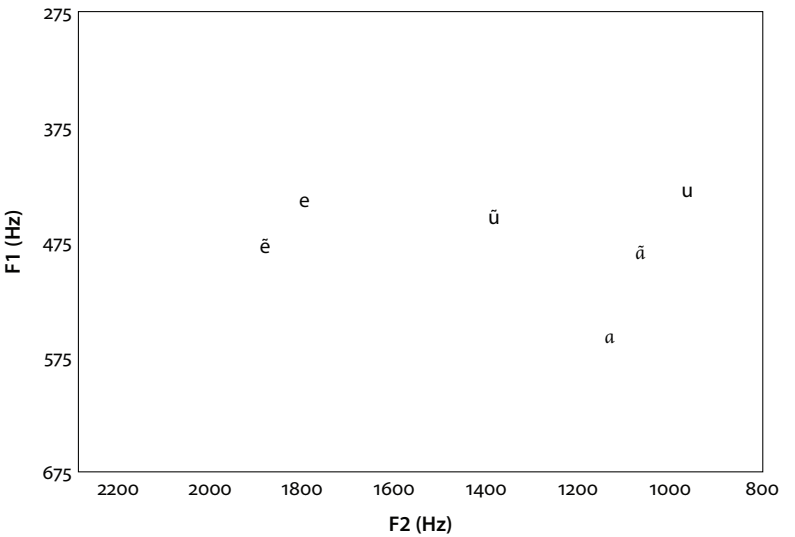

Figure 4b. Oral and nasal vowel formants (JC)

Figure 4. Values of mean F1 and F2 of nasal vowels and their oral counterparts (words taken from Table 6)

Coda consonants do not follow nasal vowels, suggesting that vowel nasality comes from a loss of final */-n/, which is not an attested coda. Some Tawrã words with final nasal appear to have inherited a nasal coda from Proto-Tibeto-Burman: /hɑ.ɹɑŋᴸ/, *harang*, *(s)ram 'otter'; /ɑhliŋᴴᴸ/, *ahling*, *luŋ, 'liver' (reconstructions from Matisoff 2003). Identification of additional reflexes of Proto-Tibeto-Burman *nasals is dependent on more historical work being done on the language.

Descriptions in the literature of Tawrã vowels are essentially the same as has been recorded here. Sources vary on whether to transcribe a high central vowel /ɨ/ (Sastry 1984a, 1984b; Blench, n.d.) or a high back unrounded vowel /ɯ/ (Jiang et al. 2013). Chakravarty (1963) transcribes {í} and describes the sound of [ɯ]. Impressionistically, the vowel in question does sound like [ɯ]. In the F1 x F2 space, the vowel occurs in a more acoustically front position than [u, o] (Figure 3a, b; Jiang et al. 2013: 29). Phonologically, the choice to transcribe /ɨ/ allows for a more symmetrical vowel system than does /ɯ/.

Sastry (1984a, 1984b), Jiang et al. (2013) and our study all identify a total of six vowels phonemes. Chakravarty (1963) transcribes seven vowels, using separate letters for /ɑː, ɑ/. Blench (n.d.) describes an additional height level of open-mid vowels /ɛ, ə, ʌ, ɔ/, however, minimal sets are not provided; hence, the differences could be allophonic, or due to speaker variation in pronunciation.

With the exception of Jiang et al. (2013), previous studies also posit nasalized vowels. Chakravarty (1963) and Sastry (1984a, 1984b) propose the same three vowel system observed here, with the exception of transcribing /õ/ where we have /ũ/. Blench (n.d.) proposes additional nasalized vowels, but does not demonstrate their contrast.

Within the monomorphemic syllable there are sequences that could be interpreted as either vowel-vowel sequences (true diphthongs), or as vowel-glide sequences. All such sequences in Tawrã involve at least one sound that could be analyzed either as a high vowel or as an approximant. Because there are no monomorphemic syllable-internal sequences of two non-high vowels, it has been decided to treat these sound combinations as glide-vowel (GV) or vowel-glide (VG) sequences, rather than as two vowels. In further support of the glide analysis is the fact that final stops and nasals do not occur tautomorphemically after final glides, which suggests that they occupy the coda slot. For both GV and VG, there are no sequences where the vowel is homorganic with the glide; e.g., */wu/, */ij/, etc., with the exception of variant forms; e.g., [gji˥hwã˩] 'sweet potato' (Table 10). The forms in Table 10 demonstrate possible GV sequences, which are also partly exemplified within Table 3.

