himalayan linguistics

A free refereed web journal and archive devoted to the study of the languages of the Himalayas

Himalayan Linguistics

Language and dialect relations in Bumthang

Mark Donohue

The Living Tongues Institute for Endangered Languages

The Bhutan Oral Literature Program

ABSTRACT

This report presents basic wordlists from seven closely related East Bodish languages from Bumthang, northern Trongsa and far eastern Wangdue Phodrang districts in Bhutan. These wordlists are analysed, with lexico-statistical comparison to other languages of the region (East Bodish, Central Tibetan, and Indic), and preliminary notes on phonological processes and sound correspondences and change within the Bumthang varieties.

KEYWORDS

Tibeto-Burman, Bhutan, East Bodish, Bumthang, dialects, lexicostatistics, sound change

This is a contribution from *Himalayan Linguistics, Vol. 19(3): 1–45.*
ISSN 1544-7502

Himalayan Linguistics, Vol. 19(3). © Himalayan Linguistics 2020
ISSN 1544-7502

Language and dialect relations in Bumthang

Mark Donohue
The Living Tongues Institute for Endangered Languages
The Bhutan Oral Literature Program

1 Introduction

The core of this article are the wordlists taken from five locations in Bumthang (བུམ་ཐང་), representing six different language varieties. In addition to the core varieties sampled within Bumthang, two additional varieties, the Trongsa (ཀྲོང་གསར་) variety of Bumthang and a wordlist of Henkha, are included. In the analysis eleven additional language varieties from the region surrounding Bumthang were included for comparison.

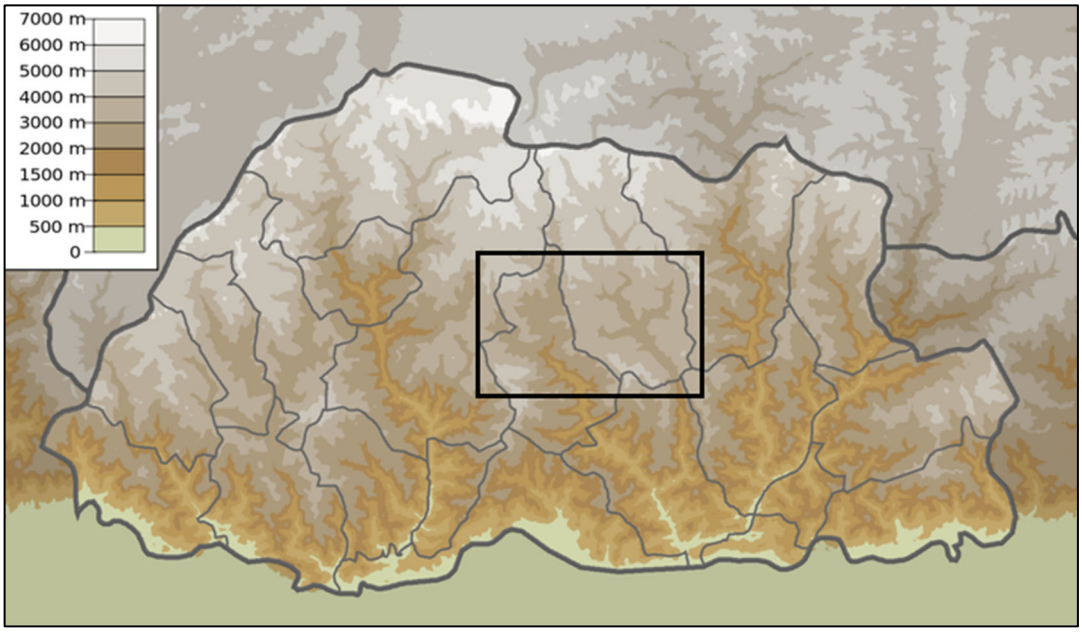
The sample from Bumthang includes the main varieties from all of the inhabited valleys in Bumthang; they are listed in Table 1, which shows the name, the locations of the villages where wordlists were elicited, altitudes, and the iso code for that language. [1] All are from the central part of Bhutan, inside the box in the middle of Map 1. As can be seen in Maps 1 and 2, there are no samples from the north of Bumthang. Following the Chinese take-over of Tibet this high-altitude part of the country has been uninhabited, following the abandonment of a trading village north of Tang that previously served as a way point for trade caravans travelling to Tibet. All of the Bumthang-speaking villages are now found below 3500m (Ura is the highest continuously inhabited village in Bhutan). Trongsa, to the west, is not part of the Bumthang highlands, lying just above 2000m, spoken around Trongsa Dzong, the fortress that controlled and taxed trade from the lowlands further down the Mangde-chhu rising up to Bumthang. The central valley of Bumthang is the widest, lowest, and longest, and so has the highest population of anywhere in the highlands. Wordlists were gathered from near Jakar town, and also from Dhur in the north. It is likely that further variation is waiting to be discovered, and the villages on the borders of the Tang and Chumey valleys, which open onto the central valley, are likely to be very interesting, in light of the lexical separation of Chamkhar from the other varieties. Chamkhar has prestige as the home to Jakar Dzong (བྱ་དཀར་རྫོང་), a large temple-fortress dominating the valley. Chumey includes a royal palace at its north-western end, and the northern end of the Tang valley includes Orgyen Chöling (ཨོ་རྒྱེན་ཆོས་ལིང་), a fortified manor that controlled trade to Tibet, and held lowlands in Lhüntse and Monggar. Dhur is found at the northern periphery of the Bumthang valley, lacking the prestige of these other regions, and being located on the slope north of a narrow part of the upper valley. Ura is the smallest and highest of the four valleys, well above rice-growing altitudes, but is at the head of a long gorge leading down to Zhemgang (གཞས་སྒང་), where

[1] The place names cited here are listed as they are commonly spelled in Bhutan, including in the communities concerned in Bumthang. These spellings are sometimes at variance from standard transliterations used by Tibetologists.

Khengkha is spoken in the lowlands and where yaks from the southern parts of Bumthang are still sometimes pastured in the winter.

Inside Bumthang				
Chamkhar	(ཆམ་ཁར་)	27.55, 90.75	2600m	kjz
Dhur	(དུར་)	27.615, 90.665	2900m	kjz
Chumey	(ཆུས་མའི་)	27.493, 90.71	2810m	kjz
Tang	(ཐང་)	27.615, 90.885	2750m	kjz
Ura	(ཨུ་ར་)	27.48, 90.91	3160m	kjz
Peripheral to Bumthang				
Trongsa	(ཀྲོང་གསར་)	27.52, 90.5	2050m	kjz
Henkha[2]		27.45, 90.37	2380m	neh

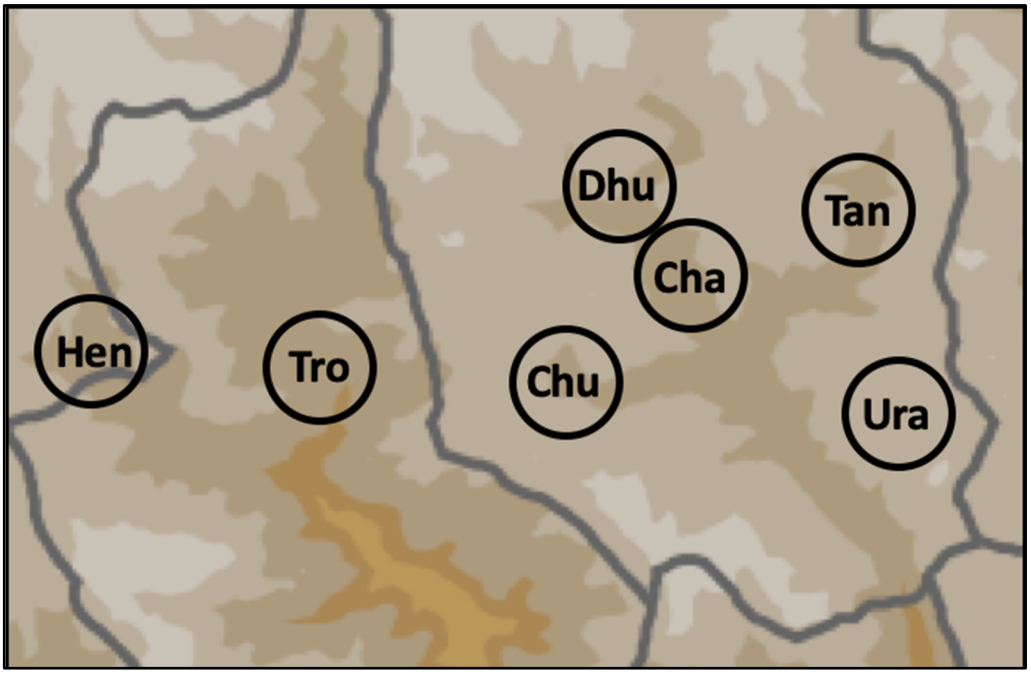
Table 1. Different language varieties included in this study



Map 1. Location in Bhutan

Previous work on Bumthang was initiated by van Driem (1995/2015), and includes work by Donohue and Donohue (2016, 2019), Peck (2017), and Wyatt (2017). Relevant work on the classification of Bumthang and languages related to Bumthang includes Shafer (1954), Michailovksy and Mazaudon (1994), van Driem (2007), and Hyslop (2013). It is generally agreed that Bumthang is most closely related to Khengkha and Kurtöp, to its south and north-east, respectively, and this study supports that conclusion.

[2] Also known as Nyenkha, 'Nyenkha, (Upper) Mangdep, Mandebi-kha, Mangdebikha, Mangde, Henke, Hen Kha and Phobjip (in Dzongkha, འབྲོག་ཕ་). The multiplicity of names almost certainly reflects a more complex linguistic situation than has so far been described (see Bosch 2016). The term 'Henkha' is used here, as that is what the speakers I worked with used to refer to their language.



Map 2. The core languages in this study

2 Elicitation

The wordlists in the appendix were collected in Bhutan in October – November 2014, from informants in the different locations in Bhutan, as described in Table 1. Data from the Ura dialect was also collected in Thimphu, Bhutan, and in Canberra, Australia. Data from the Tang dialect was also collected in Bouddha, Nepal.

2.1 Lexicon

Some words proved to be difficult to elicit, and so are only partially attested in the wordlists in the appendix, and were not included in the analysis later.

Words that were difficult to successfully elicit can be judged by their relatively poor representation in the wordlists; some words are only listed for Ura dialect, the result of their later elicitation in Canberra.

Many verbs proved hard to elicit accurately, complicated by verbal inflection and the use of different auxiliaries, or of the use of a generic word rather than a specific one, or a very specific word rather than a generic one. Other words that proved to be difficult include:

Body parts:

Forehead, cheek	(often 'face, head, hair, eyebrow, skin')
Chin	(often 'jaw, mouth, beard, head')
Palm (of hand)	(often 'chest')
Fat (on body)	
Urine, Faeces	(often informants were not willing to discuss this, or else indicated a nearby toilet for my use)

Human terms:

son	(same as 'boy')

	daughter	(same as 'girl')
Animals:		
	buffalo, ox, calf	(confused with 'cow, ox, bull')
	horn	
	monkey	(species confusion, and not present in the highlands)
	insect	(difficulties with a generic term)
	flea	(difficult to elicit as distinct from 'louse')
	firefly	('fire insect')
	leech	(not present in the highlands)
Plant terms:		
	bark (tree)	(same as 'skin')
	eggplant	(confusion about identity, rare in the highlands)
	flower	(not widely present in the highlands)
Natural world:		
	dust	('ash, dirt')
	mud	('dirt, water')
	valley	('land, village, field')
	air	(not separate from 'wind')
	lightning	(phrasal expression with 'sky')

These words are listed where known, mostly from Ura dialect, but they were not included in the lexicostatistics presented in section 2.5.[3]

The terms for grains are illustrative of the complications caused by semantic shift when evaluating lexical similarities. Presented in more detail in the appendix, Table 2 traces semantic shifts between various grains in the different language varieties. The word for 'millet' in Chumey is *nas*. This apparently corresponds to Ura *nat*, except that the regular correspondence of a final *-s* in Chumey is a final *-s* in Ura as well. No other language in this sample uses *nas* to refer to millet, but *nas* is found in Ura to mean 'black barley', and in Trongsa to refer to 'wheat', for which Tang has *nat*. Further, generic barley is *nat* in Chamkhar, *nas* in Trongsa, and *neʂ* in Henkha. Chamkhar *nam* 'millet' might be connected to this lexical set.[4] Henkha appears to have a loan from Dzongkha in *ʈɛ*, shifting the sense from 'rice' to 'millet'. Tibetan and Dzongkha *ʈɛ* 'rice',

[3] The list of words used for analytical purposes is: 'head', 'hair', 'eye', 'ear', 'nose', 'mouth', 'tooth', 'tongue', 'neck', 'arm', 'elbow', 'finger', 'breast', 'belly', 'leg', 'knee', 'skin', 'blood', 'bone', 'flesh'; 'person', 'mother', 'father', 'husband', 'wife', 'elder sister (eZ)', 'elder brother (eB)', 'younger sister (yZ)', 'younger brother (yB)', 'language', 'name'; 'I/me (1SG)', 'you (2SG)', 'he/him (3SG.M)', 'she/her (3SG.F)', 'we/us (1PL)', 'you (2PL)', 'they/them (3PL)'; 'bird', 'egg', 'dog', 'tail', 'horse', 'yak', 'cow', 'goat', 'chicken', 'snake', 'fish', 'fly', 'louse', 'mosquito', 'spider', 'ant'; 'tree', 'leaf', 'roots', 'thorn', 'seed', 'banana', 'salt', 'potato', 'paddy'; 'cooked rice', 'millet', 'wheat', 'milk', 'chilli', 'garlic', 'onion', 'pumpkin', 'sugarcane', 'butter'; 'ground', 'sand', 'mud', 'water', 'snow', 'ice', 'cloud', 'rain', 'sky', 'sun', 'moon', 'day', 'night', 'today', 'tomorrow', 'yesterday', 'star', 'fire', 'smoke', 'ash', 'mountain', 'mountain pass'; 'house', 'roof', 'village', 'thread', 'ring', 'path', 'knife', 'axe', 'bow', 'arrow', 'target'; 'above', 'below', 'front', 'behind', 'left', 'right'; 'black', 'white', 'red', 'green', 'yellow', 'big', 'small', 'old', 'new', 'hot', 'cold', 'good', 'bad', 'wet', 'dry', 'long', 'short'; 'one', 'two', 'three', 'four', 'five', 'six', 'seven', 'eight', 'nine', 'ten', 'twenty', 'hundred'; 'see', 'hear', 'know', 'speak', 'sleep', 'wake up', 'wake up.TR', 'go', 'climb', 'descend', 'come', 'arrive', 'walk', 'run', 'sit', 'stand', 'wash', 'dig', 'eat', 'drink', 'die', 'kill', 'hit', 'give', 'laugh', 'cry', 'burn (intr.)', 'cook', 'no(t)'; 'who', 'what', 'where', 'when', 'how many', 'why'. Semi-colons mark the division into different semantic fields; divisions other than those shown here are also possible.

