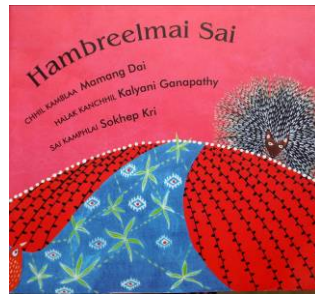# Proposal for a writing system for Kman [=Miju], a language of Arunachal Pradesh

Roger Blench
Kay Williamson Educational Foundation
8, Guest Road
Cambridge CB1 2AL
United Kingdom
India mobile 00
Voice/ Ans (00-44)-(0)1223-560687
Mobile worldwide (00-44)-(0) 7847-495590
E-mail rogerblench@yahoo.co.uk
http://www.rogerblench.info/RBOP.htm

This printout: Tezu, March 18, 2015

TABLE OF CONTENTS

TABLES

Please note

Writing systems are often known as 'scripts' in India. However, 'scripts' refer to systems of characters such as Devanagri, Chinese, Arabic or Cyrillic. Writing systems or orthographies are a series of conventions within a script, in this case Roman script. The term 'script' will not be used in this document.

The first part of the paper uses the special characters of the International Phonetic Alphabet (IPA). These allow linguists to write down the sounds of the world's languages. Learning to read your language *does not require* you to learn IPA characters, just to able to pronounce the examples. Through this you can read the words in Idu using the conventions agreed.

MAJOR PRINCIPLES OF WRITING SYSTEMS

The objective is to write down an unwritten language in a way that can easily typed and read and is also easily taught to learners

The procedure is as follows;

- The language is analysed according to scientific linguistic principles
- The principles are set out in an orthography document [this document]
- The orthography document makes suggestions for ways of writing the sounds of the language
- It also sets out problems in writing and choices to be made

The community is responsible for;

- Forming a language development committee which is representative of speakers of the language
- Creating a 'paper trail', that is a document recording the decisions of the committee and the reasons for those decisions
- Discussing the options for a writing system and making choices, giving reasons for those choices

Then;

- The language development committee then forwards its decisions back to the author of the orthography document and the document is updated with this information
- The language development committee receives this revised document, reads and approves it

Next steps

- A trial primer must be created, a booklet which teaches new readers how to write their language
- This should *not* be printed until it has been tested
- When it has been tested, some problems will surely be found. The primer and possibly the orthography will need to be updated
- A final version of the primer can be printed and tested on a larger scale
- Preliminary books can be printed
- At the same time a first version of the dictionary can be circulated, to help new readers write their language

REMEMBER

Kman is not like any other language in the world, and especially unlike Hindi, Nepali or Assamese. It has quite different sounds and quite different grammar. Therefore there is no reason to write it in the same way. Kman must have its own writing system which reflects the language.

1. Introduction: the Kman and the term 'Mishmi'

The term 'Mishmi' is used in the travel literature as far back as the early nineteenth century to refer to three distinct peoples, the Kman, Tawra and the Idu. These three peoples live in northern Arunachal Pradesh, northeast India, around the towns of Roing and Tezu, and it villages as far as Tibet and Myanmar. Culturally speaking, the Tawra have been historically grouped with the Kman, as the Mishmi, and not with Idu. However, Idu and Tawra are undoubtedly related, but Kman is not part of the same linguistic group. The Kman people are commonly known as Miju, but this is incorrect. The spelling 'Kaman' is also used.

The Kman language has been little described. The first record of Kman appears to be Robinson (1856) which is quite accurate for the period, and his transcriptions are still recognisable today. Needham (1886) is the first specific vocabulary of Kman (under Miju) and the first to compare it with Tawra. The only modern publications on the language from the Indian side are Das Gupta (1977) and Boro (1978). These were said to be 'practical' guides and the transcription of Kman is highly inaccurate by modern standards. Although there are few speakers on the Chinese side of the border, there have been several publications on 'Geman', as the language is known. These include Sun (1983, 1991, 1999) and most importantly, Li (2003) which is a full-length description of the language. As these are wholly written in Chinese, they are not very useful to speakers in India.

Kman has undergone an intriguing development in terms of its orthography; a local system of writing used for communication on Facebook has developed which is also used in a children's book (Dai et al. 2013). While functional, this system is internally highly inconsistent and preparing teaching materials based on it would be difficult. A linguistically more well-founded publication is Krisikro (2006) which covers both Kman and Tawra. Although printed, it has never been used. Separately, a lexical guide has been published covering both Kman and the neighbouring Tawra [Taraon] language (Tawsik 2014).

