Number building in Tibeto-Burman languages

Martine Mazaudon

Number building in Tibeto-Burman languages

Martine Mazaudon, Lacito-CNRS, France

Summary

The vast majority of the three hundred or so modern Tibeto-Burman languages have decimal numeral systems, like most contemporary languages, and more specifically like their large influential neighbours, the Sinitic and Indo-Aryan languages. Like most languages again, their favorite way of constructing higher numbers is by multiplication and addition.

On inquiry some of them reveal less common bases, 20, 12, 5 and even 4 as the building blocks of their number systems. These are used in parallel with a decimal system, mixed into it, or as the unique system of the language. Some are only traces and some are full-fledged systems, like Dzongkha, which can compute all numbers up to 160.000 (20^4) including all intermediate numbers in its vigesimal system. Only the Maya language of Central America has been described with a comparable complexity.

Principles of number-building other than addition and multiplication, like fractions inside a number, or overcounting, which have become rare in other parts of the world, are also attested, although not always recognized by descriptors.

A complete typology of number systems has much to learn from Tibeto-Burman languages. More field-research is urgently needed to collect these fast disappearing systems, which foreign educators and local speakers alike unfortunately regard as a hindrance to socio-economic development.

1. On the endangerment of number systems

Many Tibeto-Burman languages are, as we know, endangered. Some are more imminently threatened with extinction, due to the very small number of their speakers, or to socio-economic conditions which discourage their speakers from continuing to use them. Some language communities fight actively for the preservation of their language, and languages which have the support of a large community of speakers, and which have national or quasi-national status and a vast literature, like Burmese, Tibetan or Dzongkha, do not seem likely to disappear in the near future.

Nevertheless, when languages are spoken in contact with other languages more prestigious or more widespread than they are, some parts of their structure are threatened faster than other parts. Technical, administrative or religious words can easily be borrowed or copied along with the techniques, or concepts they refer to. In areas like the Tibeto-Burman area, where multilingualism has been in effect for thousands of years, the semantic fields of words in different languages may come to be exactly the same, and the grammatical structure of sentences become so similar that only the words themselves seem to remain different. This is well known and needs no elaboration.

Numbers, on the other hand, have sometimes been considered as part of core vocabulary, that part of the vocabulary which remains the most stable over time, and indeed in Indo-European they are a very reliable source of examples for historical comparative reconstruction. But this stability is not found everywhere. The Thai numbers, for instance, once thought to be derived from the same ancestral forms as the Chinese numbers, and so used to argue for the genetic relationship of these two large language groups, have been shown to have been borrowed from Chinese into Thai at an early date (Benedict, 1942).

Even when number words are not borrowed, number structure can be. This can happen naturally, by contact in the market place, where prices have to be debated. It can also, at least since the last century, be the result of a voluntary transformation of the system, with a view to "modernization". Examples of willful changes of the syntax of a language are much rarer.

Number systems, on the other hand, can be considered as a part of a language vocabulary, but they are also a sort of intellectual technical tool, used in everyday life and in higher mathematics. This could be the reason why they are especially susceptible to influence from outside languages. Whatever the case may be, among Tibeto-Burman languages, we observe a very interesting diversity of structure, which was gradually receding in previous centuries and is disappearing dramatically fast nowadays. Some structures are already only traces, and for several languages, it is only in descriptions written in the nineteenth and early twentieth century that we find evidence for what were once complete functional systems, of a structure different from the dominant decimal numeration.

2. What is a number system

The number system is an integral part of a language and should be described like all other parts of the lexicon from the point of view of its syntactic construction and its internal morphology. It is also, from a different perspective, a technical tool, built along regular principles to help man measure his environment[1]. The number system is taught to children in a more directed and systematic manner than e.g. the names of animals. This may be part of the reason why the simple arithmetic principles which underlie the building of numbers as a conceptual tool are most often transparently apparent in their names. I say "most often" because it is not always so. In Nepali for instance, number names up to one hundred have evolved in their pronunciation to such extent that the old structure cannot be seen without properly scientific reconstruction, for instance *baunna* 'fifty-two' is not immediately understandable when you know the words for 'two' *dui* and for 'fifty' *pacas*. In most Tibeto-Burman languages this is not the case and the original structure is readily visible.