[4] We can also note Chöcangacakha *na*, Dzala *nè* 'wheat'. The forms are not related to Tibetan *muntʃa?* 'millet', but have a probable relative in Tibetan *nɛ* 'barley' (written ནས་ nas).

written འབྲས་ 'bras, is related to the words for 'buckwheat' in Table 2, but also to 'rice in field (paddy)' in Tang, Ura and Trongsa, and also 'unhusked rice' in Ura.

The Dhur word for 'paddy', *ʈʊŋ*, is also their word for rice grain (husked or unhusked), but in other varieties this form refers to 'husked rice' exclusively. The lack of rice growing in Dhur, which is at too high an altitude for this to be practical, means that all rice is imported in the form of purchased (husked) rice, and the term has spread to mark all reference to uncooked rice. Chumey, Tang and Trongsa allow *iba* to refer to paddy rice, but in Chamkhar the related term *iβa* refers to unhusked rice (which is freshly harvested from the field). Chumey has a specialised term *ʧaraj* for reference to unhusked rice, but the related term *ʤaraj* means 'barley' in Ura, for which Tang uses *ʧaræa*. The wide range of reference for each of these terms indicates that the reference of the term is to a cultural practice ('cultivation of grain for consumption'), and that the reference of each term to a specific type or stage of grain is subject to change under contact conditions.

The clear relationship of the various buckwheat terms to the Tibetan term, which has wider currency as a reference to rice,[5] strongly indicates that the term was an early loan into Bumthang, but taken to mean 'staple grain food', which at that time was buckwheat.[6] The wide range of referents for the *nas* term also indicates that the term was likely borrowed into at least some of the varieties considered here, and spread to refer to different grains as dictated by local agricultural practices. In addition to these loans which have transferred their reference, we also see endemic terms such as *iba* and *ʧaraj* surviving into the modern languages.[7] Due to the at least partial multilingualism in local varieties that most speakers in the area have, as well as almost universal ability (to some level) with Dzongkha and Tibetan, it is likely that the reference of many of these terms is not fixed for some speakers.

	millet	paddy	buckwheat	unhusked rice	husked rice	barley	black barley	wheat
Chamkhar	nam		branma	iβa	ʈʊŋ	nat		kar
Dhur		ʈʊŋ	braŋma	ʈʊŋ	ʈʊŋ			kar
Chumey	nas	iba	bratm	ʧaraj	ʈʊŋ			gɔ
Tang		iba, mrat	branma		ʈʊŋ	ʧaræa		nat
Ura	kɔŋbɔ, (nat)	mras	brasma	mras	ʈʊŋ	ʤaraj	nas	kar, gɔ
Trongsa	brakma	mras, iba	branma	ʈʊŋ	ʈʊŋ	nas		nas
Henkha	ʈɛ		brɛm		ʧʰʊŋ	neʂ		kaɾ

Table 2. Selected grain terms

[5] Proto-Tibeto-Burman *b-ras 'bear fruit, rice', Chinese 糲 'brown rice, unpolished rice' (Mandarin lì, Old Chinese *m·raːds (Zhengzhang 2000, 2003), *[r]ˢat-s (Baxter and Sagart)), rGyalrong *mbras* (Suzuki et al. 2016). See also van Driem (2012).

[6] Further evidence that borrowing is implicated in the structure of the lexicon examined here is the appearance of doublets in many languages.

[7] Other relevant Proto-Tibeto-Burman reconstructions (drawn from Matisoff 2003) include *zəy 'barley', *ra 'buckwheat', *ma-y/*mey 'rice, paddy', *dzya 'rice', and *ka 'grain of rice'. While *ka might be related to *kar* 'wheat' in the languages considered in Table 1, the other reconstructions are not likely antecedents of any of the terms recorded.

This brief exploration of grain terms shows the level of complexity that long-term language contact, both internal and external, can have on the lexicon.

2.2 Transcription

The transcription used is in most cases broadly phonetic, rather than necessarily representing underlying phonological contrasts, because exhaustive phonological analyses of most of the language varieties discussed have not yet been completed. A number of fricatives are noted which are not contrastive with stops (e.g., β for b, ɣ for g, in intervocalic position adjacent to non-high back vowels). The sound systems of the varieties are transcribed with the (predominantly IPA) symbols shown in Figure 1. Note that the high front rounded vowel was transcribed as ü, not y, to avoid any possible confusion with the palatal glide. The umlauted ö is occasionally used for a lower front rounded vowel, when the exact quality was variable. All of the apical stops are dental, though this is not indicated in the transcriptions, as it is not contrastive. Similarly, the retroflex plosives are released with weak non-grooved fricative: ʈʰ̪ɻ̝̊, ʈɻ̝̊, and ɖɻ̝̊.

pʰ	tʰ	tsʰ	tʃʰ	ʈʰ	kʰ	qʰ		i	ü		u		
p	t	ts	tʃ	ʈ	k	q	ʔ	ɪ			ʊ		
b	d		dʒ	ɖ	g			e	ö	ə	o		
m	n		ɲ	ɳ	ŋ			ɛ	œ		ɔ	ɔ̃	
ɸ		s	ʃ	ʂ	x		h			a			
β		z	ʒ	ʐ	ɣ								
	r l ɹ			ɻ									
			j				w						

Figure 1. Symbols used in transcription

Likely phonological equivalents for different symbols in the languages in the sample are listed in Table 3. For example, it is very unlikely that [ɸ] is contrastive with [pʰ] in the languages concerned, since alternations between [ɸ] and [pʰ] have been observed elsewhere in the data collected on Bumthang. Similarly there is a great range of allophony of the single contrastive rhotic segment, influenced by the position in syllable, the quality of the adjacent vowel(s), and the tone; regardless of this, all of the different realisations can be considered to be equivalent phonologically. The high vowels show lax allophones when they occur in closed syllables. The only occurrence of the retroflex nasal [ɳ] in the data is preceding a retroflex plosive, clearly representing assimilation. The mid vowels probably do not show true variation; there are two contrastive mid vowels in at least some of the languages considered here, but the transcription is not consistent, and lexemes that vary between varieties solely on the basis of the height of the mid vowel should not necessarily be considered to be distinct. Similarly, given the (at least partial) correspondence of voicing with pitch height or inter-sonorant position, the difference between (for example) *s* and *z* or *p* and *b* in the transcriptions should not be taken as being categorial.

p^h	ɸ		
b	β		
p, t, ʧ, ʈ, k, ɸ, s, ʃ, ʂ, x	b, d, ʤ, ɖ, g, β,z, ʒ, ʐ, ɣ		
ɲ	ŋj		
n	ɳ		
k	x	q	ʔ
r	ɻ	ɹ	ɾ
k^h	q^h	x	
i	ɪ		
u	ʊ		
a	ə		
(e	ɛ)		
(o	ɔ)		

Table 3. Likely equivalences

Most notably, tone is not directly transcribed. This is because the tonal system of at least one variety of Bumthang (Ura) is very complex, involving more than eight contrasts which are susceptible to change due to intonation and the presence of tonally-signalled morphology, and so not amenable to annotation in a survey context.[8] As far as it makes sense to speak of stress in a language with contrastive tone, stress is initial, and trochaic.

2.3 Analysis

While the core Bumthang group of dialects spoken in the Bumthang highlands have a very high level of lexical correspondences, the realisation of these apparent cognates can vary considerably. Sound correspondences are rarely categorial between the varieties described here, and will be dealt with in the following sections.

2.3.1 Erratic correspondences

These most likely represent the combination of local innovations combined with extensive inter-dialectal borrowing induced by contact and marriage with groups outside the range of the dialect(s) concerned. In Table 4 we can see the treatment of an initial *kr- cluster in three words well attested in the different wordlists; while it is clear that Ura is most conservative in preserving the *kr- cluster, and Chumey is most prone to the reduction of *kr and *gr, and is probably the centre of this innovation there is no clear pattern in the distribution of its reflexes. In all of the varieties in Table 4 except for Chumey and Henkha the onset clusters *kr* and *gr* are attested, but with different lexical distributions. The form *ʈu* in Henkha is probably a loan from Dzongkha.

[8] For instance, five levels are attested in the Ura Bumthang words *nu*^{55} 'soot', *lɔŋ*^{44} 'wind', *mi*^{33} 'person', *ŋɔj*^{22} 'silver', *jam*^{11} 'path'. The same dialect also has contour tones, including those attested in *na*^{52} 'head of queue', *ɲam*^{343} 'fat', and *wam*^{334} 'bear'. While segmentally distinct, all of the consonants are sonorants, and there is nothing in the different segments involved that would be responsible for the different pitch contours.

	'wake up (tr.)'	'hair'	'village'	'six'	'elbow'
Chamkhar	r	r	kr	gr	gr
Dhur	r	r	kr	gr	k
Chumey	r	r	r	r	r
Tang	r	r	kr	gr	gr
Ura	kr	kr	kr	gr	gr
Trongsa	r	kr	kr	gr	(tiwa)
Henkha	r	hr	(saʃː)	(ʈu)	(tsɪktɔ)
	kruŋ-	kra	krɔŋ	grok	grumaŋti

Table 4. Reduction of |kr| and |gr|

In Table 5 we can see different patterns of lenition with *pʰ. While this is allophonic is many varieties, the segment is normally realised as a fricative in Chamkhar. Intervocalically a voiced fricative is attested in some varieties (Chamkhar, Tang, Dhur), and a voiceless fricative is more common in other varieties that do not show initial lenition (Chumey, Ura). The Trongsa [w̃] represents the collapse of the entire second syllable (pʰaŋ); the full form of the word in Trongsa is [naw̃]. The final velar nasal is preserved in nasalisation, and the bilabial stop passed through a lenition chain of pʰ > ɸ > β > w. We assume that the vowel of the second syllable was first nasalised, then assimilated to the preceding *w*.

	'spider'	'forehead'	'nose'	'tail'
Chamkhar	ɸ	ɸ	β	ɸ
Dhur	pʰ		ɸ	
Chumey	pʰ	pʰ	pʰ	ɸ
Tang	pʰ		β	β
Ura	pʰ	pʰ	pʰ ~ ɸ	pʰ
Trongsa	pʰ		w̃	
Henkha			p	
	pʰrumzaŋ	pʰaj(laŋ)	napʰaŋ	mipʰaŋ

Table 5. Lenition of bilabial segments

Pre-coronal fronting is found erratically in many of the varieties described here, though it is not categorial, nor is it consistently conditioned by the following segment.[9] Table 6 shows correspondence sets in which a back rounded vowel, *ɔ o* or *u*, is fronted preceding a coronal segment, *t n* or *s.* This is most unambiguously illustrated with 'young', showing *o* in most varieties appearing as *ö* in Dhur and Tang preceding an *n*, and 'kill', with *u* corresponding to Henkha *ü* preceding *t*. Both of these words illustrate pre-coronal fronting, but the languages in which is occurs are not the same: there is no fronting in 'kill' as pronounced in Dhur or Tang. We must also note that in Dhur the conditioning environment has been lost, through assimilation of the

[9] Pre-coronal fronting is also a feature of the historical phonology of Central Tibetan, and given the social use of Tibetan in the Himalayas it is not unlikely that the sporadic and erratic appearance of pre-coronal fronting in languages such as those described here is in part due to this influence.

*n to the place of the following (historical) nominaliser, and so we must assume that the resulting front rounded vowel has been reinterpreted as an underlying contrastive segment. In 'knee' we see fronting in Dhur, Tang, and Chumey, with raising of the vowel in Tang, presumably through assimilation as a form of limited vowel harmony with the following high vowel. Similar variation in vowel height is found with 'black', with fronting appearing now in Chamkhar and Trongsa.

We can also see the palatalisation of the initial nasal in a pattern that is not dependent on the fronting of the vowel, though the two do coincide in Tang.[10] In 'wash' only Tang shows fronting, and in an interesting twist a non-etymological -p has been added to the end of the verb root. In left' we see a situation similar to that in 'young', but with even greater opacity in the conditioning environments, and the spreading of front rounding to the initial glide, regardless of the fronting of the intermediate vowel. (This correspondence set also shows great confusion in the form of the vowel, a point that will be addressed further later.)