This working paper[1] is prepared for consideration by the Kman community. It first describes Kman using the system of the International Phonetic Alphabet (IPA) and then discusses how the sounds of Kman can be written in a practical orthography.

2. Phonology

2.1 Consonants

Kman consonants are shown in Table 1;

Table 1. Kman consonants

	Bilabial	Labio-dental	Alveolar	Retro-flex	Palato-alveolar	Palatal	Velar	Glottal
Plosive	p, pʰ b		t, tʰ d				k, kʰ g	ʔ
Fricative		[f]	s z		ʃ			h
Affricate			ts dz		ʧ, ʧʰ ʤ			
Nasal	m		n			ɲ	ŋ	
Flapped			r					
Lateral			l	ɭ				
Approximants	w	ʋ		ɽ		y		

Unusual are the retroflexes [ɽ] and [ɭ], which are contrastive but await further phonetic investigation. /ɭ/ is rare, although heard in;

---

bicycle	ɭyárl
frying pan	ɭyaŋa

it may possibly be a free variant of /l/.

/s/ and /ʃ/ appear to be in free variation in many words, and are probably not distinct phonemes. /ʧ/ and /ʤ/ are written 'c' and 'j' throughout this document. c and cʰ are sometimes in free variation with tɕ and tɕʰ.

/pʰ/ is in free variation with /f/, which may be influence from speaking Indo-Aryan languages. Boro (1978) includes an aspirated series of voiceless stops, i.e. bʰ, gʰ etc. but these are spurious.

/z/ is in alternation with /ʤ/, but some speakers prefer [z] for particular words, so it may well be acceptable to write it as part of the alphabet, but it is unlikely to be a distinct phoneme. The affricates /ts/ and /dz/ are clearly sounded in many words including /tsɔ́ŋ/ 'person', but locally this is written *cong*, which loses the contrast with /ʧ/.

The labio-dental approximant is an unusual phoneme for this area, but occurs in;

blood	ʋì, áʋì
snake nest interior	ʋì

Kman has occasional long consonants in the centre of words;

ear infection	táppám
fish	áŋŋà

A key feature of Kman are the many consecutive sequences of consonants without vowels. Some important examples are;

Table 2. Kman CC and CCC sequences

Gloss	Kman
ankle	tln
bamboo sp.	máy mbrt
cough, productive	tt
cup	wadn
grass	tfn
green	krm
partridge	mnm
snake	ʈl
snake nest entrance	rbn
storm	t.mk òwáŋ
wild buffalo	tl.wì

This is very counter-intuitive to anyone raised on the traditions of writing European languages; we expect vowels to follow consonants. So attempts to write the language usually insert extra vowels. For example, the name of the language is often locally spelt 'Kaman' and in the Chinese source 'Geman'. However, speakers do not pronounce a vowel and it is not necessary to write it.

2.2 Vowels

Kman probably has nine vowels, shown in Table 3;

Table 3. Kman vowels

Vowels	Front	Central	Back
Close	i	ɨ	u
Close-Mid			o
Open-Mid	ɛ	ə ʌ	ɔ
Open		a	

It is uncertain that the two central vowels /ɨ/ and /ə/ are distinct.

Nasalised vowels and long vowels are marked in the two existing sources Das Gupta (1977) and Boro (1978). However, in striking contrast to Tawra and Idu, these are in fact absent. Phonetic nasalisation of final vowels is often heard due to a prior nasal, and sometimes final -ŋ is in free variation with a nasalised vowel. As in Idu, there are occasional sequences of two similar vowels, often with different tones, but these do not amount to a systematic contrast in length.

Diphthongs

Kman has occasional sequences of two dissimilar vowels. These usually have contrastive tones and are pronounced separately. Examples are;

    cut with knife   súì
    buffalo, wild   tlúì

2.3 Tones

Like most regional languages, Kman is tonal, but the tone-system is relatively simple. There are two level tones, High (´) and Low (`), a rising (ˇ) and a rare falling (ˆ) tone. Glide tones also arise from VV sequences. For example;

    cut with knife   súì
    buffalo, wild   tlúì

Examples of three-way contrast are quite rare on single syllable words. One example is;

    High   fruit   sí
    Low   fungus   pásì
    Rising   rat, general   sǐ

Table 4 shows examples of both rising and falling tones in Kman. As you can see, they are all associated with either modified consonants (i.e. C + w,y) or syllables with final -y-, -w.