Table 10. Glide-vowel sequences

	/j/	*y*	
/i/	*		
/e/	/tɑpjeᴴ/	*tapye*	'female animal'
/ɨ/	/bjɨᴴ/	*byü*	'run'
/ɑ/	/njɑ/	*nya*	'face'
/u/	/gjũᴴᴸhwã/ ~ [gjiˀ hwã˩]	*gyũhwã, gyihwã*	'sweet potato'
/o/	/pjoᴸ/	*pyo*	'to paste, paint'
	/w/	***w***	
/i/	/kwi/	*kwi*	'heat'
/e/	/twe/	*twe*	'tie loosely'
/ɨ/	*		
/ɑ/	/tɑᴸwɑ/	*tawa*	'honeybee'
/u/	*		
/o/	/woᴴᴸ/	*wo*	'snag, hang'

4. Syllable structure

Sastry (1984a, 1984b) and Chakravarty (1963) claim that lexical words have a minimum of two morae. However, we find a difference in morpheme count between /ɑ/ [ɑˤ] *a* 'child' (4) and /ɑː/ [ɑːˠ] *aa* 'child:REF' (3).

The maximal Tawrã monomorphemic syllable is CGVX(T), where C can be any of the sounds in the consonant table, except that, as shown above, /ʔ/ *q* only occurs word-initially before glides (except predictably and non-distinctively before vowels). G can be any of /r l w j/ *r, l, w, y, ü*, and X can be either a glide consonant, nasalization on the vowel, a nasal consonant, or a stop (represented here by K). Lexical morpho-syllables must contain a vowel and a tone; mid tone is assigned by default and is not marked phonemically. The syllable structure VK has not been attested. Attested syllable types are exemplified in Table 11.

Table 11. Syllable types

V (T)	/ɑ/	[ʔɑ³³]	*a*	'child
VN (T)	/ɑŋᴸ/	[ʔɑŋ²¹]	*ang*	'house'
CVX (T)	/gã/	[gã³³]	*gã*	'wheel'
CVN (T)	/ɹiŋ/	[ɹiŋ³³]	*ring*	'sun'
CVK (T)	/tʰɑkᴴᴸ/	[tʰʌk⁵²]	*thak*	'bite'
CVG (T)	/thɑᴴᴸ-l̲ij̲/	[tʰɑ⁵⁵-lij²¹]	*tha-lüy*	*eat*-PFV ('S/he ate.')
CV (T)	/pu/	[pu³³]	*pu*	'pound with fist'

Table 11. (*continued*)

CGVX (T)	/pʰlãᴸ/	[pʰlã^{21}]	*phlã*	'stone'
CGVN (T)	/ljaŋᴴᴸ/	[ljaŋ^{52}]	*lyang*	'evening'
CGVK (T)	/kwagᴴᴸ/	[kwɑk^{52}]	*kwak*	'dog'
CGVG (T)	/sjejᴴᴸ/	[sjej^{52}]	*shey*	'fruit'
CGV (T)	/njɑ/	[njɑ^{33}]	*nya*	'face'

In addition to full syllables, there are also unstressed, neutral vowel /ɨ/ syllables that occur in word-initial position, yielding a "sesquisyllabic" structure (Matisoff 1973: 86). For consonant sequences that can form clusters, there are then three degrees of temporal proximity, as can be seen with sequences of /k/, /l/ (Table 12).

Table 12. Degrees of juncture between consonants

consonant cluster:	/klɑjᴴᴸ/	*klay*	'excreta'
light syllable:	/kɨlɑᴴᴸ/	*küla*	'official (n.)'
full syllable:	/kɑlɑŋᴴ/	*kalang*	'long'

In some cases there is free variation, as in /tɨlɑᴴᴸ ~ tɑlɑᴴᴸ/ *tüla ~ tala* 'musk deer'.