	'knee'	'day'	'young'	'black'	'wash'	'left'	'kill'
Chamkhar	punmuŋ	ŋjɛn	ʃunbɔ	ŋündila	kʰrɔ-	œβalɔk	sut-
Dhur	pœnmuŋ	nɛnʈa	ʃjömba	ɲɔndi	kʰra-	jœmba	sut-
Chumey	pœnmɔ	nɛnʈa, nœnʈa		ɲünde	kʰrɔ-		sut-
Tang	pünmuŋ	nɛnʈa	ʃönbɔ	ɲündi	kʰröp-		sut-
Ura	*puspuŋ*	nɛnʈa	ʒonbu	ɲɔndi	kʰrɔs-	ɛmbu	sut-
Trongsa	putmo	nɛm	ʒonmula	ŋündi	kɔ-	üɛnmɔ	sut-
Henkha	*ŋatɔ*			*nakʰɛ*	ʈʰo-	œŋkatɔ	sü-
	pɔn-	nɛn-ʈa	ʒon-	ŋɔndi	kʰrɔs-	j[ɛ/ɔ]n-	

Table 6. Pre-coronal fronting

It is worth working through some of the sound changes that have applied to the forms in Table 6. In Figure 2 we can see that the form for 'knee' in Tang, *pünmuŋ*, is the result of pre-coronal fronting applying to the first vowel, and regressive vowel height assimilation. The Dhur form lacks vowel height assimilation, but has undergone the same pre-coronal fronting. In Chamkhar there is no pre-coronal fronting, only vowel height assimilation. Chumey displays the results of both processes, plus the deletion of the final nasal coda, but the vowel height assimilation operates progressively, not regressively.

[10] Similarly, Ura *notpa* 'throat' corresponds to Chamkhar *nœtpa*, showing pre-coronal fronting of both the vowel and the nasal.

	'knee'					
Original	pɔnmuŋ		pɔnmuŋ		pɔnmuŋ	
PCF	pœnmuŋ	Dhur			pœnmuŋ	
VHA	pünmuŋ	Tang	punmuŋ	Chamkhar	pœnmɔŋ	
NCD					pœnmɔ	Chumey

PCF: Pre-coronal fronting
VHA: Vowel height assimilation
NCD: Nasal coda deletion

Figure 2. Sound changes in selected etyma from Table 6: 'knee'

In Figure 3 we see the palatalisation extending to the onset. The Tang and Chamkhar forms exhibits onset palatalisation as well as pre-coronal fronting on the vowel and vowel height assimilation. In Dhur and Ura we find only onset palatalisation. The Dhur form for 'young' has undergone pre-coronal fronting, onset palatalisation, and nasal assimilation (of the nasal to the following plosive, in place), thus removing the environment which conditioned the palatalisation in the first place.

	'black'				'young'	
Original	ŋɔndi		ŋɔndi		ʒon-bV	Ura
PCF	ŋœndi				ʒœn-bV	Tang
VHA	ŋündi	Trongsa				
OP	ɲündi	Tang	ɲɔndi	Dhur, Ura	ʒjœn-bV	
	ɲündila	Chamkhar				
NA					ʃjömba	Dhur

OP: Onset palatalisation
NA: Nasal assimilation

Figure 3. Sound changes in selected etyma from Table 6: 'black' and 'young'

Fronting is also found preceding a palatal segment. The same complications that have been seen with the previous tables are found with pre-palatal fronting as well. Some of the forms for 'large bamboo (sp.)' exhibit pre-palatal fronting, and in two cases the Tautosyllabic high vowel-high glide sequence is simplified though the loss of the glide, thus removing any synchronic conditioning factor for the change, and phonologising the front rounded vowel. With 'rain' the form with a front rounded vowel *and* the conditioning palatal segment is not found, only the forms with tautosyllabic high vowel simplification (*jœ* in Tang). We find the spread of the rounding to the onset in Chamkhar, and a form in Trongsa, *üɛ*, which shows the operation of a constraint against the appearance of two rounded segments in the one syllable.[11,12] The forms transcribed as [kʰwi] are frequently pronounced as [kʰɥi], showing the spread of palatalisation to

[11] The transcription [ü] in these cases (pre-vocalic) represents an alternation between IPA [y] and [ɥ].

[12] This constraint is justified elsewhere in Bumthang; while *wa*, *we*, *wɛ* and *wa* are attested in the Ura dialect, **wɔ*, **wo* and **wu* are not found. Asymmetrically, with the front glide *ju*, *jo*, *jɔ* and *ja* are attested, but also *jɛ* and *je*, while **ji* is not permitted.

the glide, showing that in these languages two high front segments can occur, as long as one of them is rounded and one is not. In Trongsa the operation of the tautosyllabic rounding simplification constraint results in the preservation of the rounding, but the loss of the high front unrounded vowel, [kʰü].

	'dog'	'water'	'bamboo (large)'	'rain'	'that'	'you.SG'
Chamkhar	kʰwi	kʰwe	ruj	üœ	utuj	wɛt
Dhur	kʰwi	hwe	rü	jœ	tʰü	wɛt
Chumey	kʰwi	kʰwe	ɻü	jœa	tü	wɛt
Tang	kʰwi	kʰwe	rüj	jœ	utuj	wɛt
Ura	kʰwi	kʰwe	ruj	joj	utuj	wɛt
Trongsa	kʰü	kʰœ	ruj	üɛ	*bɔt*	üɛt
Henkha	ʧʰi	kʰɛ	ɻü	ü	utü	*ji*
	kʰwi	kʰwe	ruj	joj	utuj	wɛt

Table 7. Pre-palatal fronting

	'bamboo'		'rain'		'you.SG'	
Original	ruj	Ura	joj	Ura	wɛt	
PPF	rüj	Tang	jœj			
THVS	rü	Dhur	jœ	Tang		
OR			üœ	Chamkhar		
TRS			üɛ	Trongsa	üɛt	Trongsa

PPF: pre-palatal fronting
THVS: tautosyllabic high vowel simplification
OR: onset rounding
TRS: tautosyllabic rounding simplification

Figure 4. Sound changes in selected etyma from Table 7: 'bamboo', 'rain' and '2SG'

Finally, there are some instances of the idiosyncratic appearance of front rounded vowels without a clear conditioning environment. The vowel in 'eat' is front and rounded in Tang and Henkha. The vowel alternation in 'small' might reflect and original *tʃunku-, with palatalisation spreading from the *tʃ* onset in most cases, and nasal assimilation to *ŋ* before the velar stop. Under this explanation the Henkha form lacks the palatalisation of the vowel to *i*, and the Trongsa form underwent pre-coronal fronting prior to the operation of nasal assimilation. The forms for 'when' are clearly related, with the exception of Henkha, which if cognate exhibits the lenition of the internal *b > w, followed by the vowel in the first syllable assimilating in height and rounding to the *w*, after being fronted before the coronal *r/*t.

	'eat'	'small'	'when'
Chamkhar	su-	tʃĭŋgula	aɾβa
Dhur	su-	tʃɪŋkula	aɾβa
Chumey	zu-	tʃĭŋgula	aɾβa
Tang	zœ-	tʃĭŋgula	aɾba
Ura	zu-	tʃɪŋkula	aɾba
Trongsa	zu-	tʃŭŋku	aɾba
Henkha	zü-	tʃʊŋku	üwɛ
	zu-	tʃunku- ?	a[r/t]ba

Table 8. Erratic front rounded vowels

*2.3.2 Regular correspondences, and the retention of *l*

The lenition of *l > *j* has been cited (Shafer 1954, Michailovksy Mazaudon 1994, Hyslop 2013) as a defining characteristic of the Core Bumthang group, consisting of Bumthang, Kurtöp and Khengkha. Data such as that in Table 9 supports this assertion, showing the reflexes of Proto-Tibeto-Burman *l in different positions in Bumthang (Ura dialect and others) (drawing on Matisoff 2003) (further examples can be found in the wordlists).

PTB	*g-l(y)ak	'hand/arm'
Bumthang	jak	(various)
PTB	*lam	'road'
Bumthang	jam	(various)
PBodish	*l-ŋa	'five'
Bumthang	jaŋa	(various)
PTB	*pral	'forehead'
Bumthang	pʰaj-laŋ	(various)

Table 9. PTB *l > j

As with many of the other sound changes discussed in this section, the sound change is not categorial. In contrast to the data in Table 9, Table 10 presents evidence of the retention of PTB *l as *l*.

PTB	*s-ley/*s-lay	'tongue'
Bumthang	li	(various)
PTB	*s-la/*lap	'leaf'
Bumthang	lamba	(various)
PTB	*la	'moon'
Bumthang	la	(various)
PTB	*liŋ	'field'
Bumthang	leŋ	(various)
PTB	*ləy	'bow'
Bumthang	li	(various)

Table 10. PTB *l > l

A further examples of the retention of *l is found in the ergative suffix *-le*.

Table 11 lists some regular correspondences between the different varieties presented here. Note that all of these processes are tendencies with variation in the other varieties; for instance, paddy rice is *mras* in Chumey and Ura, but *mrat* is heard in Ura on occasion. Final vowel loss is not a regular feature of Ura, but *kʰawa* 'chicken' is only ever heard as *kʰaw* in the compound *kʰawtɛ* 'chicken's egg', showing *-CV# > -C#. Final plosives (-p, -t and -k) are preserved in all of the highland varieties, but -k is lost in the personal perfective: the root *kʰrak-* 'arrive' can be realised as (for example) *kʰrakna* 'arrive (impersonal perfective)', or *kʰrak* 'arrive (impersonal irrealis)', but *kʰra:* 'arrive (personal perfective)' (Ura dialect).

	*-s	*kʰ	*l-	*kʰr	*ʈ	*-CV#	*-T	*kr-
Chamkhar	-t	kʰ	j-	kʰr	ʈ	-CV#	-T	kr-
Dhur	-t	h	j-	kʰr	ʈ	-CV#	-T	kr-
Chumey	-s	kʰ	j-	kʰr	ʈ	-CV#	-T	r-
Tang	-t	kʰ	j-	kʰr	ʈ	-CV#	-T	kr-
Ura	-s	kʰ	j-	kʰr	ʈ	-CV#	-T	kr-
Trongsa	-t	kʰ	j-	k	ʈ	-C(V)#	-Ø	k-
Henkha	-t	kʰ	l-, -n-	ʈʰ	tʃ	-C#	-Ø	(h)r-

Table 11. Some regular correspondences

2.4 Lexical comparisons

The wordlists were coded for 'plausible nested cognacy'. The 'plausible' part of the terms refers to the coding of different words according to whether the surface forms could reasonably be taken as representing variants of the same form, whether it was shared through inheritance of through diffusion (and so not necessarily 'cognate' in the sense usually used by linguists). The 'nested' part of the label refers to the fact that for some words, in which clear sound changes unify the correspondence sets, pairs of words could be reasonably judged as being more or less closely related.

		Cognacy			
	'six'	|grok|	|tsʰa|	|gr|>|ʈ|	|gr|>|r|
Chamkhar	ɡrək	1	0	0	0
Dhur	ɡrək	1	0	0	0
Brokkat	ʈu	1	0	1	0
Chumey	rək	1	0	0	1
Tang	ɡrək	1	0	0	0
Ura	ɡrək	1	0	0	0
Henkha	ʈu	1	0	1	0
Dzongkha	ʈu	1	0	1	0
Nepali	tsʰa	0	1	0	0

Table 12. Plausible cognate coding with 'six'

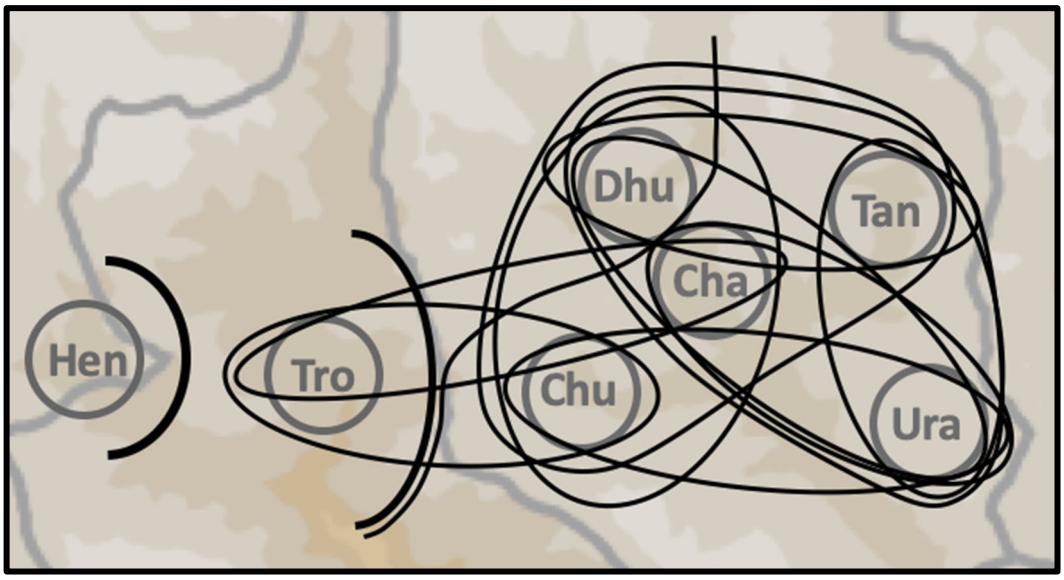
		Cognacy			
	'eight'	|gjad|	|aʈh|	|gj|>|dʒ|	|-ad|>|ɛ|
Chamkhar	dʒat	1	0	1	0
Dhur	dʒat	1	0	1	0
Brokkat	dʒɛ	1	0	1	1
Chumey	dʒat	1	0	1	0
Tang	dʒat	1	0	1	0
Ura	dʒat	1	0	1	0
Henkha	kɛ	1	0	0	1
Dzongkha	dʒɛ	1	0	1	1
Nepali	aʈh	0	1	0	0

Table 13. Plausible cognate coding with 'eight'

Following this coding, the degree of similarity between each pair of languages was calculated. Since different semantic fields show different patterns of diffusion, following the general examination of lexical similarity, and visualisations of that data, selected pairs of languages will be examined for lexical similarity by different semantic fields.