Table 4. Kman glide tones

Gloss	Kman
money	ápây
cousin's father	kâw
valley	gyûŋ
wind	băwŋ
jaw	găy
navel	ŏw
kidney	gădn

One problem is whether words which consist entirely of consonants (Table 2) can be considered to have tone. The answer appears to be 'sometimes'. For some words, a floating tone is clearly heard between the

consonant and the one that follows it. Elsewhere, the tones seem to be have been neutralised. This could be treated as a sequence of low tones, but it seems likely that the loss of vowels ends by bleaching tonal contrast.

3. Writing system proposal

3.1 Consonants

Table 5 shows the main consonants of Kman with suggestions as to the most usual way of writing them. The difficult consonants, where a writing solution is to be discussed, are marked with a shaded line.

Table 5. Kman consonants

IPA	as in	Gloss	Written	as in
p	pàŋ	first	p	pang
pʰ	pʰùn	door	ph	phun
b	bici	chili	b	bici
t	táthì	saliva	t	tathi
tʰ	tʰànàt	resting place	th	thanat
d	dʌ́k	belly	d	duhk
ts	tsɔŋ	person	ts	tsong
dz	dzàwpà	tree sp/	dz	dzawpa
k	ki	I, me	k	ki
kʰ	hàŋ kʰìm	toad	kh	hang khihm
g	gíl	cucumber	g	gil
h	hamak	tasty	h	hamak
s	sá	child	s	sa
ʃ	shabre	sword	sh	shabre
ʧ	ʧe	to split	c	ce
ʧʰ	ʧʰay	sincere	ch	chay
ʤ	ʤal	comb of cock	j	jal
z	zk	nail	z	zt
m	mál	floor	m	mal
n	nâw	mother	n	naw
ɲ	ɲù	tongs	ny	nyu
ŋ	ŋit	to know	ng	ngit
r	rɔ́k	hand	r	rohk
ɽ	tɽù	brother's wife	rh	trhu
l	lǎp	leaf	l	lap
ɭ	ɭyárl	bicycle	lh	lhyarl
w	wadn	cup	w	wadn
ʋ	ʋì	blood	v	vi
y	yùn	clan names	y	yun
ʔ	mùʔùl	to burp	'	mu'ul, mu-ul

A particular issue in Arunachal Pradesh writing systems is how to write final -y and -w. It is very common for these to be written -i and -o. So;

gǎy   'jaw'   is written   gai
naw   'mother'   is written   nao

Using -i and -o gives the impression of a double vowel, which is not the case. So the use of -w and -y is recommended.

Kman also has rare long consonants, typically in the middles of words, which are written as double consonants. For example;

fish   áŋŋà

If we were to write this, it would come out as 'angga'.

Kman has numerous modified consonants or consonant clusters where an initial such as k-, kh, g-, t-, th- etc, are followed by laterals (l), rhotics (r), bilabials (w) and palatals (y). Examples are;

Examples of these are given below;

l	back	glɔ̀k
	leg	plá
	tongue	bláy
	burn n.	máy phlɛ̀
	sprain, dislocation	tákhlɔ́t
r	backbone, spine	brŭgn
	chest	gròŋ
	fish nest	krùl
	match	máy khrùt
	tree sp. I	cíprùŋ
y	hair	shyàm
	nerve	gyâw, gyân
	field hut	kyàp
	frying pan	ḷyaŋa
	clay	nyápén
	eclipse	tá phyɔ̀p
	wild cat	k.tyùŋ
	copper	khyɔp
w	mole	mwínyɔ̀
	paralysis	hámbwèy
	flea	cáŋgwɔ̀n

A problem for writing Kman is that the spoken language has many words beginning with k-, g- etc. which are distinct syllables and do not form a cluster with the following consonant. Where the next consonant is something which does not combine with k- or g-, the pronunciation is fairly obvious. The name of the language, Kman, is a good example. This looks unusual, and sometimes it is written with a following vowel, for example in the language name, often written 'Kaman'. In reality, there is no vowel, so it is just as effective and shorter to write 'Kman'. Some other examples are;

kshik	child
ksit	good
kma ne	although
ktam	six
nàw ktì	mother's younger sister
kgìn	two
knìn	two

The problem comes when writing words where k- , g- etc. are followed -y, -w, -l and -r, since they can either combine with them to form a cluster, or be pronounced separately. If we do not mark this in writing, then confusion can result. Linguists write a full stop after the consonant which is pronounced separately, in order to mark this division. Some examples of this type of contrast are;