5. Tones

Tawrã, like numerous Tibeto-Burman languages of Northeast India, has up to one tone per lexical morpheme, which can consist of more than one syllable (Morey 2008; Post 2015). Tawrã tonal phenomena can be described with reference to three tonemes, H(igh), L(ow) and HL (falling). Not every morpheme has a specified tone. If no tone is assigned to a syllable (Ø), then it receives a default M(id) pitch. For example, the toneless initial syllables of sesquisyllabic words are realized on the default Mid pitch. The Mid pitch assignment itself is not a toneme. Tones are transcribed phonologically with raised H, L, or HL. The purpose of this transcription is to clearly communicate the tonal categories both to linguists and to Tawrã speakers. The International Phonetic Association (1999) recognizes two types of pitch notation, a series of diacritics above the corresponding letter (e.g., [á] for high pitch) and Chao tone letters based on the concept of height on a musical staff (e.g., [a˥] for high pitch). In our phonetic transcriptions, we use the Chao tone letters to more transparently indicate the phonetic realization of the tone categories.

Table 13 shows the distribution of the three tonemes on Tawrã mono- and disyllabic morphemes, along with the pitch patterns of toneless morphemes. There are only a few possible tone assignments on trisyllabic morphemes (Table 14); morphemes longer than three syllables have not been attested.

Table 13. Tonal possibilities on mono- and disyllabic morphemes

	H	HL	Ø	L
	/hɑᴴ/ [hɑ↗] 'thigh'	/njɑᴴᴸ/ [njɑ↘] 'wool strip'	/njɑ/ [njɑ→] 'face'	/njɑᴸ/ [njɑ↘] 'feel pain'
Syll 1	/gɑᴴ.ʔɑ/ [gɑ↗.ʔɑ→] 'nearby'	/suᴴᴸ-jɑ/ [su↘-jɑ↓] 'to boil'	/hɑlo/ [hɑ→lo→] 'moon'	/kɑᴸdiŋ/ [kɑ↓diŋ→] 'star'
Syll 2	/hɑdzɑᴴ/ [hɑ→dzɑ↗] 'king'	/tɑmjumᴴᴸ/ [tɑ→mjum↘] 'monkey'		/tɑtɕʰoŋᴸ/ [tɑ→ tɕʰoŋ↘] 'furniture'

For both H and L tones on monosyllables, the speaker's pitch often begins in the mid-range and then slides upward or downward toward the target ('thigh', 'feel pain'). Low-toned syllables tend to have creaky voice. In some cases, creaky voice surfaces during high-falling syllables. The high-falling tone /ᴴᴸ/ typically descends to the mid range [↘], when it occurs in a pre-pausal position ('wool strip', 'monkey'), but can fall to Low when there is a following syllable on which the L is realized ('to boil'). The four Tawrã tones on monosyllables are represented in the following time-normalized figure (Figure 5). All F0 figures in this study are time-normalized averages of three utterances spoken in isolation by the second author. Time-normalized data were used because it has not been possible to find sets of words that differ only by tone, and that exemplify the four categories. Thus, there are segmental differences which could affect duration, skewing the results. Isolation utterances were used due to the communication styles of consultants (frame sentences were produced inconsistently). For the purpose of recording, words were spoken individually in order to avoid list intonation.

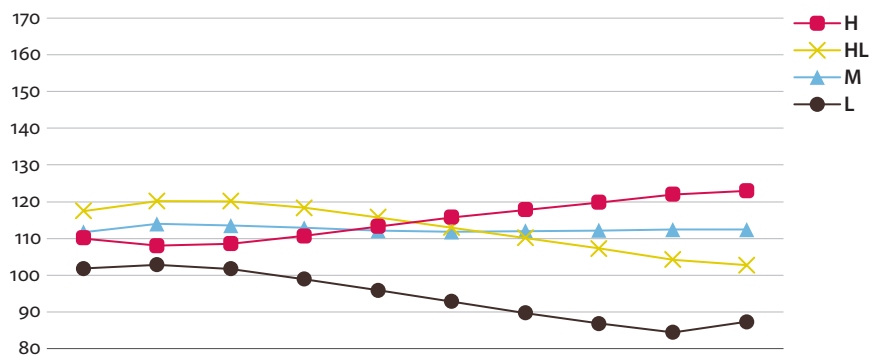

Figure 5. F0 tracing of contrasting tones in Hz on monosyllables containing [ɑ] (three utterances by JM)