2.4.1 Lexical isoglosses

The lexical map of the Bumthang highlands is one that is full of clear divisions. Map 3 shows the lexical isoglosses that can be found in the data. Isoglosses which define only a single variety in the highlands have not been noted on the map; while Henkha and Trongsa have large numbers of individual-identifying isoglosses (44 for Henkha, 14 for Trongsa), there are many less in the highland varieties. Chamkhar and Tang are each defined by only one isogloss (Chamkhar has the form *kʰɐptiŋtiŋ* 'hot', while all other variety's forms are related to *tsʰan, Tang *jir-* 'run', vs. *dʒuk-*), Ura by two (*kar* 'run', *ʃiŋ burum* 'sugarcane'), Chumey by three (*tɔkpa* 'leg', *ʈadmɔ mu* 'wet', *bɔr-* 'dig'), and Dhur by four (*paspa* 'body hair, fur', *ama* 'wife', *niʃɔ* 'twenty', and *lɐktɔktɔk* 'good', of which 'wife' and 'twenty' are loans from Central Tibetan). Isoglosses that link two or more languages against the remainder are shown on the map; there are very few pairs of language varieties in the highlands that cannot be linked by lexical isoglosses, while there is a large bundle of isoglosses separating the highlands dialects from the Trongsa dialect.



Map 3. Lexical isoglosses amongst the seven languages exemplified

2.5 Lexicostatistics

The lexicostatistical relationships between the different languages are shown in Figure 5, which displays lexical similarity as a percentage.[13] We can firstly see a dramatic break between Henkha and the other languages in the sample; this is *a priori* unsurprising, given the greater distance between Henkha and the other languages. Trongsa is the next most lexically distinct language, again reflecting its relative isolation. Trongsa occupies a westerly position compared to the other Bumthang varieties, is not in the high uplands that form the core of the Bumthang region, and was a separate polity from the highland varieties. The strikingly higher percentage of shared forms with Tang, the most distant of the highland Bumthang varieties, most likely reflects shared retentions at the edges of the Bumthang spread. We have already seen the reasons why Trongsa is distinct from the highland varieties; within the highlands Tang was always a powerful local polity, at the northern end of a long valley quite removed from the main rice-growing region in the Chamkhar valley, and so more isolated from lexical changes that (we hypothesise) spread across the highlands. The highlands present a lexical core, consisting of Tang, Ura, Dhur Chumey, which all show lexical relationships at levels above 80%. Most striking is the exclusion of Chamkhar, which is geographically central to the highlands region (see Map 2), and which must be transited in travel between any two of the other sample locations, except between Ura and Tang (for which a high-altitude pass exists). Indeed, linking the varieties together requires us to place Chamkhar in a peripheral position, as shown in Figure 6 (where the thickness of the lines corresponds to the percentage of shared lexical items as shown in Figure 5).

[13] Here two forms are counted as 'similar' (as opposed to 'different', on a binary scale) if there is any trace of (plausible) cognacy when comparing two forms. All of the first eight rows in Table 12 were coded as showing similarity, regardless of the subsequent nestings.

	Chamkhar	Dhur	Chumey	Tang	Ura	Trongsa	Henkha
Chamkhar							
Dhur	82						
Chumey	82	83					
Tang	83	88	87				
Ura	82	85	87	87			
Trongsa	79	80	78	84	78		
Henkha	60	65	60	64	57	64	

Figure 5. Lexicostatistical relationships between the seven varieties

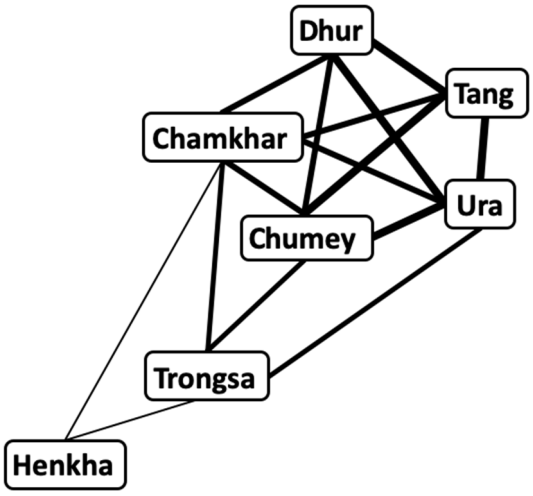


Figure 6. Lexical chains in the core seven varieties

Another way to display the lexical relationships that hold between the dialects is to plot them according to the lexical distance between each pair of languages, following multi-dimensional scaling. Figure 7 shows the relative position of the different languages to each other according to lexical distance, as measured by cumulative nested plausible cognacy measures; in this figure the where the distance displayed between points correlates strongly to the lexical distance, at $r^2 = 0.94$.[14] In this figure languages that are very similar to each other will be plotted very close, while the more distant the lexical relationship, the more distant they are in the figure. The correlation to the position of the different languages in space is strong (compare with the locations in Map 2), with the major exception being the position of Chamkhar, which is central to the highlands group in Map 2, but peripheral to it in Figure 7.

[14] The non-metric MDS analysis was implemented using the solver function in excel, with the optimal solution showing the greatest correlation of rank-ordered distances in the figure to the rank orderings of lexical distances.

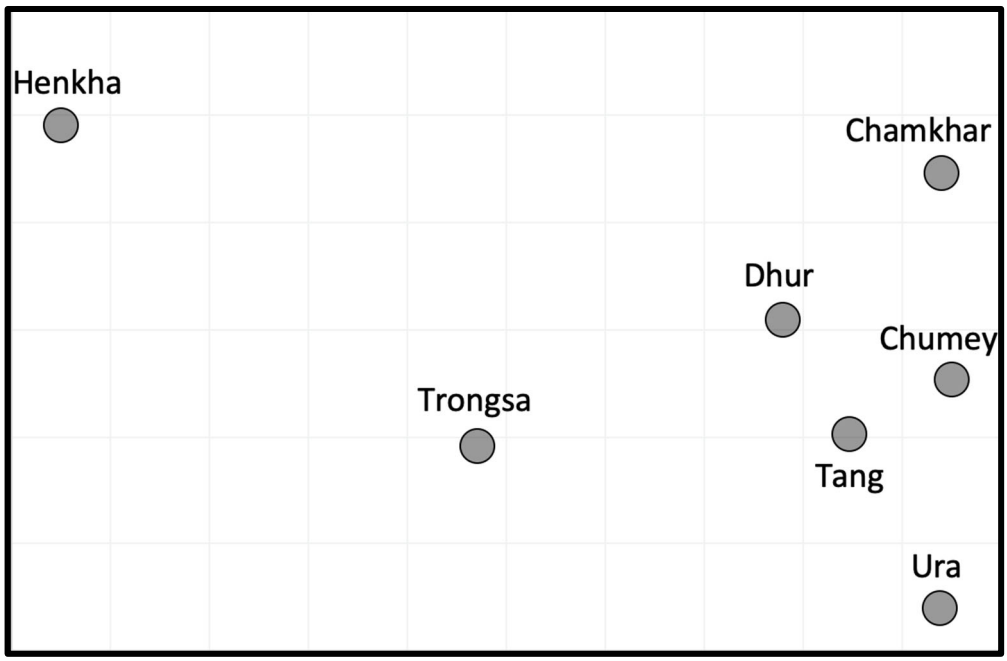


Figure 7. Plotting of language varieties following MDS analysis of lexical data

The answer to the apparent conundrum of the lexically peripheral position of the geographically central Chamkhar variety can be found when we examine the lexical cognacy found with the Central Tibetan varieties relevant to the area, shown in Figure 8. (Standard) Tibetan is relevant as a proxy for the liturgical language (Chöke) that is widely known (to some degree) in Bhutan; Dzongkha, which can be considered a (in some ways conservative) southern variety of Tibetan is the national language of Bhutan, and both a subject and a medium of instruction in schools. Brokkat and Chöcangacakha are local varieties spoken in northern Bumthang (Dhur village) and western Monggar (to the east of Bumthang), respectively. Brokkat is in close contact with the Dhur dialect of Bumthang thought their co-residence in the same village, and Chöcangacakha is in contact with Ura, the inhabitants of which pasture their livestock on the edges of the Chöcangacakha-speaking region in Monggar. Chöcangacakha in particular shows evidence of having adapted towards the East Bodish linguistic area, while Brokkat is not as lexically affected.

The highest cognacy with a Central Tibetan variety is found in Henkha, which is 46% cognate with Dzongkha. Henkha is spoken immediately adjacent to the eastern edge of the area in which Dzongkha is the native language of the Ngalops, and as a minority language without traditional political prestige it is not surprising that so many Dzongkha words have entered the Henkha vocabulary. The similarly high cognacy that Henkha has with Brokkat reflects that fact that Brokkat is closer to Dzongkha than it is to the other Central Tibetan varieties included in this comparison.

Trongsa shows a similar pattern to that seen in Henkha, but is slightly less extreme, averaging 39.5% with the four Central Tibetan varieties presented here. The only other high (above 40%) scores are between Chöcangacakha and the Bumthang varieties, though the score between Dhur and Brokkat is as high as between Trongsa and Dzongkha. Dhur has a higher value with Brokkat than the other highland languages because Brokkat is spoken in the same village as Dhur, and a number of Tibetan-isms have entered the language through contact with

the previously nomadic Brokkat speakers. Beyond this,[15] we also note that Chamkhar has lower percentages than the other languages with the Central Tibetan languages, averaging under 31%. The central position of Chamkhar in the Bumthang highlands has isolated it to some extent from loans originating in Tibetan-related varieties that surround Bumthang, and the lexically peripheral position of Chamkhar is due to its conservatism, rather than innovation.

	Chamkhar	Dhur	Chumey	Tang	Ura	Trongsa	Henkha
Chamkhar							
Dhur	82						
Chumey	82	83					
Tang	83	88	87				
Ura	82	85	87	87			
Trongsa	79	80	78	84	78		
Henkha	60	65	60	64	57	64	
Brokkat	35	39	36	38	36	**42**	**44**
Chöcangacakha	35	**41**	**41**	**41**	**42**	**41**	37
Dzongkha	30	37	32	35	33	39	**46**
Tibetan	23	30	25	27	27	30	31
Average	30.75	36.75	33.5	35.25	34.5	38.0	39.5

Figure 8. Comparison with Central Tibetan varieties

Dzongkha	27.48, 89.9, widespread	dzo
Standard Tibetan	29.65, 91.12, widespread	bod
Brokkat	27.66, 90.65 (Dhur village)	bro
Chöcangacakha	27.53, 91.14	cgk

Table 14. Central Tibetan languages

2.6 Wider comparisons

In addition to the Central Tibetan languages, a further set of languages were compared to the data presented here: the remaining East Bodish languages, and the other two languages spoken in Bumthang, Nepali and Hindi. Table 15 lists these languages, and Map 4 shows their locations (as well as the locations of the Central Tibetan languages) (Chöcangacakha is abbreviated to Chö.). Figure 9 shows the lexical similarities found between each pair of languages; the upper left part of the figure is the same as Figures 5 and 8, but Khengkha and Kurtöp have been included before Henkha, and the remaining three East Bodish languages interrupt the figure between Henkha and the Central Tibetan languages.

[15] The remaining values will be discussed in more detail in the next section.

East Bodish		
Khengkha	27.145, 90.69	xkf
Kurtöp	27.85, 90.97	xkz
Chali	27.32, 91.26	tgf
Dakpa	27.38, 91.82	dka
Dzala	27.79, 91.37	dzl
Indic		
Nepali	widespread	nep
Hindi	widely dispersed	hin

Table 15. Additional languages considered

Map 4. Locations of additional languages in Bhutan

The lexical relationships between the different languages in Figure 9 show the Bumthang languages in their fuller context. While the somewhat peripheral status of Chamkhar within Bumthang has already been seen, the data in Figure 9 allow us to see that the Bumthang varieties so far discussed are more similar to each other than they are to Khengkha and Kurtöp, though without the lexical bridge provided by Tang, Trongsa would be considered to be, if anything, more distant from Bumthang than either of Khengkha or Kurtöp. Henkha is closer to Bumthang (and Khengkha and Kurtöp) than Chali, and also shows some evidence, in its higher percentage of similarity with (more distant) Kurtöp than with neighbouring Khengkha, that suggests that the shared lexemes are retentions which have been lost in the Bumthang highlands. Chali is more lexically similar to the eastern group of East Bodish, Dzala and Dakpa, which are geographically closer. Rather than showing lexical similarities with Dzala, to its immediate east, Kurtöp shows an unexpectedly high percentage of shared vocabulary with Chöcangacakha, downriver from the Kurtöp position at the top of the valley. Lexically, Dakpa is only related to the other East Bodish languages through its connection with Dzala. The internal relationships of the Central Tibetan languages have already discussed, but Figure 9 allows us to see the degree to which

Chöcangacakha is lexically related to the eastern Bumthangic languages which are close to it, and Brokkat is lexically closer to the western ones which are closer to it.