2.1 The base of a number system

The base of a number system can be envisioned as the main building block of the system. It is the first number, reached in counting, which is used to build higher numbers by being itself counted in a regular manner. So for instance in a perfect decimal system, whose building block is the number '10', we would have simple non-compound names for all numbers below ten, and compounded names above. The powers of the base (10^2, 10^3, 10^4, etc.), called bases of higher order, or of higher rank, alone should receive new names, not compounded from the names of lower numbers. Thus, if we write simple names in italics and compounds in roman, for a perfect decimal system we should expect the following:

one, two, three, four, five, six, seven, eight, nine, ten (10^1)
ten-one, ten-two, ten-three, ten-four ….two-ten
two-ten-one, two-ten-two …three-ten
three-ten, four-ten … nine-ten, *hundred* (10^2)
one-hundred, two-hundred..., nine-hundred, *thousand* (10^3)
one-thousand, two-thousand ..., nine-thousand, *ten-thousand* (Chinese *wan*) (10^4)
one-(ten-thousand), two-(ten-thousand) …, *hundred-thousand* (I.A. *lakh*) (10^5)
one-(hundred-thousand) …, *million* (10^6)

Such a perfectly transparent system is found in Chinese, with simple names for bases up to 10.000, and all intermediate numbers transparently expressed, with no morphological change, by the operations of multiplication and addition. — In most languages, as in Nepali or in English, the passing of time has obscured the etymology of the intermediate number names, and one or the other of the powers of the base is missing (10^4 and 10^5 in English).

---

[1] This conceptual aspect as it is reflected in the language is what will interest us here. For a thorough study of the etymology and comparative reconstruction of the roots the reader is refered to Matisoff (1997).

Similarly, a 5-based, or quinary, system would ideally have simple names for numbers from 1 to 5, then for ‘25’ (5x5), and ‘125’ (5x5x5). The names for ‘10’, ‘20’, and ‘100’ should be compounded.

In the vigesimal system of Dzongkha, the national language of Bhutan, we find simple names[2] for the powers of the base up to the fourth power of the base (160.000).

Table 1. Names for the base 20 and its powers in the Dzongkha vigesimal system

1 khe		= 20	(20^1)
1 ɲiɕu	= 20 khe	= 400	(20^2)
1 kheche	= 20 ɲiɕu	= 8 000	(20^3)
1 jãːche	= 20 kheche	= 160 000	(20^4)

2.2 Auxiliary bases

When the main base of a number system is high, like ‘20’ and even more so ‘60’, auxiliary bases are used for the construction of numbers smaller than the base. Vigesimal systems very often have an intermediate base ‘10’, used to build numbers between 11 and 19. The vigesimal system of Aztec, the ancient language of Mexico, had two auxiliary bases ‘5’ and ‘10’ (Menninger, 1969: 62-63). Sumerian, which had sixty as its main base, used 6 and 10 as auxiliary bases (Menninger, 1969: 163sqq; Ifrah 1981: 56 sqq). Although the Tibetan and the Chinese calendars count years in five cycles of twelve years, which makes sixty years, we have not found any sexagesimal system in Tibeto-Burman. The other bases we have mentioned are all found in one or the other of the Tibeto-Burman languages.

These auxiliary bases are used only in the intervals between the multiples of the main base; they give way to the multiples of the main base whenever these are reached. For instance in a quinary system, you would say ‘3 fives’ for ’15, ‘4 fives’ for twenty, but in a quinary-vigesimal system, although you say ‘3 fives’ for ‘15’, and although the language possesses the material to construct ‘4 fives’, it is not done and the simple word ‘twenty’ is used. Multiples of auxiliary bases are not found in parallel with those of the main base.

2.3 The origins of the bases and of their names

Counting, measuring, and grouping objects are similar and distinct activities[3]. The base of a system is the first level where a grouping is reached in the enumeration of numbers. This can be heard in the intonation of a person counting: the reaching of the base or of a multiple of the base is marked by a strong stress and a pause; clearly the speaker has reached a resting place, a place where he can grasp all the objects counted until then in one secure bundle.

The word ‘bundle’ is precisely the source of the Dzongkha name for ‘twenty’, the base of its system. The word is *khal* in classical Tibetan, pronounced regularly as /khe/ in Dzongkha. Its original meaning was

“1. burden, load […]; 2. bushel, a dry measure = 20 *bre*, therefore = a score, or 20 things of the same kind” (Jäschke 1881).

Such semantic shift from a measure word to a number word is common, and reflects the cognitive position of the round numbers as means to envision globally a group of objects as opposed to the intermediate numbers.