For disyllabic morphemes, one syllable in a word may be marked with a tone, unless both syllables are toneless, and thus have a Mid pitch by default (/hɑlo/ 'moon'). Figures 6a and 6b show pitch patterns on the abovementioned disyllables with tone on the first or second syllable. The default pattern of Ø-Ø (Mid-Mid) is given in both figures. Voicing of sounds produces F0 (acoustic pitch); thus, every voiced syllable has F0. From a phonological perspective, only three tone categories (plus toneless) are required to delineate the pitch differences that can distinguish meaning.

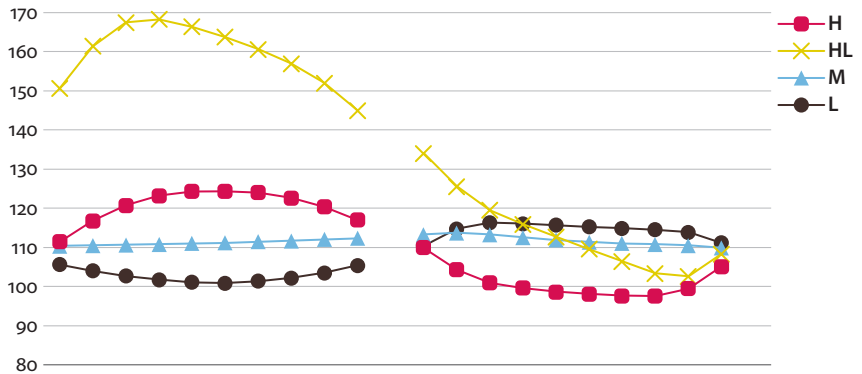

Figure 6a. F0 tracing of contrasting tones in Hz on disyllables with tone on the initial syllable (three utterances by JM)

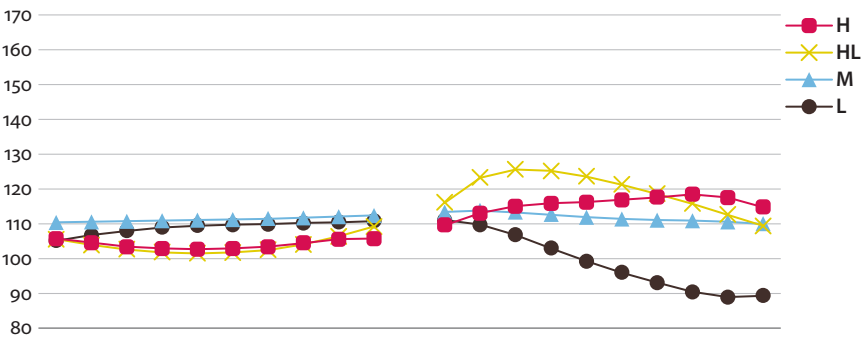

Figure 6b. F0 tracing of contrasting tones in Hz on disyllables with tone on the second syllable (three utterances by JM)

Trisyllables, which are not common, have only three patterns, in which either the first syllable has a H tone, or the second syllable has either a H or a HL tone (Table 14).

Table 14. Tawrã trisyllabic tone patterns

HL on σ1	/tʰaᴴᴸkala/	[tʰa˥˩ka˧la˨˩]	*thakala*	'shadow'
H on σ2	/katsabᴴba/	[ka˧tsab˥˩ba˧]	*katsabba*	'tortoise'
HL on σ2	/adajᴴᴸtjaŋ/	[a˧daj˦˩tjaŋ˨˩]	*adaytyang*	'scorpion'

High tone on the first syllable is the most common tone pattern on trisyllables. Falling tone on the second syllable yields a L tone on the final syllable ('scorpion'). The form 'tortoise' has medial geminate /bb/ *bb*, a sound which has not been attested elsewhere in the lexicon, but is clearly articulated nonetheless. The word appears to have been borrowed from Indo-Aryan (e.g, Hindi कछुआ /kɑtʃʰuɑ/), which could account for the lone occurrence thus far of geminate /-bb-/.