	Cha	Dhu	Chu	Tan	Ura	Tro	Khe	Kur	Hen	Chl	Dza	Dak	Bro	Chö	Dzo	Tib	Nep	Hin
Cha																		
Dhu	82																	
Chu	82	83																
Tan	83	88	87															
Ura	82	85	87	87														
Tro	79	80	78	84	78													
Khe	75	78	80	81	80	75												
Kur	73	80	78	79	78	74	74											
Hen	60	65	60	64	57	64	61	60										
Chl	48	51	52	53	53	50	54	60	43									
Dza	44	49	45	49	49	48	45	46	48	51								
Dak	32	36	33	35	33	33	35	37	37	42	59							
Bro	35	39	36	38	36	42	40	39	44	30	38	25						
Chö	35	41	41	41	42	41	40	46	37	35	42	33	60					
Dzo	30	37	32	35	33	39	34	37	46	29	37	26	70	61				
Tib	23	30	25	27	27	30	29	37	31	25	37	27	59	53	59			
Nep	4	4	6	5	5	5	5	4	3	3	5	2	4	6	6	6		
Hin	5	5	5	5	5	5	5	5	4	3	5	2	4	4	7	7	61	

Figure 9. Wider lexicostatistical relationships of the languages

Abbreviations: Cha: Chamkhar, Dhu: Dhur, Chu: Chumey, Tan: Tang, Ura: Ura, Tro: Trongsa, Khe: Khengkha, Kur: Kurtöp, Hen: Henkha, Chl: Chali, Dza: Dzala, Dak: Dakpa, Bro: Brokkat, Chö: Chöcangacakha, Dzo: Dzongkha, Tib: Tibetan, Nep: Nepali, Hin: Hindi.

In Figure 10 the lexical relationships between the East Bodish languages described in Figure 9 are presented graphically. Dhur and Chamkhar are peripheral to a core Bumthang cluster consisting of Chumey, Tang, Trongsa and Ura. The next most close pair of languages are Khengkha and Kurtöp, which from a lexical perspective are sisters to the Bumthang varieties. Henkha joins the Bumthang group, but is not so closely linked, lexically, to the more distant Khengkha and Kurtöp, which only form a lexical group with Henkha when Chali is also included (that is, Khengkha and Kurtöp are approximately equally related, lexically, to both Henkha and Chali). Chali is loosely lexically associated with Dakpa, the language to its immediate east, and Dakpa is weakly linked, lexically, to Dzala, to the north-west, though Dzala does not share a close lexical relationship with Chali. The lexical cohesion of the East Bodish languages is low when Dzala and Dakpa are included, but it is still significantly higher than their lexical relationship with Central Tibetan.

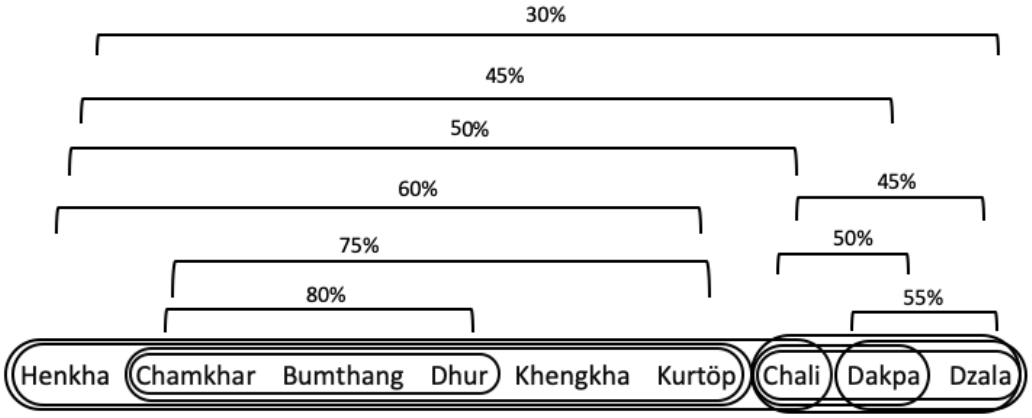


Figure 10. Lexical relationships between East Bodish languages

The spread of Central Tibetan lexical influence is shown in Figure 11, in which the lines correspond to lexical similarity values in Figure 9. The Central Tibetan languages considered here have a conservative core around Tibetan and Dzongkha, with Chöcangacakha most removed from this core, and Brokkat intermediate in its lexical distance. The closeness of Henkha to Central Tibetan reflects the position of Henkha on the eastern border of Wangdue Phodrang (དབང་འདུས་ཕོ་བྲང་), where Dzongkha is spoken natively. Trongsa is immediately east of the Henkha region, and next in terms of Central Tibetan influence, but Dhur is (Map 2) at the opposite end of the Bumthang range, at the upper end of the Bumthang valley. The Central Tibetan influence here comes from contact with Brokkat, the previously nomadic herding group that occupies half of the village of Dhur.

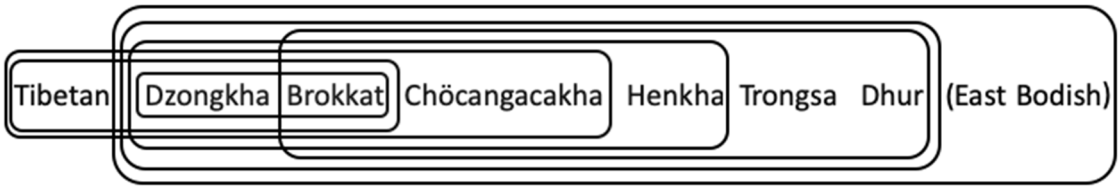


Figure 11. Central Tibetan lexical influence

If we confine ourselves just to the Central Tibetan varieties, it is clear that Dzongkha and Brokkat are very close (70%), and that Dzongkha, Brokkat and Chöcangacakha group together

with about 60% lexical similarity, and Tibetan is also close (60%) with Dzongkha and Brokkat. Chöcangacakha is not significantly close to Tibetan, indicating that while Brokkat is a (relatively) recent split from the Southern Tibetan group that contains Dzongkha, Chöcangacakha is not significantly closer to Dzongkha than it is to (Lhasa) Tibetan, and probably represents a Tibetan variety with a significant independent history.

Figure 12. Lexical relationships within the Central Tibetan varieties surveyed

2.7 Comparison by semantic fields

The following figures present the results of lexicostatistical analysis of semantic domains. In Figure 13 we see the results of the analysis of body part terminology; compared to the values seen in Figure 9 (abbreviations are the same as for Figure 9), all of the values are higher, especially so for the Central Tibetan languages. It is striking that Chöcangacakha shows a high rate of cognates with the Bumthangic languages to its west, as high as the range it shows for the Central Tibetan languages; as noted earlier, both Ura and Tang dwellers traditionally pasture animals in the Monggar lowlands, providing ample opportunity for contact. The Chöcangacakha body part lexicon provides a bridge between the Central Tibetan list of terms and the East Bodish set through the additional of East Bodish terms in a predominantly Central Tibetan system (such as Chöcangacakha *prenmuŋ* 'finger, Brokkat *pɹimaŋ*, Dhur *premaŋ*, Ura *pɻimaŋ*, compared to Dzongkha *dzimu*, Tromowa *dzumu*, and Lhasa Tibetan *dzuku*). Brokkat, despite the speakers inhabiting the same village as the Dhur dialect of Bumthang, shows no assimilation in terms of body part terminologies, and this difference in lexical diffusion is almost certainly related to the difference between the settled, agriculture-oriented Chöcangacakha speakers and the still semi-nomadic Brokkat speakers.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	95											
Chu	95	100										
Tan	95	100	100									
Ura	95	95	95	90								
Tro	90	95	95	95	85							
Chl	75	80	80	80	70	75						
Hen	65	65	65	65	60	65	60					
Bro	42	42	42	42	42	42	37	37				
Chö	61	61	61	61	56	56	44	39	59			
Dzo	50	50	50	50	50	45	40	50	68	56		
Tib	47	47	47	47	47	47	42	47	72	59	79	

Figure 13. Bumthang and surrounds: lexicostatistical relationships for body parts

When we examine the terms for humans and human kin in Figure 14, we find a number of surprising results. There is in general a high level of similarity between the East Bodish languages and the Central Tibetan languages, with the exception of Chamkhar and Chali. Chali is to the east, removed from intensive contact with Dzongkha, and Chamkhar is, as noted earlier, lexically conservative, and lacks many Central Tibetan loans. Showing the same relationship that was seen with body part terminologies in Figure 13, there is evidence of special assimilation between Brokkat and the Bumthang varieties, and between Henkha and Trongsa with Central Tibetan generally. Tang also shows a higher relationship with Central Tibetan in kin terminology.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	36											
Chu	45	73										
Tan	36	91	82									
Ura	55	73	82	82								
Tro	55	91	82	91	82							
Chl	36	45	45	45	45	45						
Hen	40	80	80	80	80	90	40					
Bro	44	78	78	78	78	89	22	89				
Chö	40	70	70	70	70	80	40	80	78			
Dzo	36	82	73	82	73	82	45	80	78	70		
Tib	27	55	45	55	36	55	27	60	56	50	55	

Figure 14. Bumthang and surrounds: lexicostatistical relationships for human and kin terms

Examining the terminologies for animals, we considerable convergence between Brokkat and Dhur, with Dhur borrowing the Central Tibetan terms at a greater rate than is found with other Bumthang or East Bodish varieties. We find, for example, Dhur *ba* 'cow' corresponding to Brokkat *bạ* and Dzongkha *pạ*, showing Central Tibetan influence on Dhur, but at the same time Brokkat *pɔ* 'snake' is a loan from East Bodish *pɔ*, only distantly cognate with Dzongkha *bü*, showing a loan from East Bodish into Brokkat. Henkha shows a significantly higher level of similarity with Central Tibetan, through loans from Dzongkha (e.g., *tʃʰi* 'dog', Dzongkha *rotʃʰi*, compared to general East Bodish *kʰwi*; also *pạ* 'cow', see above).

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	77											
Chu	93	85										
Tan	88	85	93									
Ura	94	85	100	100								
Tro	88	85	93	88	94							
Chl	75	85	87	88	88	75						
Hen	60	75	64	53	60	67	47					
Bro	60	70	60	60	60	60	60	70				
Chö	33	42	36	40	40	40	40	43	60			
Dzo	31	38	27	38	38	38	25	53	70	67		
Tib	31	46	33	38	38	31	31	40	80	47	56	

Figure 15. Bumthang and surrounds: lexicostatistical relationships for animal terms.

Plant terminologies show great cohesion amongst the Bumthang languages, including Trongsa, while the non-plateau Central Tibetan languages considered *all* show greater similarities with the East Bodish languages than they do with each other, particularly Lhasa Tibetan. Of the languages considered here only Chali, in a separate side-valley to the east of the Kuri river in Monggar, has a significantly different set of terms associated with plants.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	89											
Chu	78	78										
Tan	89	89	89									
Ura	78	78	89	89								
Tro	78	78	89	89	89							
Chl	44	44	56	44	56	56						
Hen	78	78	67	78	67	67	44					
Bro	67	67	67	78	56	67	33	67				
Chö	50	50	63	50	63	63	63	38	25			
Dzo	67	67	56	67	56	56	44	67	56	50		
Tib	25	25	25	25	25	25	38	25	38	43	38	

Figure 16. Bumthang and surrounds: lexicostatistical relationships for plant terms

When we confine the plant terms to consumed foods, we values that are on average 10% lower than for plants in general. The Central Tibetan languages show a clear division between Lhasa Tibetan, spoken on the Tibetan plateau, and the other three languages in the Bumthang valleys. Chöcangacakha in particular shows the use of many of the same food terms that are found in south and east Bumthang. Henkha has borrowed most of its food terminology from Dzongkha (e.g., *tɔ* 'cooked rice', *datsɪ* 'cheese'), and there is some evidence that Dhur too has borrowed plant terms (from Brokkat?) that are cognate with the Dzongkha terms (e.g., *ŋaŋla* 'banana', cf. Brokkat *ŋala*, Dzongkha *ŋaḻa*).

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	100											
Chu	78	88										
Tan	88	88	89									
Ura	78	75	80	78								
Tro	67	88	70	89	60							
Chl	63	63	44	44	56	56						
Hen	63	75	56	63	44	56	50					
Bro	29	33	29	29	29	29	29	50				
Chö	29	33	43	50	43	43	33	57	80			
Dzo	38	50	33	33	22	44	33	75	57	50		
Tib	11	13	10	11	10	20	22	11	29	14	33	

Figure 17. Bumthang and surrounds: lexicostatistical relationships for food terms

The highland Bumthang languages cohere together in their terms for the natural world, while Trongsa appears to have loans from Central Tibetan, such as *saŋmo* 'cloud', cognate with Dzongkha *sâmu*, compared to Bumthang *sartʃak ~ sartʃa ~ sɛrtʃɛ̀ ~* Henkha *sɛkɛ*, or *mukpa* (we note that Brokkat and Chöcangacakha both appear to have a loan from Bumthang or Kurtöp with *mukpa* 'cloud').

