DIAGNOSING A CONTACT HISTORY FOR TSUM

Dubi Nanda Dhakal, Mark Donohue, Bhojraj Gautam and Naijing Liu

Tsum, a language if the Nepal-Tibet borderlands in upper Gorkha, has been classified as a Central Tibetan language, closely related to Kyirong. The lexicon and nominal morphology (both forms and functions) of the language support this classification. We argue that, based on evidence from tonal phonology and verbal morphology, that there is a conflicting signal in the data. Given the known ease with which lexical material can be borrowed, we argue that the evidence of the verbal agreement system, and the complicated (and contradictory) tone system should be taken as more indicative of the language's origins. We suggest that the search for a single origin is not meaningful, and that the Tsum data strongly suggests that the phenomenon of the mixed language is more frequent than has been generally described, and that the Himalayas, with its long history of social mixing and blending, has a large concentration of such languages, many of which will prove to be the last reflections of since submerged linguistic ecologies that predate the dispersal and dominance of Tibeto-Burman and Indo-Aryan in the region, just as the modern isolates Kusunda and Burushaski more directly reflect such earlier language areas and genealogies.

Keywords: contact, mixing, ecology, dispersal

1 Language contact and language family affiliation

Language contact is often invoked to 'explain' those features of a language that cannot be easily ascribed to material that the language inherited from a prior ancestor, allowing for changes that accrued along the way. We advance the idea that 'language inheritance' is not, in fact, a simple statement, and that a language can be best affiliated in more than one direction. We suggest that different parts of the Tsum language have different origins, and that Tsum can be accurately described as a mixed language, the result of early contact, with an ultimately non-Tibetan origin. The question remains as to which languages have a prior relationship with Tsum, through genealogical affiliation or a period of social and linguistic contact.

2 Tsum

Tsum is a Tibetan language of upper Gorkha, spoken in the Tsum valley, at the very border of Nepal and Tibet in the greater Manaslu region[1]. The valley is connected by a valley joining the main Budhi Gandaki valley, and thence Gorkha to the south, which is accessible year round, and via three high altitude passes that allow seasonal access to Tibet. It is generally classified as firmly embedded within Central Tibetan, as 'Tibeto-Burman, Bodic, Bodish, Central Tibetan, Kyirong-Kagate' (also see Bradley 1997). That classification presumes a relationship that relies on the existence of a common ancestor (as established through comparative linguistics). To be a 'Central Tibetan, Kyirong-Kagate' language implies that the relationship is such that the languages present shared retentions that cannot be attributed to chance or contact; to be descended from a common ancestor involves 'regular uninterrupted intergenerational language transmission'. Since 'regular' has not been defined other than circularly, it is hard to assess this formulation, but the essence of the quote is that the community has a shared social history with other languages that it is related to, and the use of that statement is that that social history is reflected in the language. For instance, at a simple level looking at the lexicon, we can see from the data in Table 1 that there is clear evidence for the Tibetan varieties of Lhasa, Kyirong, Tsum and

[1] The Tsum language is not listed as a different language in the national census 2011 (CBS 2012) in Nepal but it has been included in Ethnologue (2012). The Government of Nepal has listed the Tsumpas as 'Siyars' in the list of its nationalities. There have been only some previous works available in the Tsum language, viz. Dhakal and Donohou (2015), Liu (2015), and Donhouu and Dhakal (2016). Some words are listed and compared in Webster (1997).