When morphemes are combined to form words, they often retain their tone specifications: /weᴴ-boᴴᴸ/ *webo* 'and then' from /weᴴ/ 'that' and /boᴴᴸ/ 'go'. Similarly, /pjɑᴴᴸ/ 'bird' and /lãᴸ/ 'tooth' combine to form /pjɑᴴᴸ lãᴸ/ 'beak'. For this reason, disyllabic compounds and disyllabic morphemes have different surface pitch patterns, even if the underlying tones are the same; e.g., monomorphemic /tɑᴴᴸʔã/ [tɑ˥ʔã˨˩] *taqã* 'light (weight)', in which the L tone of HL is assigned to the second syllable, versus the compound /pjɑᴴᴸ-tsʰɹi/ [pjɑ˦˩tsʰɹi˧] *pya tshri* 'small bird (sp.)' from /pjɑᴴᴸ/ [pjɑ˦˩] *pya* 'bird' and /tsʰɹi/ [tsʰɹi˧] *tshri* 'bird (sp.)', where the second morpheme retains its toneless specification, and surfaces with a Mid pitch.

Not all compounds have surface tones that can be clearly traced to the surface tones of the individual morphemes. For example, /tuᴸ/ 'mustard' + /tɕɑj/ 'oil' yields [tu˨˩tɕaj˧ ~ tu˦˩ tɕaj˧] 'mustard oil' (Blench, n.d.). We leave the analysis of the tone rules of compounds for a future work.

In terms of previous works, Chakravarty (1963) recognizes the existence of tone, but says that the details remain to be worked out. Sastry (1984a, 1984b) also

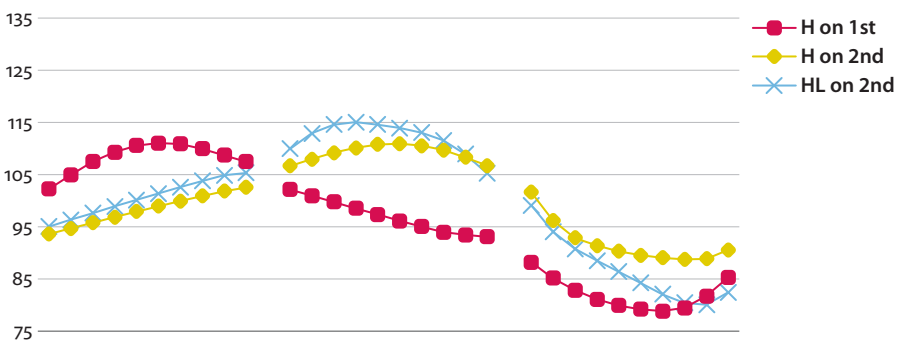

Figure 7. F0 tracing of contrasting tones in Hz on trisyllables (three utterances by JM)

identifies four tones, which he calls 'Level'/unmarked (our Mid/unmarked), Rising (our High), Falling (our Low), and Rising-Falling (our Falling) tones. His analysis is close to, but not exactly the same as ours, as can be seen in Table 15 (note that 'bite (as snake)' and 'to ax' are the same lexeme).