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	94											
Chu	83	83										
Tan	86	94	83									
Ura	90	100	83	100								
Tro	72	76	59	72	78							
Chl	38	39	33	43	43	39						
Hen	73	79	64	73	73	71	53					
Bro	47	47	35	41	47	44	35	54				
Chö	33	33	25	40	40	38	40	45	64			
Dzo	38	39	28	33	38	56	29	53	65	67		
Tib	38	39	28	38	38	50	24	40	65	60	67	

Figure 18. Bumthang and surrounds: lexicostatistical relationships for natural world terms

Material culture terms do not show the pattern of assimilation under contact conditions that are frequently reported. Rather, we see localised patterns in the distribution of material culture terminologies.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	89											
Chu	91	78										
Tan	89	88	78									
Ura	70	67	80	75								
Tro	55	56	55	67	60							
Chl	45	44	55	56	60	45						
Hen	44	44	44	50	44	44	33					
Bro	25	25	25	29	25	38	13	38				
Chö	0	0	0	0	0	0	0	25	50			
Dzo	27	33	27	33	20	45	18	44	75	50		
Tib	18	22	18	22	20	36	9	33	38	25	55	

Figure 19. Bumthang and surrounds: lexicostatistical relationships for tool terms

Terms denoting property concepts show a clear division between the Central Tibetan languages and Bumthang and the other East Bodish varieties, with only a slight level of assimilation found in Brokkat and Chöcangacakha.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	94											
Chu	83	83										
Tan	86	94	83									
Ura	90	100	83	100								
Tro	72	76	59	72	78							
Chl	38	39	33	43	43	39						
Hen	73	79	64	73	73	71	53					
Bro	47	47	35	41	47	44	35	54				
Chö	33	33	25	40	40	38	40	45	64			
Dzo	38	39	28	33	38	56	29	53	65	67		
Tib	38	39	28	38	38	50	24	40	65	60	67	

Figure 20. Bumthang and surrounds: lexicostatistical relationships for property terms

The numeral systems are similarly divided between the Central Tibetan languages and the East Bodish group, with limited diffusion of Central Tibetan numerals into some Bumthang varieties.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	82											
Chu	91	82										
Tan	100	82	91									
Ura	91	91	82	91								
Tro	100	82	91	100	91							
Chl	82	64	73	82	73	82						
Hen	64	55	64	64	55	64	45					
Bro	27	18	36	27	18	27	27	45				
Chö	36	45	45	36	45	36	27	18	64			
Dzo	27	18	36	27	18	27	27	45	91	73		
Tib	27	36	36	27	36	27	18	36	82	91	91	

Figure 21. Bumthang and surrounds: lexicostatistical relationships for numerals

The distribution of cognates in the class of verbs shows a pattern very similar to that found with property concepts, with an even sharper division between the two main language groups examined.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	89											
Chu	93	86										
Tan	86	89	93									
Ura	89	85	88	85								
Tro	89	88	88	81	84							
Chl	48	46	54	52	56	44						
Hen	79	83	79	75	71	79	58					
Bro	19	31	19	19	24	25	23	22				
Chö	22	22	28	22	24	22	28	24	53			
Dzo	14	25	21	17	22	22	24	29	73	61		
Tib	3	7	7	3	0	7	14	17	54	56	52	

Figure 22. Bumthang and surrounds: lexicostatistical relationships for verbs

Finally, with the set of interrogatives, we find a very strong Bumthang highlands group,and a strong Bhutan Central Tibetan group.

	Cha	Dhu	Chu	Tan	Ura	Tro	Chl	Hen	Bro	Chö	Dzo	Tib
Cha												
Dhu	100											
Chu	100	100										
Tan	100	100	100									
Ura	83	83	83	83								
Tro	83	83	83	83	83							
Chl	0	0	0	0	0	0						
Hen	67	67	67	67	67	67	0					
Bro	0	0	0	0	0	0	0	0				
Chö	33	33	33	33	33	33	0	33	100			
Dzo	0	0	0	0	0	0	20	0	33	33		
Tib	0	0	0	0	0	0	17	0	0	0	60	

Figure 23. Bumthang and surrounds: lexicostatistical relationships for interrogatives

In Figure 24 we examine the percentage of shared vocabulary in the different semantic fields examined between Brokkat and three Bumthang varieties, as well as two Central Tibetan varieties, drawn from the preceding figures. It is clear that the vocabulary in the 'body parts', 'food', 'tools', 'properties', 'numerals', 'verbs' and 'interrogatives' semantic fields are most clearly indicative of the language's Central Tibetan origin, while 'plants' (but not food plants) shows clear convergence to the Bumthang lexicon.

	body parts	human and kin	animals	plants	food	natural world	tools	properties	numerals	verbs	interrogatives	overall
Dhur	42	78	70	**67**	33	47	25	47	18	31	0	39
Chamkhar	42	44	60	**67**	29	47	25	47	27	19	0	35
Ura	42	78	60	**56**	29	47	25	47	18	24	0	36
Chöcangacakha	**59**	78	60	25	**80**	64	**50**	**64**	**64**	**53**	**100**	60
Dzongkha	**68**	78	70	56	**57**	65	**75**	**65**	**91**	**73**	**33**	70

Figure 24. Brokkat and other languages: lexicostatistical relationships by semantic fields

Similar figures could be constructed for other languages, but the overall patterns of differentiation across different semantic fields can be seen in Figures 13 – 23.

3 Conclusions

The wordlists presented here offer a source of lexical information for the different dialects of the Bumthang language that has not previously been available. The lexical relationships within these languages, combined with an analysis of sound correspondences, indicate a very long-term process of internal contact and dialect mixing. External contact with Central Tibetan languages has resulted in a lexical profile with those languages that is nearly as close as it is with the distant East Bodish languages Dzala and Dakpa, though we have the perhaps surprising result that the most prestigious variety of Bumthang, that of the central valley around Chamkhar, is the most lexically distinct from the other Bumthang valleys' dialects.

Differentially examining the lexicon by different semantic fields has shown that we can better understand the nature of the lexical relationships between the languages, showing how some semantic fields are better at showing the genealogical source of the language, while other semantic fields show different patterns of susceptibility to, or resistance to, borrowing.

References

Baxter, William H.; and Laurent Sagart. 2014. *Old Chinese: A New Reconstruction*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199945375.003.0004

Bosch, André. 2016. *Language contact in Upper Mangdep: a comparative grammar of verbal constructions.* Honours thesis, The University of Sydney. https://www.academia.edu/27358758

DeLancey, Scott. 2008. "Kurtöp and Tibetan". In Brigitte Huber; Marianne Volkart; and Paul Widmer (eds.), *Chomolangma, Demawend und Kasbek: Festschrift für Roland Bielmeier zu seinem 65. Geburtstag*, 29–38. International Institute for Tibetan and Buddhist Studies GmbH. DOI: https://doi.org/10.4000/abstractairanica.39099

Donohue, Cathryn; and Mark Donohue. 2016. "The variable use of the ergative case in Bumthang". *Language* 92 (1): 179-188. DOI: 10.1353/lan.2016.0004

Donohue, Cathryn; and Mark Donohue. 2019. "Extended Ergativity in Bumthang". *Himalayan Linguistics* 18(2): 57-72. DOI: https://doi.org/10.5070/H918243953

Hyslop, Gwendolyn. 2013. "On the internal phylogeny of East Bodish". In Gwendolyn Hyslop; Stephen Morey; and Mark Post (eds.), *North East Indian Linguistics*, 91-110. Foundation Books. https://www.academia.edu/4232991/

Matisoff, James. 2003. *Handbook of Proto-Tibeto-Burman : System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley: University of California Press. Retrieved from: https://escholarship.org/uc/item/19d79619

Michailovsky, Boyd; and Martine Mazaudon. 1994. "Preliminary notes on the languages of the Bumthang group". In Per Kvaerne, ed., *Tibetan Studies: Proceedings from the Sixth Seminar of the International Association for Tibetan Studies, Fagernes 1992*, Vol. 2: 545–557. Oslo: The Institute for Comparative Research in Human Culture.

Peck, Naomi. 2017. Predication, reference and modification: Major word classes in Bumthang, a Tibeto-Burman language. Honours thesis, Australian National University.

Shafer, Robert. 1954. "The Linguistic Position of Dwags". *Oriens, Zeitschrift Der Internationalen Gesellschaft Für Orientforschung* 7: 348–356. DOI: https://doi.org/10.1163/1877837254X00071

Suzuki, Hiroyuki; Keita Kurabe; Kazue Iwasa; Satoko Shirai; Shiho Ebihara; and Ikuko Matsuse. 2016. "A Geolinguistic Analysis of the 'Rice' Category in Tibeto-Burman". In Mitsutaki Endo (ed.), *Studies in Asian Geolinguistics II*, 37-51. Tokyo: Tokyo University of Foreign Studies [Research Institute for Languages and Cultures of Asia and Africa].

van Driem, George. 1995. *Een eerste grammaticale verkenning van het Bumthang, een taal van Midden-Bhutan, met een overzicht van der talen en volkeren van Bhutan (A first grammatical introduction to Bumthang, a language of central Bhutan, with an overview of the languages and peoples of Bhutan)*. Leiden: Centrum voor Niet-Westerse Studiën. (Republished in translation as: van Driem, George. 2015. *Synoptic grammar of the Bumthang language*. Himalayan Linguistics Archive 6. 1-77. Retrieved from: http://escholarship.org/uc/item/6h50n7v5 on 12/14/2015.)

van Driem, George. 2007. "Dzala and Dakpa form a coherent subgroup within East Bodish, and some related thoughts". In Roland Bielmeier; and Felix Haller (eds.), *Linguistics of the Himalayas and beyond*: 71-84. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110968996.71

van Driem, George. 2012. "The ethnolinguistic identity of the domesticators of Asian rice/ L'identité ethnolinguistique des domestiquateurs du riz asiatique". *Comptes Rendus Palevol* 11 (2-3): 117-132. DOI: https://doi.org/10.1016/j.crpv.2011.07.004

Wyatt, Thomas. 2017. *Syntactic Change in the Verbal System of Bumthang, a Tibeto-Burman Language of Bhutan*. Honours thesis, Australian National University. https://www.researchgate.net/publication/322400343

郑张尚芳 [Zhengzhang, Shangfang] (translated by Laurent Sagart). 2000. *The Phonological system of Old Chinese*. Paris: École des Hautes Études en Sciences Sociales.

郑张尚芳 [Zhengzhang, Shangfang] (translated by Laurent Sagart). 2003. 上古音系 [Shànggǔ yīnxì] (Old Chinese Phonology). Shanghai: Shanghai Educational Publishing House (上海教育出版社).

Mark Donohue
mhdonohue@gmail.com

Appendix

Bumthang and Henkha wordlists

Abbreviations used: eB: elder brother; yB: younger brother; eZ: elder sister; yZ: younger sister.

	English	Chamkhar	Dhur	Chumey	Tang	Ura	Trongsa	Henkha
1	Head	gujuŋ	gujuŋ	gujuŋ	gujuŋ	gujuŋ	gujuŋ	gunu
2	Forehead	ɸajlaŋ		pʰaj		pʰajlaŋ		
3	Hair	ɹa	ɹa	ɹa	ra	kra	kra	hra
4	Eye	mɛk	mɛk	mɛk	mɛk	mɛk	mɪŋ	mɪ
5	Ear	na	na	na	na	na	napɪŋ	nɪlz
6	Nose	naβaŋ	naɸaŋ	napʰaŋ	naβaŋ	napʰaŋ	naw̃	napa
7	Cheek		hʊr			kurtoktom	ŋur	
8	Chin		kamlaŋ			tʃum	kjaj zur	maŋkʰa
9	Lips	tʃɪ	tʃi	tʃi	tʃi	tʃi	ʃop	mamʃi
10	Mouth	kʰa	ha	kʰa	kʰa	kʰa	kʰa	kʰa
11	Tooth	kwa	kwa	kwa	kwa	kwa	kɔ	ʔwa
12	Tongue	li	li	li	li	li	li	li
13	Breast	dʒu	dʒu	dʒu	dʒu	dʒu	dʒu	nʊm
14	Stomach	dʒoma	dʒɔma	dʒɔma	dʒuma	dʒoma	dʒuma	pɔw
15	Shoulder	pɔŋma	pɔŋma	pɔŋma	pɔŋma	pɔŋma		
16	Arm (upper)	jak	jak	jak	jak	koksila, jak	jak	la
17	Arm (lower)	jak	jak	jak	tʰɔ	tʰomali, jak	jak	la
18	Elbow	grumaŋti	kumaŋ	rumaŋti	grumaŋdi	gumaŋti	tiwa	tsɪktɔ
19	Finger	primaŋ	premaŋ	pɻimaŋ	pɻimaŋ	pɻimaŋ	pruma	tsimɔ, tsɪktʃu
20	Fingernail	sima	sima	sima	sima	sima	simu	tsɪm
21	Leg (upper)	tawa	kü, tawa	ŋartɔŋ, tɔkpa	tawa	tawa / gwi	tawa	tɔw, pɛx