Sherpa, as well as the Bumthang language of northern Bhutan and the Tamang language of eastern Nepal, form a group, based on the numerals examined, a group that excludes Nepali (and other Indo-Aryan languages). Further, within the Tibeto-Burman group (and based purely on the inspection of the six numerals in Table 1), we would probably judge that Bumthang is more closely related to the Tibetan languages than is Tamang, and that the Tibetan varieties are more closely related to each other than to Bumthang (despite Bumthang being much closer, geographically, to Lhasa than any of the Nepalese languages in the table). Nonetheless, the closely related languages are not identical; while the numerals 1 – 5 are identical in each of Lhasa Tibetan, Kyirong Tibetan, Tsum and Sherpa, we can see variation in the numeral 'seven', in terms of vowel height, roundedness, and voicing of the initial stop. If we accept that the data in Table 1 can be used to argue for a language (subgrouping) relationship, then we must also accept that some degree of variation is acceptable even at a very low level. (Further, we note the replacement of Proto-Tibeto-Burman *g-nis 'seven' with *b-dun as one of the lexical definitions of the Tibetan group, and the unrelatedness of the Bumthang numeral *zɔn* 'two' to the other lexemes in the Table, Tibeto-Burman or otherwise.)

Table 1: A selected comparison of numerals

	One	Two	Three	Four	Five	Seven
Lhasa	tɕik	ɲi	sum	ɕi	ŋa	tyn
Kyirong	tɕik	ɲi	sum	ɕi	ŋa	tœn
Tsum	tɕik	ɲi	sum	ɕi	ŋa	tun
Sherpa	tɕik	ɲi	sum	ɕi	ŋa	din
Bumthang	thek	zɔn	sum	ble	jaŋa	nis
Tamang	kii	ɲii	som	pli	ŋa	ɲis
Nepali	ek	dwi	tin	tsar	pãts	sat

Examining other lexemes, and a different selection of Tibetan languages, similarly reveals phonetic differences between the languages in some words (such as 'fingernail' and 'tooth'), but not in others (such as 'arm' and 'leg'). The phonetic variation in 'skin' is greater than seen in 'fingernail' and 'tooth', but even with this variation we must admit that there are two cognate sets, one involving *pakpa*, *paba*, *pako* and *pabu*, and one involving *gawa* and *kowa*. While the variation between *pakpa* and *paba* is greater than that between *so* and *sa*, it is still plausible variation, while *pakpa* is too different from *kowa* to be explained by phonetically plausible sound changes. Similarly, in the case of 'face' there are three cognate sets, one attested in Lhasa, Kyirong and Yolmo, another only (in this sample) in Tsum, and one more in Nubri. Importantly, there is no obvious explanation for the appearance of the different cognate sets; this, too, is part of the level of accepted variation within even a small group of closely related languages.

Table 2: A selected comparison of the lexicon

	Lhasa	Kyirong	Tsum	Yohlmo	Nubri
'arm'	lakpa	lakpa	lakpa	lakpa	lakpa
'finger-nail'	simu	semu	semo	semu	senmu
'leg'	kaŋba	kaŋba	kaŋba	kaŋba	kaŋba
'face'	toŋba	toŋba	dʒɛ	toŋba	toŋba
'tooth'	sɔ	so	sa	so	so
'skin'	pakpa	kowa	gawa	paba, pako	pabu

In other cases we can identify the source of the variation. Given the variation seen in the food items in Table 3, the terms for 'chilli' in all the languages but Lhasa Tibetan can trivially be assessed as originating in Nepali, in the words *khursani* 'chilli' and *marica* 'pepper'.

Table 3: A comparison of other selected elements in the lexicon

	'chilli'	'potato'	'mango'
Lhasa	sipɛn	ʃɔkɔ	am
Kyirong	k^horsani	he	amgor
Tsum	kortsan	kjɛza	ãp
Yolmo	martsi	he	ãp
Nubri	kortsʌn	doa	ãp

The terms for 'mango' all show either *ãp* (from Nepali) or *am* (from Hindi), again showing evidence of a prior loan relationship. With 'potato', however, we again have lexemes with no obvious external source; he is attested in a number of languages of Gorkha, but without a clear external source. Given the widespread distribution of the lexeme in the region, at least some of the languages must have acquired them through

means other than 'regular uninterrupted intergenerational language transmission'.

Given that we have established that evidence for contact between two languages (including loan words) is also evidence for a prior social relationship between the communities that speak (or spoke) those languages, we have to consider what shared lexical cognates (or, perhaps, plausible cognates) represent, and how we should interpret loan words. Rather than describe the loan words are distractions from the process of understanding the 'true' genealogy of a language, we should examine suspected loan words as evidence of a type of historical social relationship different from that attested in shared cognates that are not loans.