Table 15. Tone minimal sets in Sastry (1984b: 13, 14) and our transcription

Sastry	Our form	Gloss
píya	/pjɑ^{HL}/	'bird'
pìyà	/pi-jɑ^{L}/	'to bite' (as snake)
piyà	/pi-jɑ^{L}/	'to axe'
pîyà	/pi^{H}-jɑ^{L}/	'to pack'
nyáŋ	/njɑŋ/	'to push'
nyàŋ	/njɑŋ^{L}/	'daughter in law'
nyaŋ	/njɑŋ^{H}/	'room'
kɨnyâŋ	/kɨnjɑŋ/	'nine'
kânínpá	/kɑniŋ^{H}pɑ/	'opium cloth'
pà	/pɑ^{L}/	'axe'
pâ	/pɑ^{H}/	'to cross'
pa	/pɑ^{H}/	'yeast'
--	/pɑ/	'draw water'

Progressive aspect is marked with verbal affix /-ʔ/, which causes creakiness on a preceding vowel, and also causes mid-toned words to have a rising pitch: [pʰlɑ̃˦] *phlã* 'cook', [pʰlɑ̰ʔ ˥] *phlãq* 'cooking'.

Jiang et al. (2013) identifies High, High-Falling, Mid-Rising, and Mid-Falling tones. Blench (n.d.) lists High, Low, Rising, and Falling tones. Neither of these two publications proposes a neutral, toneless specification. Blench (n.d.) provides one minimal pair: /lá/ 'beak', /là/ 'tooth'. Unfortunately, these are the same morpheme: /lã^{L}/ 'tooth, beak'. The forms in Jiang et al. (2013; e.g., 33–34) do not exactly match what is found in the dialects of Arunachal Pradesh; e.g., their form /hɑ^{55} lo^{55}/ corresponds, but is not identical to, to our mid-toned /hɑ lo/ 'moon'. Because tonal oppositions are not presented in minimal pairs in Jiang et al. (2013), it is not yet clear if the difference recorded between Chinese and Indian dialects is due to dialectal variation, or some other reason.

None of the published sources make reference to the use of tone to indicate a grammatical morpheme. Definite marking appears to be marked by a combination of vowel length and tone (see Examples 3, 4). However, grammatical analysis is still ongoing, and the role of tone within morphosyntax is not yet settled.

6. Orthography

The authors, in collaboration with other speakers, began to work on standardizing Tawrã orthography in 2009, culminating in a presentation of the orthography to the CALSOM in 2016. In 2020, a primer was produced (Manyu et al.), and the government of Arunachal Pradesh approved the teaching of Tawrã in schools. Although there is widespread fluency in Hindi and Assamese, most speakers are not comfortable writing in Devanagari scripts. The Mishmi speakers of Tawrã and of Kman (Miju) expressed a desire to use Roman letters with values similar to those of English, so that the orthography could also provide a bridge to English language learning. Because these two populations are in contact with each other, there was also a desire expressed to have the transcription systems for the two languages to be as similar to each other as possible.

There was some controversy surrounding the transcription of the vowel /ɨ/; it was decided that *ü* is easy to type on a phone, and doesn't introduce the confusion that using a different Roman letter might cause. Various proposals for the representation of vowel nasalization were considered. The final consensus was to use tilde *ĩ, ã, ũ*. No other diacritics were desired.

One of the purposes of this study is to introduce to the wider Tawrã speaking community both the orthography and the reasoning for the underlying phonological analysis. Furthermore, because this study provides native speakers with the data and argumentation given above, it serves as a check on the validity of the phonological analysis and the orthographic proposal. Additional exemplification of the transcription system demonstrated here may be found in Manyu, Chai, Chaitom, Tayang, Tega, Thalai and Ama (2020).

It should be noted that Tawrã orthography is not exactly phonological, although most letters (and letter combinations like *th* for /tʰ/) correspond to distinctive sounds in the language. One counter example is that the letter *ü* is used to transcribe both the vowel /ɨ/ (/maᴸ sɨ/ *masü* 'have a cold'), along with the similar sounding consonant [ɯ] (/bwiᴸ/ [bɯi˩] *büi* 'dance').

In line with Bird's (1999: 83) observation that "tone marking degrades reading fluency and does not help to resolve tonally ambiguous words," we have found that Tawrã speakers find tone marking to be difficult to master, and a distraction to reading fluency. Thus, tones are not marked in the orthography, although they are available in phonologically transcribed reference materials.