22	Leg (lower)	tawa	ŋartɔŋ, tawa	tɔkpa	tawa	nartɔŋ	tawa	tɔw, ŋala
23	Ankle	tiŋma		tɛkɔlɔŋ	tɛɣɔlɔŋ	takolıŋ	zur, kɔla	tsıktɔ
24	Skin	pakpa	pakpa	pakpa	pakpa	pakpa	pakpa	pɔɣɔ
25	Body hair	pɔ	paspa	pɔ	pɔ	puspai	pɔ	pɔ
26	Bone	rotpa	ɻutpa	rɔza	rœtpa	rospa	rɔspa	rɔtɔ
27	Chest	prandɔ	braŋtɔ	braŋdɔ	braŋdɔ	braŋta	braŋkɔ	praŋkɔw
28	Heart	neŋma				neŋ	nɛŋka	
29	Blood	kak	kak	kak	kak	kak	kah	ka
30	Urine					zeŋma		
31	Faeces					cok		
32	Knee	punmuŋ	pœnmuŋ	pœnmɔ	pünmuŋ	puspuŋ	putmo	ŋatɔ
33	Neck (front)	ɲœtpa	takpa	takpa	takpa	notpa	takpa	ügɔ
34	Neck (back)	takpa	takpa	takpa		takpa	takpa	taq
35	Liver	tʃʰimbo				ʐuwa	dʒɔma naŋtʰɛ	
36	Name	mɛŋ	mıŋ	mıŋ	mıŋ	meŋ	min	miŋ
37	Man	bɔdʒa	poʃa		bodʒa	pʰodʒa	bɔs	pusaj
38	Woman	minmboza	ame		mımboza, bɔmıt	memboʒa	ama	amaj
39	Child	oŋa	ɔŋa	oŋa	oŋa	oŋa	oŋa	ʊptʃi
40	Father		apa	aβa	apa	apa	apa	ap
41	Mother		ama	ama	ama	ama	ama	am
42	eB		atʃo	nɔ	adʒɔ	atʃo	atʃɔ	adʒɔ
43	yB		nɔ	nɔ	nɔ	nɔ	nɔ	nʊdʒo
44	eZ	aʃɛ	tʃʰɛ	aʃɛ	tʃʰɛ	aʃe	atʃɛ	aʒi
45	yZ		nɔmıt	nɔmɛt	nɔmet	nɔmet	nɔmɛ	nɔmɛ
46	Son					bo		
47	Daughter					bome		
48	Husband	makpa	makpa	mɔkpa	makpa	makpa	makpa	

49	Wife	nesa	ama	nɛmɔ	nɛmɔ	ŋamo	nɛsaŋ	
50	Boy			bɔza		boza		
51	Girl			bɔmɛt		bome		
52	Mother's mother		ʔajɔ	aja	ajla	aja	aja	aŋgje
53	Father's father		mɛmɛ	mimi	mɛmɛ	mɛma	mɛmɛ	akpa
54	Child's child			dımıt	dımıt, diβu	dımıt, diβu	ɔŋa tʃıŋku; tıpu, tımu	dibam, dip
55	Mother's brother					aʃaŋ		
56	Mother's sister					amem		
57	Father's brother					aku		
58	Father's sister					nena		
59	Wife's sister's husband					atʃʰo, nɔ		
60	Friend		tʃarɔ	tʃʰarɔ	tʃarɔ	gola	gɔn	
61	Person	mi	mi	mi	mi	mi	mi	mi, pɔstsɛ
62	I	ŋat	ŋat	ŋat	ŋat	ŋat	ŋat	ŋa
63	You	wɛt	wɛt	wɛt	wɛt	wɛt	üɛt	ji
64	He/She	kʰit, gun	hıt	gɔn	kʰit	gɔn	kʰit	kʰi
65	We	kɔβa, pura	ŋet	ŋjɛra	ŋet	ŋir	ŋet	jit
66	You (plural)	win	wıŋambɔ	ın	wın	win	üɛt rugɔ	jit
67	They	gə̃je	gonegambɔ	kɔnɛ	gɔne	gɔne	bɔt	ozɛ mi
68	Fish	ɲa	ɲa	ɲa	ɲa	ɲa	ɲa	ɲa
69	Chicken	kʰawa	hawa	kʰawa	kʰawa	kʰawa	kʰaɣa	qʰɔw
70	Egg	kawti, te	hawtɛ	kʰawate	kʰawdı	kʰawte	tɛ	tɛ
71	Cow	nor	ba	nɔr	nɔr	nor	nɔr	pa̤
72	Buffalo					mahe		
73	Milk	dʒu	dʒu	dʒu	dʒʊ	dʒu	dʒu	tʃu
74	Goat	lele	ratsʰa, lɛlɛ	lɛlɛ	raza	ratsa, lɛlɛ	lɛlɛ	lɛlɛ
75	Horn					ro		

76	Tail	miɸaŋ		ɲiɸaŋ	miβaŋ	miphaŋ	dʒuma	dʒʊm
77	Dog	kʰwi	kʰwi	kʰwi	kʰwi	kʰwi	kʰü	tʃʰi
78	Snake	pɔ	pɔ	pɔ	po	pɔ	pɔ	pɔ
79	Monkey					pra		
80	Mosquito	zendum			sɛŋdɔm	zendom	dzenduŋ	
81	Ant	ʂɔŋ	bɻuktula	bruktula	bruptula	broktola	buktüla, dʒɔmɔ	
82	Spider	ɸrumzam		pʰrumzaŋ	pʰrumzaŋ	pʰromsaŋ	pʰrumsaŋ	patsʰam
83	Bird	tʃəwja	dʒiwɔ	dʒawja	dʒawja	dʒawja	dʒa	dʒɪdʒʊ
84	Wing	ʂɔkpa	ʃokpa	ʃokpa	ʃokpa	ʂɔkpa	ʃokpa	ʃok
85	Louse	sek		sɛk	sɛk	sɛk	sɛk	si
86	Chicken louse		ɻüit					
87	Yak	jak	jak	jak	jak	jak	jak	ja
88	Female yak	bɻe		brɛ		bɻe		
89	Fly	braŋ		braŋ	braŋ	braŋ	braŋma	bram
90	Horse	ta	ta	ta	ta	ta	ta	ta
91	Rat	ɲiwa	niwɔ	ɲiwa	ɲiwa	ɲiwa	nɛ	ŋɪ
92	Pig	pʰak	pʰak	pʰak	pʰak	pʰak	pʰak	pʰa
93	Cat	ʒimbuta		ʃɪmnɛ	ʃimbula	ʃimbala	dʒɪmdʒa	ʃumbal
94	Ox	bari			baɾi	laŋo	bari	laŋ
95	Bull	toɣa	baɾi	baɾi	baɾi	laŋgu		
96	Calf	bawja				bawja		
97	rooster	kʰadir				khari		
98	Insect					ʐoŋ		
99	Dragonfly	namɖu				dʊndʒaɸaŋ		
100	Butterfly	piβliŋ	piβiliŋ	pipilɪŋ	pipilɪŋ	pipeliŋ	tʃiŋla, pipila	tʃɪmla
101	Firefly					gami ʐoŋ		
102	Leech	pat				pat		
103	Food	ʒego			sani	ʃeɣɔ	zama	

104	Banana	ŋala	ŋaŋla	tʃae	ŋaŋla	tʃej	tʃaj, naŋla	ŋaḻa
105	Betel nut	pan, dɔma		dɔm	dɔma	dɔma	dɔma	
106	Chewing betel nut		dɔm	kʰadʒi	kʰadʒu			
107	Betel leaf		pan		pan	pan	pan	panɛ
108	Lime	tsun		tsʊn	tsʊn	tsʰʊn	tsuna	
109	Wheat	kar	kar	gɔ	nat	kar, gɔ	nas	kaɾ
110	Millet	n'am		nas		kɔŋbɔ	brakma	ʈɛ
111	Paddy	leŋ	ʈʊŋ	iba	iβa, mrat	mras	mras, iba	sem
112	Unhusked rice	iβa	ʈʊŋ	tʃaraj		mras	ʈʊŋ	
113	Husked rice	ʈuŋ	ʈʊŋ	tʊŋ	ʈʊŋ	ʈʊŋ	ʈʊŋ	tʃʰʊŋ
114	Cooked rice	zama	zama	zama	zama	zama	zama	tɔ
115	Potato	ki	ki	ki	ki	ki	ki	ki
116	Peanut					badam		
117	Chilli	baŋgala	baŋkala	baŋgala	paŋgala	baŋgala	baŋgala	baŋgal
118	Garlic	ʈhaɖu	ʈhaʈu	ʈɛɖʊk	ʈhadʊ	ʈhadu	ʈhaɖu	tʃɔɣɔp
119	Sweet potato	ki ŋambala		ki ɲɔksa			ʒim	kʰjowa ŋam
120	Maize (corn)	aʂam		aʃʃam	aʃam	aʂam	tʃadama	kjatʊm
121	Buckwheat	branma	braŋma	bɾatm	branma	brasma	branma	brɛm
122	Bamboo (big)	ruj	rü	ɻü	rüj	ruj	ruj	ɻü
123	Bamboo (small)	mek				mɛk	mɛ	miks
124	Cheese	pʂɹum	pʰʃrum	pʰrʊm	pʰrʊm	pʰrum	pʰrʊm	datsı
125	Ginger	saɣa	saxa	saxa	saxa	saka	saɣa	saxa
126	Turnip					bawa		
127	Radish	jawa	jawa	jawa	dʒawa	jawa	dʒawa	
128	Spring onion	tsɔŋ	tsoŋ	tsoŋ	tsoŋ	tsoŋ	tsoŋ	ma

129	Greens (spinach)	pœsœ	pɛsœ, pœsœ	pets^{h}e	pɛts^{h}ɛ			
130	Pumpkin	kaɣo̯ru	kakur	kako̯ra	kako̯rʊ	kaɣo̯ru	kakuri	kakʊɻ
131	Beans	ʃɛβɛn	ʃɪbɪn	ʃɪpɪn	ʃjɛβɛn	ʃaβaj	namɪt, ʃɛpɛn	sɛmtʃu
132	Sugarcane		kuntʃa	kutʃa	guʒa	ʃiŋ burum	guta	kutʃa
133	Onion	pjɛs		pɛs	pjɛɛs	pjes	gɔkpa	
134	Cauliflower					metokopi		
135	Tomato					lambenda		
136	Cabbage					kopi		
137	Oil					makho, rupot		
138	Salt	ts^{h}a	ts^{h}a	ts^{h}a	ts^{h}a	ts^{h}a	ts^{h}a	ts^{h}a
139	Flesh	ʂa	ʃa	ʃja	ʃja	ʂa	ʃa	ʃja
140	Fat (on meat)				got			
141	Seed	sɔn	sɔn	sɔn	sɔn	sɔn	sɔn	sœn
142	Bark						(pakpa)	
143	Barley	nat			tʃaræa	dʒaraj	nas	neʂ
144	Butter					mar		
145	Sun	ŋjɛ	nɛ	nɪ	nɛ	na	nɛ	nɛsɛ
146	Moon	la, lataŋ	la	lataŋ	la	laj	la	lɛw
147	Sky	nam	nam	nam	nam	nam	la, nam	nam
148	Star	karma	karma	karma	karma	karma	karma	kam
149	Rain	üœ	jœ	jœa	jœ	joj	üɛ	ü
150	Water	k^{h}we	hwe	k^{h}we	k^{h}we	k^{h}we	k^{h}œ	k^{h}ɛ
151	River	k^{h}we mɔtʃ^{h}u		k^{h}wɛ ɣaŋ	k^{h}we	matʃ^{h}u	k^{h}œ juruba	k^{h}ɛ
152	Cloud	sartʃak	sɛrtʃɛ	sartʃak	sartʃa	sartʃa	saŋmo	sɛkɛ
153	Rainbow					gɔŋʒam		
154	Wind	ʃalɔŋ, luŋ	ʃalɔŋ			lɔŋ	ʒɛlɔŋ	nilzu

155	Stone	gɔr	kɔr	gɔr	gɔr	gɔr	gɔr	gɔr̥
156	Sand	bema	dʒɛm	bema	bɛma	bama	bɛŋma, dʒɛm	tʃɛm
157	Mud	dam			dam	dam	dam	
158	Dust					tsʰuma		
159	Tree	sɛŋ	sɛŋ	sɛŋ	sɛŋ	sɛŋ	sɛŋ	sɪŋ
160	Leaf	lamba	lamba	lamba	lamba	lampa	lamba	tɛma
161	Root	rado	sɛŋ ratɔ	rato	radɔ	ra	tarɔ	ra(dɔ)
162	Thorn	tsaŋ	tsaŋ	tsaŋ	tsaŋ	tsaŋ	tsaŋ	tsɔŋ
163	Flower	mento				mento		
164	Soil	sa	sa	tʰaŋ	sa	sa	sa	sa
165	Mountain (g'ang)	kaŋ	gaŋ	katpa	ri	gaŋri	kaŋ	
166	Mountain (ri)	ri			ri	ri	ri	
167	Mountain Pass	ja	ja	ja	ja	ja		
168	Snow	ka	ka	ka	ka	ka	ka	ka
169	Lake			tsʰɔ	tsʰɔ	tsʰɔ	kʰœ dʒamtsa	tsʰɔ
170	Ice	kʰitpa	kʰɪkpat	kikp	kitpa	kitpa	mɪk	kʰi
171	Frost	tʃʰawa		tʃʰakpa	tʃʰawa	tsʰawa	tʃʰakpa	
172	Air					lɔŋ		
173	Forest	buŋ	buŋ	bʊŋ	bʊŋ	buŋ	tsʰɛma, bum	
174	Valley	tʰaŋ				krɔŋ		
175	Cliff	bɻak	bɻak	bɻak	bɻak	bɻak	brak	
176	Village	kroŋ	kroŋ	rɔŋ	kroŋ	kroŋ	kroŋ	saʃ:
177	House	maɛ	mɛ	mɛ	mae	maj	maj	mɛ
178	Roof	tʃaʃɔ	tʃʰimtɔk	tʃaʃu	mʊŋma	maklam	ʃap	titʃɛ
179	Door					ko		
180	Broom					pʰiksaŋ		