5

k.l versus kl

    k.lìn   as opposed to   klòŋ   klii taŋso
    five   insect   sweeper

k.r versus kr

    k.ràу   but   kròŋ   krùl
    block for cutting meat   classifier for longhouses   fish nest

k.y versus ky

    k´.yìn   but   kyàp
    two   field hut

k.w versus kw

    k.wí   pigeon   but   kwì   wax

A related problem is where t + h sequences can be confused with aspirated t, i.e. tʰ. So we have the word;

fungal infection   t.hàt

but;

place where you stop on a journey   thànàt

There are many of these unusual breaks in pronunciation. For example;

    wild buffalo   tl.wì

Where the 'tl' forms a cluster, but the 'w' following does not combine with them but is pronounced distinctly.

Writing a full stop between two letters to mark the different syllables, may lead to confusion with other punctuation. A more practical solution might be a dash. So;

    k-lin   but   klong
    five   insect

and

    k-ray   but   krong
    block for cutting meat   classifier for longhouses

Exactly how difficult this is to read and which words will need to be marked in this way must be tested on readers.

Kman has some unusual changes in context which require decisions with the rules of the writing system. One of the principles of orthography is the 'consistent image', i.e. the notion that a word should look the same wherever it occurs, otherwise the reader will think two or more distinct words are meant. Take the example of the word for 'stone'. If it is cited in isolation it is pronounced;

lʌ́gn

If it is given as part of a compound, it is pronounced;

boulder   lʌ́ŋ móy

If it is followed by a classifier, it is pronounced \lòwŋ\, as in;

lòwŋ   bràt   k.sàm
stone   CLF   three

These are all the same word, and these are all allomorphs of the same underlying word. We know on comparative grounds that *lʌ́ŋ* is likely to be the original form, and the other two later transformations. For ease of reading, it is almost certainly best to keep the original form, which would be *luhng* in the orthography, and let the reader make adjustments. There are a number of other words showing these unusual consonant reversals, where /n/ plays a role. For example;

kidney   gădn
cup   wadn
insect general   klògn   klòŋ in compounds

The principle of a consistent image should be kept under review.

3.2 Vowels

Five of the Kman vowels correspond to English vowels as in Table 6;

Table 6. Kman vowels

IPA	Example	Gloss	Written
a	áŋŋà	fish	a
e	kɛpmɔ	ten	e
i	iŋ	ear	i
o	oway	husband	o
u	ùí	to pick up	u

and can be written like that. The 'e' sound is usually 'ɛ', i.e. the sound found in English 'net', 'head' and so on, but it is not distinct from /e/ in Kman.

However, Kman has four vowels which do not have a way to be written in the English alphabet. These are /ɔ/, /ɨ/, /ə/ and /ʌ/.

The vowel /ɔ/ sounds like the 'o' in English 'not', 'shot' and 'pot', which is diffeent from the sound in 'boat', 'float' and 'throat'. Some common words in Kman with this vowel include;

IPA	Gloss
glɔ̀k	back
rɔ́k	hand
áphɔ̀	shoulder
sɔ́	liver

This sound will need to be distinguished from 'o'. Some languages in Africa just keep the IPA symbol ɔ but this is unlikely to be popular in India. It depends on how receptive the audience is to unfamiliar characters. Some earlier works on Kman proposed 'ó' and 'ö'. The first one should not be used as it can be confused with a tone-mark, but 'ö' (called an umlaut and used in German) is possible. Other possibilities are following the 'o' with a 'h' or using an underline. This set of words would then be written as follows;

7

IPA	Gloss	with -h	with umlaut	with underline
glɔk	back	glohk	glök	gl<u>o</u>k
rɔ́k	hand	rohk	rök	r<u>o</u>k
áphɔ̀	shoulder	aphoh	aphö	aph<u>o</u>
sɔ́	liver	soh	sö	s<u>o</u>

The vowel /ʌ/ sounds like the vowel in English 'cut', 'nut' etc. Common words in Kman with this vowel include;

IPA	Gloss
dʌk	belly
dʌp	nipple
hámbʌl	mat
mʌnyɔ̀ŋ	elephant

Possible writing solutions include;