7. Conclusions and further directions

The summary of Tawrã phonology presented here attempts to systematically demonstrate the basic phonological categories present in the variety of Tawrã spoken in and around Tezu, Lohit District, Arunachal Pradesh. By combining native-speaker intuitions with linguistic analysis, we have cleared up some errors and gaps in the literature. Furthermore, the present study demonstrates the systematic basis for the newly codified orthography of Tawrã.

8. Bidialectal wordlist

In collaboration with Professor Di Jiang (江荻), we present a wordlist of about 2500 Tawrã words, collected both in Tibet and in Arunachal Pradesh, with glosses in English and Chinese. The list is available at https://www.webonary.org/tawra.

Acknowledgements

The authors wish to acknowledge the generous assistance of native speakers Baham Tindya, Shiva Tayang, Jabralum Chaitom, and the Cultural and Literary Society of Mishmi. The first author also received assistance from Kishor Dutta, Uta Reinöhl, Mathew Ebenezer, Mathew Thomas and Wycliffe India. Wen-Chi Yeh and Chiawei Wang assisted in manuscript preparation.

Funding

Funding for this research was generously provided by Ministry of Science and Technology Grant (Evans) #100-2628-H-001-008-MY4.

References

Aiyadurai, Ambika. 2011. The Meyor: A least studied frontier tribe of Arunachal Pradesh, Northeast India. *The Eastern Anthropologist* 64(4): 459–469.

Bird, Steven. 1999. When marking tone reduces fluency: An orthography experiment in Cameroon. *Language and Speech* 42: 83–115.

Bird, Steven. 2001. Orthography and identity in Cameroon. *Written Language & Literacy* 4(2): 131–162. https://doi.org/10.1075/wll.4.2.02bir

Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. James A. Matisoff, contributing editor. Cambridge: Cambridge University Press.

Blench, Roger. n.d. *Proposal for a writing system for Tawra, a language of Arunachal Pradesh*. https://www.academia.edu/12868198/Proposal_for_a_writing_system_for_Tawra_a_language_of_Arunachal_Pradesh (accessed 21 June 2020).

Blench, Roger. 2015. *The classification of Meyor (Zakhring)*. http://www.rogerblench.info/RBOP.htm

Blench, Roger. 2017. *A dictionary of Tawrã, a language of Arunachal Pradesh*. http://rogerblench.info/Language/NEI/Mishmi/Tawra/Tawra%20Dictionary.pdf

Chakravarty, L. N. 1963. *A dictionary of the Taraon language*. Shillong: Philology Section, Research Department, Northeast Frontier Agency.

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

Driem, van George. 2007. Endangered languages of South Asia. In Matthias Brenzinger (ed.), *Language diversity endangered*, 303–341. Berlin & New York: Mouton de Gruyter.

International Phonetic Association and International Phonetic Association Staff. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jiang, Di, Daqin Li & Hongkai Sun. 2013. Darangyu yanjiu [A study of Taraon]. Beijing: Nationalities Publishing House.

Konow, Sten. 1902. Note on the languages spoken between the Assam Valley and Tibet. *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 1902: 127–137.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2015. *Ethnologue: Languages of the World*, Eighteenth edition. Dallas, Texas: SIL International. Online: http://www.ethnologue.com.

Lindblom, Björn & Ian Maddieson. 1988. Phonetic universals in consonant systems. In Larry M. Hyman, Victoria Fromkin & Charles N. Li (eds.), *Language, speech, and mind: Studies in honour of Victoria A. Fromkin*, 6278. Taylor & Francis.

Luce, Gordon H. 1944. *Hiroi-Tarao lexicon by C. Ginson prepared for Robert Shafer, dated 12/11/1944*. In Luce Collection, MS 6574-7, 029a–029b. Canberra: National Library of Australia.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. Southern California Occasional Papers in Linguistics 1: 73–95.

Matisoff, J. A. 1996. The cognate noun/verb construction in Lahu. *Linguistics of the Tibeto-Burman Area* 19(1): 97–101.