3 A methodology for assessing language contact: phonology

The methodological question we face when we move beyond the lexicon is how to decide which language traits are evidence for which kinds of prior social relationship. While not as simple as it can be when examining lexical items (such as 'chilli' and 'mango' in Table 3), we can develop a procedure for determining suspicious typological traits, and thus plausibly suspecting a prior social relationship other than 'regular uninterrupted intergenerational language transmission'.

Table 4: Typological traits associated with Central Tibetan

Phonology	CVC syllables (simple) tonal contrasts two or three VOT contrasts in stops front rounded vowels?
and not:	implosives, complex clusters, etc.
Morphosyntax	SOV clausal order ergative case marking (related to the genitive?) aspect/evidentiality/egophoricity inflection
and not:	agreement, tense, grammatical gender, etc.

For instance, when considering the Central Tibetan subgroup, we expect to find a number of traits, and extreme deviation from this will be suspicious. The traits listed in Table 4 are discussed below.

Many of the traits that are typical of Central Tibetan are also typical of Tsum. The maximal syllable size in Tsum is CVC, the stops show a three-way VOT contrast in stops (identical to Kyirong Tibetan, its northern neighbour), and the language has a seven vowel system, *i e a o uy ø*, including two front rounded vowels. While, however, Tsum does possess a tone system, it is not one that we would expect of a Central Tibetan language. Further, the tone system of Tsum is not internally coherent. Table5 shows the contrasts found on monosyllables inTsum; there are eight contrastive pitch contours (Liu 2015). We might predict (if tone is assigned to the word) that there would be eight contrasts on disyllables; we might predict (if tone is assigned to the syllable) that there would be 64 (8 x 8) contrastive patterns; alternatively, due to phonological restrictions or sandhi changes reducing contrasts, there might be some number of contrasts across the two syllables which is less than 64 and greater than (or equal to) 8. In fact, words of more than one syllable have exactly two contrasts, LH and HH.

Table 5: Monosyllabic pitch contour contrasts

High	lower High	Mid	Low
			
[55]	[44]	[33]	[21]

Table 5: continues…

Falling	late Falling	Rising	late Rising
			
[51]	[541]	[244]	[224]

The difference between the behaviour of tone in monosyllabic and polysyllabic words is suggestive of some complex series of phonological developments. Returning just to the

fact that there are eight contrasting tones, we can show that this alone can be enough to make us suspect a prior contact event.

We can examine the number of contrasting tones in a sample of 31 (tonal) Tibetan languages, drawn from descriptions of languages in India, China, Nepal and Bhutan. Some languages are described as having only two contrastive tones, some with three or four, and a few with more than that. The average number of tones in these languages is 3.4, with a standard deviation of 1.4. This means that it is not surprising to find between 2.0 and 4.8 tones; up to 6.2 contrastive tones falls within two standard deviations of the mean, and three standard deviations takes the total up to 7.6.

Table 6: Contrastive tones in the tonal Tibetan languages

Languages	No.	Contrastive tones
Brokpa, Denjongkha, Dolpo, Drokpa, LamjungYohlmo, Lhasa Tibetan, Lhomi, Mustang, Sherpa, Syuwa, Yohlmo	11	2
Dongwang, Kyirong_Tibetan, Spiti	3	3
Batang, Dege, Dingri, Dzongkha, Gagatang Tibetan, Gyalsumdo, HaaZongkha, Humla, Jirel, Mugom, Purik, Sherpa_Hile, Shigatse	14	4
Chöcangacakha	1	5
Nubri	1	6
Tsum	1	8
Average		3.4
Standard Deviation		±1.4

The point is that we can evaluate the level of *unusualness* of a tone system, calibrated to this subgroup of Tibeto-Burman; up to four tones is not unusual, and requires no more explanation than the rest of the subgroup. We might say that up to six tones is within the realm of possibility; but having more than 6.2 tones is expected in only 2.2% of the sample; having eight or more tones, more than three standard deviations from the mean, is expected in only 0.1% of the sample. Given that the sample is 31 languages, we expect only 0.031 languages to have this number of contrastive tones, assuming the population we are examining is normally distributed The presence of a language such as Tsum in this sample, that far from the mean, indicates that we are *not* dealing with normally distributed data, and suggests that we should treat the tonal system of Tsum as not representing normal developments from within Tibetan.