181	Mortar					togor		
182	Pestle					tomʃɛlɛŋ		
183	Pata (knife)	pataŋ	pataŋ	pataŋ	padaŋ	pataŋ	pataŋ	pata
184	Sheath (pata)	ʂup		sʊp	ʂʊp	ʂʊp		
185	Kitchen knife	jurba	jürʁba	juɾβa		katila	katila, rüp, zoɾwa	kutʃu
186	Axe	tari	tari	tari	tari	tari	tari	taɾ
187	Rope	taɣpa	takpa	takpa	takpa	takpa	takpa	tap
188	Hammer					towa		
189	Bow	li	li	li	li	li	ʒu	lzi
190	Arrow	mjɛwa	mɛwa	ɲa	mɛwa	ɲa	da	mra
191	Target	ba		pa̤	pa	ba	ba	
192	Quiver	surɛ				surɛ	daj ʃup	
193	Thread					kronman		
194	Needle					khap		
195	Ring	laxam		dʒəskam		dʒəskam	tsɔdʊŋ	
196	Path/road	jam	jam	jam	jam	jam	jam	lam
197	Field				leŋ			
198	Fire	ɡami	ɡami	ɡami	ɡami	ɡami	ɡami	ɡam
199	Smoke	duwa	towa	duwa	duwa	duwa	duwa	mixu
200	Ashes	tʰabdʒa	taptʃi	taptʃa	taptʃa	tʰondʒa	taβaj	kɔtʰɛw, tʰɛw
201	Gold					sɛr		
202	Silver					ŋoi		
203	Above	namdo, dʒaɛdo	jaw	namdɔ	dʒaɛdɔ	dʒaɛdo	dʒaɛ	tʰɔxar
204	Below	sutu	mœw	ɾaɾa	sʊttʊ	suttu	suttu	wak
205	In front	dɔŋɔ	tɔŋɔ	dɔŋɔ	dɔŋɔ	dɔŋɔ	dɔŋgwœ	dɔŋkaɾ
206	Behind	kʰaɛdɔ	kjædɔ	kajdɔ	kʰaɛdɔ	kajdɔ	kʰajdɔ	kɛsik

207	Inside	naŋɔ	naŋɔ	naŋɔ	ŋaŋɔ	naŋɔ	naŋ	nasunaŋ
208	Outside	paŋna, bitu	pitʊ	pɪttɔ	pitʊ	bitʊ	paŋɔ	tʃʰɛlɔk
209	Tibet			pœ	pœ	bot	bɔtpa	pœp
210	India			ɪndia	dʒakar(pa)	dʒa	tʃaharpa	tʃaxap
211	This	tsʰaɛ	mɛ	tsʰaɛ	osaɛ	tsaɛ	oʃo	o
212	That	utuj, seŋ	tʰü	tü	utuj	utüj	bɔt	utü
213	Here	tsʰɔ	tsʰɔ	tsʰɔ	tsʰɔ	tsʰajnana, tsʰɔ	tsʰɔr	ozɔ
214	There	utuj	utuj	tu	tu	tujnana	tʰʊɾ	tʰu, bœ
215	White	kartila	harti	kʰarti	karti	karti	karti	kat
216	Black	ŋündila	ɲɔndi	ɲünde	ɲündi	ɲɔndi	ŋündi	nakʰɛ
217	Red	ʃindila	ʃindi	ʃindi	ʃindi	ʃindi	ʒʊndɪ	ʒin
218	Green	ŋundi	dʒaŋku	dʒaŋku	dʒawŋgu	dʒaŋsɛr, dʒaŋku	undi	dʒaŋkɔ
219	Yellow	sirtila	sɪɾti	sarti	sɪɾti	sɪri	sirtila	sɪt
220	Blue	dʒaŋgu	ŋunti	ŋundi	ŋundi	ŋʊndi	ŋundi	ŋʊm
221	One	tʰek	tʰek	tʰek	tʰek	tʰek	tʰek	ti
222	Two	zon	sɔn	sɔn	sɔn	zɔn	zɔn	zœn
223	Three	sum	sum	sum	sum	sum	sum	sum
224	Four	blɛ	blæ	blɛ	blɛ	blɛ	blɛ	pr̥ɛ
225	Five	jaŋa	jaŋa	jaŋa	jæŋa	jaŋa	jaŋa	laŋ
226	Six	grok	grok	rɔk	grok	grok	grok	ʈu
227	Seven	nit	nit	nit	nit	nɪs	ŋɪs	nis
228	Eight	dʒat	dʒat	dʒat	dʒat	dʒat	dʒat	kɛ
229	Nine	dɔɣɔ	dɔkɔ	dɔɣɔ	tɔɣɔ	dɔɣɔ	dogo	tox
230	Ten	tʃe	tʃʰe	tʃʰɛ	tʃʰe	tʃʰe	tʃʰɛ	kʰɛptʃɛ
231	Eleven					tʃʰware		
232	Twelve					tʃʰwanis		
233	Thirteen					tʃʰusum		

234	Fourteen					tʃʰɛbɛ		
235	Fifteen					tʃʰɛgrok		
236	Sixteen					tʃʰéŋa		
237	Seventeen					tʃʰunis		
238	Eighteen					tʃʰɛrjat		
239	Nineteen					tʃʰɛdɔɣɔ		
240	Twenty	kʰajdi	niʃɔ	kʰajtʰɛk	kʰajde	njɪʃo	kʰajtɛ	kʰɛdi
241	One hundred	kʰaj jaŋa		tʃɪkdʒa	kʰaj jaŋa	kʰaj jaŋa	kʰaj jaŋa	kʰɛdɛ
242	One thousand	tɔŋ ʈadi		tɔŋʈa tʰɛk	ʈɔŋʈa tʰɛk	toŋʈa tʰɛk		
243	Many					dʒakse-na		
244	All					saŋsaŋ, zadɔ		
245	Old (things)	dʒikpala	tʊkpa	manba	taŋbɔ	tukpɔ	manba	ŋɪp
246	Old (people)	manba	katpa		katpɔ	gakpola		katpu
247	New	sarba	saɾpa	saɾβa	sarba	sarba	sarba	sap
248	Young	ʃunbɔ	ʃjömba		ʃönbɔ	ʒonbu	ʒonmula	
249	Good	katʃin	lɛktɔktɔk	katʃɪn	katʃan	katʃɛn	katʃan	kʰɛtʃɛ
250	Bad	ʃuŋma mot	tüœ haka	katʃɪn mɪn	ʃjanbu, tamasip	matʃɛn mot	gɔkʃa	ɔpɛ
251	Wet	ʃɪrβan	ʃɪrba	ʈadmɔ mu	ʃɪrβaŋ	ʃɪrban	ʃɪrβa	ʃɪp
252	Dry	kʰam	kampan	lɔkpa	kam	komban	kampan	kam
253	Long	haŋmala	xaŋmalɔ	kʰaŋmala	rɪʃɪŋ	kʰaŋla	rɪŋʃila	ɻɪŋk
254	Short	tʰiŋgula	tʃiŋkulɔlɔ	tʰiŋgula	tʰiŋgula	tʰiŋgu	tʰiŋku	tʰʊŋkʰu
255	Hot	kʰɛptiŋtiŋ	tsʰanma	tsʰan	tsʰanma	tsʰɛnma	tsan	zɛ
256	Cold	ŋaqpa	ŋakpa	ŋakpa	kʰɪksa	ŋakpa	ŋakpa	kɪk
257	Blunt	ja mot	ʃukmɔli		ɔja mut	ja mot-sa	kʰa mɛ-tata	kʰam
258	Sharp	ja	ʃukpɔ	ʈadmɔ	ja nɔtɔ	ja	ŋöŋ tʃɔptʃɔp	kʰatʃʰɛm
259	Blind	miluŋ			lɔŋma	lɔŋma		
260	Deaf	tiβali			tiwli	tiwli		

261	Right	œmbalək	jɛba	(tu): away from river	tʰʊru (downriver)	ɛbi, tsʰɔɣələk	üɛtpa	ekatɔ
262	Left	œβalək	jœmba	(raɔ): towards river	kaksɔ (upriver)	ɛmbu, tuɣələk	üɛnmɔ	œŋkatɔ
263	Near					ʈando		
264	Far					tarénʃiŋ		
265	Big	tʃɛtpɔ, dɔŋbɔ	tʃɪkpala	dʒɪkpala	dʒikpala	dʒikpala	tʃʰetpɔ, dʒɪkpala	bɔm
266	Small	tʃiŋgula	tʃɪŋkula	tʃiŋgula	tʃiŋgula	tʃiŋgula	tʃüŋku	tʃʊŋku
267	Heavy					dʒutjan		
268	Light					dʒaŋsoma		
269	Same					tsok		
270	Different					soso		
271	Whole					randʒuŋ		
272	Broken					ʈombigé		
273	Full					nanzo		
274	Round					kirtiliŋ		
275	Eat	su-	su-	zu-	zœ-	zu-	zu-	zü-
276	Hungry					bru nat		
277	Drink	tʰoŋ-	tʰoŋ-	tʰoŋ-	tʰoŋ-	tʰoŋ-	tʰoŋ-	tʰoŋ-
278	Sleep	dɔt-	dɔt-	dɔt-	dɔt-	dɔt-	dɔt-	dü-
279	Sit	ŋit-	nit-	ni-	nit-	nit-	nis-	nit-
280	Give	bi-	bi-	bi-	bi-	bi-	bi-	dʒi-
281	Burn (fire)	bar-	bar-	baɾ-	baɾ-	baɾ-	bar-	bak-
282	Die	se-	sɛt-	sɛt-	sɛ-	sɛ-	sɛ-	sɛt-
283	Kill	sut-	su-	su-	su-	sut-	sut-	sü-bi-
284	Fly	liŋ-	liŋ-		lɪŋ-	lɪŋ-		pʰru-
285	Walk					go		
286	Run	dʒuk-	dʒuk-	tʃuk-	jiɾ-	kaɾ tʰʊŋ-	dʒuk-	dʒu-

287	Go	gaɛ-	gɛ-	gaɛ-	ga-	gae-	gaj-	ga
288	Come	ra-	ra-	ra-	ra-	ra-	ra-	ra-
289	Climb	kʰraŋ-	hak-	kʰraŋ-	kʰa-	kʰaŋ-	kraŋ-	
290	Arrive	kʰrak-			kʰra-	kʰrak-	kʰrak-	
291	Step over	kim-	gim-			jɛɾ-	kim-	
292	Speak	lap-	lap-	lap-	lap-	lap-	lap-	mɔk-
293	Hear	ŋe tʰu-	kɔ-	tu-	nɛ dut-	tʰut-	kɔs-	kɔ-
294	See	tʰuŋ	tʰuŋ	ta-, tʰʊŋ-	tʰuŋ	te-, tʰuŋ-	tʰuŋ-	tʰʊŋ, ta-
295	Know	bran-	bran-	bran-	bran-	bran-	bran	brɛŋ-
296	Swim					tʃaj tʰʊŋ-		
297	Stand					jaŋ-		
298	Tired/Sleepy	uduk-	utuk-			uduk-	uduk-	
299	Wake up (intr.)	jaŋda-	jaŋ-	jaŋ-	jaŋ-	jaŋ	jaŋ-	
300	Wake up (tr.)	ruŋ-	ruŋ-	rʊŋ-	rʊŋ-	kruŋ-	ruŋ-	
301	Wash (self)	kʰrɔ-	kʰra-	kʰrɔ-	kʰröp-	kʰrɔs-	kɔ-	ʈʰo-
302	Scratch	tʰek-, brat-	bras-	bra-	bra-	bra-	brat-	tɪ-
303	Keep	blɛk-	blɛk-	blɛk-	blɛ-	blɛk-		bɻɪ-
304	Dig	gu-	ku-	bɔɾ-	ku-	ku-	ku-	
305	Hit		tʰuŋ-	daŋ-	tʰuŋ-, daŋ-	tʰʊŋ-		
306	Laugh	ga-	ka-	ga-	kạ-	ga-	kʰit ka-	ga-
307	Cry	ŋɔ-	ŋɔ-	ŋɔ-	ŋɔ-	ŋɔ-	ŋɔ-	ŋu-
308	Cook	kʰɛr-	kʰɛr-, tsɔk-	kʰɛɾ-	kʰɛɾ-	kʰɛɾ-, tsʰɔk-	kʰɛɾ-	kʰik-
309	Who?	aɛ	aj	æɛ	aɛ	ae	aj	ɛ
310	What?	ʂa	ʂa	ʃæ	ʃja	ʐa	ʒa	ʃɛ
311	Where?	aɔ	aw	aw	aw	ao	aw	ɔ
312	When?	aɾβa	aɾβa	aɾβa	aɾba	arba	arba	üwɛ
313	How many?	akpurɛ	akpudɪ	akpa tɛ	akpa	akor	art	aʃtɛ

314	Why?	ʂa βu-zi	ʂa pu-zi	ʃa buzɛ	ʃa buzi	ʂa buzi	ʒaɣuzi	ʃɛ bu-s
315	Language	kʰa	ha		kʰa	kʰa, pʰalap	kʰa	
316	No/not	mut	mot, mɪn	mɪn, mibu	mɪn	min	me	mi
317	Day	ŋjɛn	nɛnʈa	nɛnʈa, nœnʈa	nɛnʈa	nɛnʈa	nɛm	
318	Night	san	sʊlla	sʊnla	sulla	sula	sula	
319	Morning					ŋazi		
320	Yesterday	dɛma			daŋma	daŋma		
321	Today	dusum			dusum	dusum		
322	Tomorrow	jambat			jambat	jaŋbat		nɛmbɛ
323	Week	hapta	hapta	hapta	hapta	hapta	hapta	hapta
324	Month					la		
325	Year					nɛŋ		