Morey, Stephen D. 2008. Working with tones in Northeast India-the tonal system of Numhpuk Singpho. In Stephen D. Morey & Mark W. Post (eds.), *North East Indian linguistics*, 26–44. New Delhi: Cambridge University Press. https://doi.org/10.1017/UPO9788175968431.004

Manyu, Johakso, Naina Chai, Jabralum Chaitom, Sowingso Tayang, Buglelum Tega, Khope Thalai & Baritlum Ama. 2020. *Tawrã (Digaru) Mishmi orthography (primer)*. Tinsukia: Signage & Graphics.

Mielke, Jeff. 2009. Segment inventories. *Language and Linguistics Compass* 3(2): 700–718. https://doi.org/10.1111/j.1749-818X.2008.00117.x

Modi, Yankee. 2013. The nearest relatives of the Tani group. Paper presented at the 19th Himalayan Languages Symposium. Canberra, Australian National University, 6–8 September, 2013.

Mya, Tu, U. Ko Ko, U. Aung Than Batu, U. Kywe Thein, C. J. R. Francis & U. Than Tun Aung Hlaing. 1966. *The Tarons in Burma*. Rangoon: Burma Medical Research Institute.

Needham, Jack F. 1978 [1886]. *A few Digaro (Taraon), Jiju (Mju) and Tibetan words collected during a trip from Sadiya to Rima and back in December 1885 and January 1886*. Shillong: Nishi, Toshio.

Nyunt Nunt Win. 2004. Family and kinship systems of Taron. Karoung Village. Kachin State. *The Jour. Myan. Acad. Arts & Sc.* Vol.II (Part Two) No.4 (Arts): 1–9.

Nyunt Nunt Win. 2015. The Taron: one of the hidden groups of hill ethnic groups in Myanmar. *Mandalay University Research Journal* 6: 1–10.

Post, Mark W. 2015. Tones in Northeast Indian languages, with a focus on Tani: A fieldworker's guide. In Mark W. Post, Stephen Morey & Scott DeLancey (eds.), *Language and culture in Northeast India and beyond: In honor of Robbins Burling*, 182–210. Canberra: Asia-Pacific Linguistics.

Post, Mark W. and Yankee Modi. 2011. Language contact and the genetic position of Milang (Eastern Himalaya). *Anthropological Linguistics* 53(3): 215–258.

Pulu, Jatan. 1978. *Idu phrase-book*. Shillong: Director of Information and Public Relations, Government of Arunachal Pradesh.

Pulu, Jatan. 1991. *A phrase book on Taraon language*. Shillong: Directorate of Research, Government of Arunachal Pradesh.

Rabinowitz, Alan. 2001. *Beyond the last village: A journey of discovery in Asia's forbidden wilderness*. Washington, DC: Island Press.

Rice, Keren. 2011. Documentary linguistics and community relations. *Language Documentation & Conservation* 5: 187–207.

Sastry, G. Devi Prasada. 1984a. *Mishmi phonetic reader. Phonetic Reader Series 25*. Mysore: Central Institute of Indian Languages.

Sastry, G. Devi Prasada. 1984b. *Mishmi grammar*. Mysore: Central Institute of Indian Languages.

Shafer, R. 1974. *Introduction to Sino-Tibetan*. Wiesbaden, Germany: Otto Harrassowitz.

Sun, Hongkai, Lu, Shaozun, Zhang Jichuan & Ouyang Jueya. 1980. *Menba, Luoba, Dengren de yuyan (The languages of the Menba, Luoba and Deng peoples)*. Beijing: Chinese Academy of Social Science Press.

Sun, Hongkai. 1991. Zangmianyu Yuyin he Cihui (Tibeto-Burman phonology and lexicon). Beijing: Chinese Academy of Social Science Press.

Sun, Tian-Shin Jackson. 1993. *A historical-comparative study of the Tani (Mirish) branch of Tibeto-Burman*. Berkeley: University of California, Berkeley PhD Dissertation.

Address for correspondence

Jonathan P. Evans
Institute of Linguistics, Academia Sinica
128 Section 2, Academia Rd.fs
Nangang, Taipei 11529
Taiwan
jonathan@sinica.edu.tw

Co-author information

Johakso Manyu
johaksomanyu24@gmail.com