Qualitatively, we can also find clues as to where the contact might have come from that led to the complexities on the Tsum tone system. The fact that there are two rising tones in Tsum, and also two falling tones, indicates that tone specification in Tsum is more complex than a simple sequence of H and L melodies. Another language of the region that also has complex tone specification is Kuke, the language of the Kutang region midway between Tsum and (upper) Nubri. Table 7 shows the contrast present in Kuke; a high fall, a mid fall, a low fall, and a rise-fall complex melody.

Table 7: Kuke tone contrasts

	H-fall	M-fall	L-fall	R-fall
Kuke	53	42	31	231

While the Kuke data is suggestive, it is not on its own convincing. In the following section we will examine morphosyntactic data that also indicates a prior relationship with Kuke, as indicated by the linguistic data, which is too far beyond what might be predicted from a comparison with other Central Tibetan languages.

4. Morphosyntax

As with the phonological data just examined, there are some (indeed, many) morphosyntactic traits in Tsum that are completely in keeping with expectations for a Central Tibetan language. The language has an SOV clausal order, the case-marking system is completely as expected from a Central Tibetan language (as are the phonological forms of those case-markers). The verbs, however, show inflection that goes beyond what is predicted from other Central Tibetan languages. Many descriptions of the inflectional system of these languages refers to the (controversial) notion of 'conjunct/disjunct', as illustrated in Table 8. Here the copular verbs that can be used with different persons, in different clause types, are shown (in orthographic form). In declarative sentences there is a contrast between the 'first person' form, *yin*, and the form used elsewhere,

red. Importantly, and this is what distinguishes the Tibetan system as an egophoric system rather than an agreement system, is the use of the erstwhile 'first person' form, *yin*, to mark a second person when the sentence is interrogative. Variations of this system, in which the *yin* form can be termed the 'conjunct' form, and the *red* form the 'disjunct', are found across the Central Tibetan languages, and others.

Table 8: Copulas in Lhasa Tibetan

	Declarative	Interrogative
1st person	*yin*	*red*
2nd person	*red*	*yin*
3rd person	*red*	*red*

The inflectional system of Tsum is very different from this. The sentence in (1) shows an example of the inflection. While the pronouns and their case marking do not represent anything unusual, and while the verb root is widely attested across Tibetan (and Bodish), the suffix *-tsuŋ* is not expected. It is, in fact, part of the paradigm shown in Table 9. The form of the suffix can be predicted if we know the person of the subject, and the person of the object. If the subject is first person, thenthe suffix is either *-pa* or *-po* (depending on tense). (The appearance of *-pa* as a suffix is not surprising, emerging as it does in many Tibeto-Burman languages in a number of functional roles. The *-so* suffix might be related to *-soŋ*, an aspect marker in other Tibetan varieties.) If the object is third person and the subject is not first person, then the suffix will be *-so*. Importantly, if the object is first person, then the suffix will be *-tsuŋ*, a suffix of different shape, and not analyzable synchronically.

(1) *khoŋ-kja-i* *ŋa-la* *thong-tsuŋ*.
3-PL-ERG 1SG-DA see-PST.3⩾1
'They saw me.'