Gloss	IPA	-h	umlaut	subdot	underline
belly	dʌk	duhk	dük	du̞k	d<u>u</u>k
nipple	dʌp	duhp	düp	du̞p	d<u>u</u>p
mat	hámbʌl	hambuhl	hambül	hambu̞l	hamb<u>u</u>l
elephant	mʌnyɔ̀ŋ	muhnyohng	münyöng	mu̞nyo̞ng	m<u>u</u>ny<u>o</u>ng

Apart from these, Kman has two central vowels, similar to 'a' pronounced in the centre of the mouth. One of these is known as schwa, /ə/, such as you hear in English 'bird'. Some common Kman words with schwa include;

    mud   də̀rí
    silver   ə̀páy
    again   bəgebo

Writing a schwa is more difficult. Some options that have been used in other languages are;

                ə   ə, ah, a̞, <u>a</u>

These words would look as follows;

    mud   dəri   dahri   da̞ri   d<u>a</u>ri
    silver   əpay   ahpay   a̞pay   <u>a</u>pay
    again   bəgebo   bahgebo   ba̞gebo   b<u>a</u>gebo

Some Arunachal languages have chosen to write a completely different consonant for 'ə', in the case of Galo 'v'. Since this sound is not used in Galo this is possible, although it gives the word an unfamiliar look. However, this would not be possible in Kman, since it already has a bilabial /ʋ/, written 'v'.

Another vowel similar to schwa is /ɨ/ which is pronounced higher in the mouth. The table shows some common words with this vowel, as well as some possible writing solutions.

lɨm	heart	lihm	l<u>i</u>m
brɨl	classifier for pieces of wood	brihl	br<u>i</u>l
sɨn	meat	sihn	s<u>i</u>n

However, it is perfectly possible to continue to write ɨ, as this is used in other Arunachal Pradesh languages. Galo uses 'w' but again the importance of 'w' with its normal sound in Kman excludes this possibility.

3.3 Tone

Is it necessary to write tone? Tone is marked throughout this working paper, but there are relatively few minimal pairs which can be confused by not marking the tone. At the moment, tone can be left unmarked in the writing system, but this decision should be kept under review.

4. Conclusion

This working paper is a preliminary attempt to help create a writing system for the Kman language. It gives examples of words as they are in phonetic [IPA] transcription and how they could be written in a practical system for Kman. A problem is that some words, including the name of the language now look very different from the spellings normally used. But for a writing system to be developed and taught ut must be internally consistent. Some issues are still very much open to discussion by the Language Development Committee, and when decisions are taken on these issues, the writing system document can be revised and used as a basis for written publications in the Kman language.

References

Boro, A. 1978. *Miju dictionary*. Shillong: Research Secretariat.
Campbell, G. 1874. *Specimens of Languages of India: Including Those of the Aboriginal Tribes of Bengal, the Central Provinces, and the Eastern Frontier*. Calcutta: Printed at the Bengal secretariat Press.
Das Guta, K. 1977. *A phrase book in Miju*. Shillong: Research Secretariat.
Konow, Sten 1902. Note on the languages spoken between the Assam Valley and Tibet. *The Journal of the Royal Asiatic Society of Great Britain and Ireland, 1902*: 127-137.
Krisikro, Sutik 2006. *Tawra: tako zo bra/ käman, pït sai brat*. Tezu: Graphic Systems.
Li, Daqin 2003. *A study of Geman*. [In Chinese]. Beijing: The Nationalities Press.
Macgregor, C.R. 1887. Journal of the Expedition under Col. Woodthorpe, R.E., from Upper Assam to the Irrawadi and return over the Patkoi Range. *Proc. Royal Geographical Society*, 9: 19-42.
Mills, J.P. 1952. The Mishmis of the Lohit Valley, Assam. *The Journal of the Royal Anthropological Institute of Great Britain and Irelan*d, 82(1):1-12.
Needham, Jack F. 1886. *A few Dîgârô (Târoan), (Mîjû) M'jû, and Thibetian Words*. Shillong.
Robinson, W. 1856. Notes on the languages spoken by the Mi-Shmis. *Journal of the Asiatic Society of Bengal*, 24: 307-324.
Sun, Hongkai 1991. *Zang Mianyu yuyin he cihui*. [In Chinese]. Beijing: Zhōngguó Shèhuì Kēxué Chūbǎnshè.
Sun, Hongkai 1999. On the Himalayan languages of the eastern Himalayan area in China. *Linguistics of the Tibeto-Burman Area*, 22: 61-72.
Tawsik, Somodang 2014. *Kaman-Tawrah*. Tinsukia: Chitra printers.