[2] As we will see in a moment, simple here is meant from the point of view of synchronic mathematical computation. The names of bases can be (old) compounds grammatically, as /kheche/, which is etymologically ‘a large twenty’, or even more strikingly /ɲiɕu/, which used to be 2 times 10, hence ‘twenty’, and was borrowed into the vigesimal system with the meaning ‘400’, clearly not computable from its etymological (not synchronic) components.
[3] For more details see Mazaudon (2002).

In the names of the Dzongkha powers of the base, we see more evidence of their original value as groupings; they are qualified phrases similar to the qualification of any ordinary object. 8000, /khe-che/ contains the word /che/ 'large' thus it is 'a large twenty'. In the same way the Old Italian *milione*, borrowed into French and English as *million* (10^6) is derived from the Latin word *mille*, one thousand, with the augmentative suffix *–one*; a million is an "inflated thousand". We will come back to the cognitive importance of the round numbers as wholes when we analyze the rare operation of overcounting used in several Tibeto-Burman numeral systems.

2.4 The construction of intermediate numbers

Once we have names for the main base and its powers, ideally unanalyzable in synchrony, the intermediate numbers are built by the operations of elementary arithmetic.

Multiplication, addition and subtraction

Multiplication and addition are found in all numeral systems. They both build the number to be expressed in reference to the closest lower step in the gradation. Thus in English 'thirty-two' starts from 30 and adds 2 to it, 'three hundred' starts from 100 and multiplies it by 3.

Subtraction, also called back-counting, expresses a number by reference to the next higher step in the gradation , and states how many lower level units are missing to reach that step: Latin *duo-de-vigenti* (2 from 20) 'eighteen', refers to 20 and takes away 2, *un-de-vigenti* (1 from 20) 'nineteen', refers to 20 and takes away 1.

Overcounting

Overcounting is a rarer operation. It has been reported from two areas of the globe: the Germanic North of Europe and Central America (Menninger, 1969: 76-80). Tibeto-Burman can now be added. We will illustrate its workings in Dzongkha with a drawing (figure 1 in section 3.2).

The principle is to express a number as a point on the way to a multiple of the base[4]: for example, 'six on the way to eighty'. What is surprising to those who are not used to such systems is that the starting point for counting the 6 units is not mentioned. In fact, the units are counted from the lower bound of an interval whose extent is equal to the value of the base. The upper bound of this interval (80 in our example) is specified; it is a multiple of the base. The lower bound, the starting point, is simply the next lower multiple of the base, i.e. the upper bound less the value of the base. Thus, in a decimal system, 'six on the way to eighty' means six units into the interval of ten whose upper bound is 80; the lower bound of this interval is 70, and we add 6 units to arrive at 76. In a vigesimal system the same expression would presuppose an interval of 20 and a lower bound of sixty, and have the value 66. See below in decimal Ao Naga (section 4.2):

Ao: rokyr̥ maben trok (60 not-reached 6) 'fifty-six'

In vigesimal Maya, all numbers over forty are formed by overcounting:

hun-kal	(1.20)	'20'	hun-tu-kal	(1-on-20)	'21' (addition)
ca-kal	(2.20)	'40'	hun-tu-y-oxkal	(1-on-towards-60)	'41' (overcounting)

The "half-count", a fractional overcounting

Using a fraction to express a quantity smaller than one unit is common, even if many languages do not have dedicated terms for fractions such as half, quarter, third etc. but simply say '1 part out of 2', or some similar expression. When an amount larger than a unit has to be completed by a fraction, this is most commonly done by addition, as in English 'three and a half'.

[4] Or a multiple of a power of the base. For clarity of exposition I leave aside this case for the moment.

In a number of languages fractional numbers larger than the unit are expressed with reference to the upper limit of an interval instead of the lower limit, as in overcounting. Thus corresponding to English ‘one-and-a-half’, Russian says *pol-tora* <*pol-vtora* ‘half of the second’. Latin *sestercius*, the name of a coin worth 2½ *as*, comes from *semis-tertius* ‘half of the third’.

It is very rare for fractions to be used to express not quantities smaller than the smallest unit, but fractions of an amount corresponding to a base in a numeral system. For instance in French you could say, if you want to buy fifty nails: “Donnez-moi un demi-cent de clous”, lit. ‘Give me a half-hundred of nails’, because nails are sold in groups or packages of 100 units, and so the word *cent* in this sentence is used as a measure word. It is impossible to say “Donnez-moi trois cent et un demi-cent de clous” (350), or **“**mille et un demi-cent de clous” (1050).