Table 9: Verbal agreement in Tsum

		Object person		
		1	2	3
Subject person	1		*-pa/-po*	*-pa/-po*
	2	*-tsuŋ*		*-so*
	3	*-tsuŋ*	*-so*	*-so*

When we examine the verbal agreement system of Kuke, we notice several points of commonality with Tsum. Firstly, the forms used with first person subjects is similar drawn from elsewhere (*na* is a copular verb in Kuke). Secondly, the arrange of the paradigm, in terms of where the different forms appear. Finally, the forms used for a first person object are phonologically atypical (disyllabic, rather than monosyllabic). We can analyse the *-uŋ* as marking a non-1st person subject, though it quite likely originated in a cliticised form of the first person singular pronoun (*ŋ* in Kuke), since it is clearly regular across the entire verbal paradigm. In Tsum, on the other hand, the *-tsuŋ* suffix is not regularly part of any etymological set.

Table 10: Verbal agreement in Kuke

		Object person		
		1	2	3
Subject person	1		*-na*	*-na*
	2	*-aruŋ*		*-tuŋ*
	3	*-aruŋ*	*-tuŋ*	*-tuŋ*

An explanation of the Tsum inflectional system relies on us observing the inflectional system of Kuke. Kuke is not a Tibetan language; it does share a number of material culture lexemes with Tsum and other Tibetan languages, just as it does share a number of vegetable food lexemes with Nepali; in both cases, these are most likely relatively recent loans because of their semantic domains and because they show sound 'correspondences' by identity with the donor languages. Since the languages do not share the features they do due to shared inheritance, the formal and functional correspondences between Kuke and Tsum, in light of their geographic proximity, there must have been a 'language contact' scenario of some sort at some time in the past of Tsum. Since it is shared with Kuke, we are implying that the ancestors of the speakers of Kuke and Tsum do, to some extent at least, have a shared social history.

5 Interpretation

We have seen that different parts of the grammar of the language tell different stories about the history of the language. If we accept that individual lexical items are subject to easy borrowing, then the overwhelmingly Tibetan nature of the lexicon, as suggested in earlier Tables presented in this article, does not provide

firm evidence for Tsum as having ultimately Tibetan origins. Even without the lexicon, though, there is ample evidence for a Central Tibetan origin of the Tsum language. The nominal morphology, and the segmental phonology, both point to a Central Tibetan source. The Central Tibetan perspective on the position of the different components of the language is summarised in Table 12, which also offers a non-Tibetan perspective (specifically, a perspective that follows from the comparison with Kuke, the nearby language of the Manaslu region).

Table 11: Different data leads to different conclusions

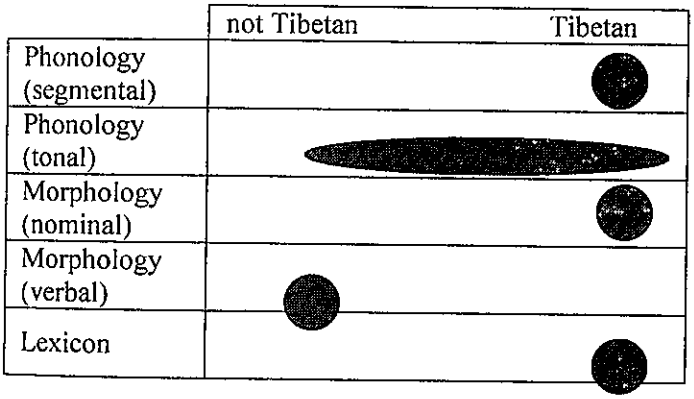

Table 12: Two perspectives on the Tsum data

	...Central Tibetan:	...Manaslu area:
Lexical inventory	conservative	radically relexified
Segmental phonology	conservative	unusual
Tonal phonology	highly atypical	somewhat atypical
Nominal morphology	conservative	highly atypical
Verbal morphology	highly atypical	conservative

Depending on the perspective, different components of data will be thought of as conservative (that is, within the 'normal' range of expected variation) or else heavily contact affected. Our question is whether we can 'weight' certain typological traits against others in terms of borrowing. Having established that lexical items are not reliable witnesses, we have to assess phonological vs. morphological evidence, and we find that this is not helpful; the split in both the phonological and the morphological fields makes this high-level division unhelpful. Examining the two phonological categories, we must assess the tonal phonology as more stable than segmental phonology. Individual segments are relatively easily dispersed as words are borrowed, but prosodic systems are more resistant to borrowing (Donohue 1997, 2012). The phonological evidence, then, leans slightly towards a Manaslu, and not Tibetan, origin of Tsum. The morphological categories split between nominal and verbal; previous work (eg., McConvell and Meakins 2005) suggests that verbal morphological systems are less prone to borrowing than nominal ones. All of this suggests that the original Tsum linguistic system is a 'Manaslu' one, and not a Tibetan one. After its original genesis the language subsequently underwent massive relexification from Tibetan, including the acquisition of the segmental contrasts associated with Tibetan words, and the nominal morphology associated with the use of those (nominal) words. Through this the original tone system was not lost, and the verbal inflection was reformed, but not restructured. It is clear that no one social history unambiguously explains the modern Tsum language, which must be regarded as a mixed language. It follows that no one 'affiliation' should be assumed for the language; one of its parents was a 'Manaslu' area language, and the other, more visible, parent was a Central Tibetan language. Any attempt to arrive at a single, unique genealogical classification does so only at the cost of forcing a simple explanation to a complex question.

References

Bradley, David. 1997. Tibeto-Burman Languages and Classification. In *Papers in Southeast Asian Linguistics No. 14: Tibeto-Burman Languages of the Himalayas*. Canberra, Pacific Linguistics (series A-86), pp. 1-72.

CBS. 2012. *National Population and Housing Census 2011(National report).* Kathmandu, Central Buereau of Statistics.

Dhakal, Dubi Nanda, and Mark Donohou. 2015.Inchoative/Causative Verb Pairs in Tsum. *Nepalese Linguistics.* 30. 45-49.

Donohue, Mark. 1997. Tone in New Guinea. *Linguistic Typology* 1.3: 347-386.

Donohue, Mark. 2012. The shape and spread of tone. In Cathryn Donohue, Shunichi Ishihara and William Steed, eds, *Quantitative approaches to problems in linguistics: studies in honour of Phil Rose*: 9-20. Muenchen: Lincom Europa.

Donohue, Mark. 2013. Who inherits what, when? contact, substrates and superimposition zones. In Balthasar Bickel, David Peterson, Lenore Grenoble and Alan Timberlake, eds., *Language Typology and Historical Contingency; in honor of Johanna Nichols*: 219-239. Typological Studies in Language 104. Amsterdam: John Benjamins.

Donohue, Mark and Dubi Nanda Dhakal. 2016. Muenchen: Lincom Europa.

Eppele, John W., M. Paul Lewis, Dan Raj Regmi, and Yogendra Prasad Yadava. 2012. *Ethnologue: Languages of Nepal*. Kathmandu: Linguistic Survey of Nepal and SIL International.

Huber, Brigitte. 2005. *The Tibetan dialect of Lende (Kyirong): a grammatical description with historical annotations.* VGH Wissenschazsverlag/International Institute for Tibetan and Buddhist Studies GmbH.

Liu, Naijing. 2015. Tsum tone: a challenge for typology and phonological description. MA sub thesis, The Australian National University.

Matisoff, James. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction.* Berkeley: University of California Press.

McConvell, Patrick and Felicity Meakins. 2005. Gurindji Kriol: A Mixed Language Emerges from Code-Switching. Australian Journal of Linguistics 25.1:9-30.

Thurgood, Graham, and Randy J. LaPolla. 2003. *The Sino-Tibetan Languages.* London: Routledge.

Webster, Jeff. 1992. A Sociolinguistic Survey of the Sino-Tibetan Dialects of Northern Gorkha, ms.

Dubi Nanda Dhakal and Bhojraj Gautam:
Tribhuvan University

Mark Donohue and Naijing Liu: The Australian National University

CONTENTS

Nepalese Linguistics

Volume 31, November 2016

Linguistic Society of Nepal
भाषाविज्ञान समाज नेपाल