Some languages use the same principle regularly in building numbers higher than the value of the base. When this occurs it is with subtractive or, mostly, overcounting structures. Classical Danish, which had a vigesimal system, expressed ‘fifty’ as *halv-tred-sinds-tyve* (modern *halvtreds*) ‘half-third-time-twenty’, which can be computed as ‘half of the interval which has the value of the base (i.e. 1/2 of 20=10)’ ‘on the way to the third twenty (i.e. 60)’, and we now know that in overcounting the beginning of the interval is ‘one value of the base’ below the expressed end of the interval (i.e. 40) (Menninger, 1969: 78).

We will see below that Dzongkha uses ‘half’ and ‘three-quarters’ in fractional overcounting to construct numbers larger than the vigesimal base. Thus we can add the ‘three-quarter count’ to Menninger’s ‘half-count’.

3. The Dzongkha number system

Dzongkha, the national language of Bhutan, has two number systems, used in different circumstances[5]. A decimal system, borrowed from Tibetan, is used in formal circumstances; a vigesimal system, based on the number ‘20’, is used in everyday life, or at least it was thirty years ago when Boyd Michailovsky and I collected it in 1977[6]. That system provides examples of several of the less common characteristics mentioned above. We will describe it in some detail before pointing out several other Tibeto-Burman languages where yet other bases are alive or traceable, or where unusual constructions are more obviously apparent. For a complete description of the Dzongkha system see Mazaudon (1982; 1985).

From 1 to 19, the numeral system of Dzongkha is purely decimal. Table 2 lists their names.

Table 2. Dzongkha number names from 1 to 19

1	ciː	11	cu-ci	10 (+) 1
2	ˈɲiː	12	cu-ɲi	10 (+) 2
3	sum	13	cu-sum	10 (+) 3
4	ʑi	14	cy-ʑi	10 (+) 4
5	ˈŋa	15	ce-ŋa	10 (+) 5
6	ɖhuː	16	cu-ɖu	10 (+) 6
7	dyn	17	cup-dỹ	10 (+) 7
8	geː	18	cop-ge	10 (+) 8

[5] For more details about Bhutan’s language situation see Michailovsky (1994).
[6] The list of vigesimal numbers was inserted in a short manual published by the government of Bhutan (Anonymous, 1977).

9	guː	19	cy-gu	10 (+) 9
10	cu-thãm ’10 full’ *or* cu	20	ɲiɕu	2 (x) 10

The names of numbers from 1 to 10 are unanalyzable roots, while numbers from 11 to 19 are formed transparently from a variant of the word ‘ten’ /cu/ (WT *bcu*) followed by the digits 1 to 9. The names of numbers from 11 to 19 are not obtained by a productive synchronic syntactic construction, but are the modern reflection of ancient compounding; there is some degree of vocalic harmony, as in ‘15’, with /ce/ as the form of ‘10’ instead of /cu/, and tonal and consonant changes as in /ɲiɕu/, with /ɕ/ instead of /c/ and a low tone instead of the high tone found in the simple word for ‘two’.

From twenty on, two parallel systems exist. One is decimal, and extremely similar to that of Tibetan. That is the system which was first obtained from elicitation. It is used in formal circumstances and in writing. As we worked on a complete description of the numeration, we eventually came across doublet forms, which revealed a different construction of the numbers. It then appeared that there existed a complete vigesimal system for all numbers up to the highest values. Table 3 shows the two systems in parallel.

Table 3. The two numeral systems of Dzongkha over twenty

	Decimal		*Vigesimal*	
20	ɲiɕu	2 (x) 10	khe ciː	20 (x) 1 “one score”
30	sum-cu	3 (x) 10	khe pɟhe-da ˈɲiː	20 ½- 2 “half score to 2 (scores)”
40	ʑi-p-cu	4 (x) 10	khe ˈɲiː	20 (x) 2 “two scores”
50	ˈŋa-p-cu	5 (x) 10	khe pɟhe-da sum	20 ½- 3 “half score to 3 (scores)”
etc.				
100	ɟa-thampa *or* cik-ɟa	100 full 1 (x) 100	khe ˈŋa	20 (x) 5 “five scores”
200	ɲi-ɟa	2 (x) 100	khe cuthãm	20 (x) 10-full “ten scores”
300	sum-ɟa	3 (x) 100	khe ceŋa	20 (x) 15 “fifteen scores”
400	ʑi-p-ɟa	4 (x) 100	ɲiɕu ciː	400 (x) 1 “one ɲiɕu”

It will be immediately noted that the same word / ɲiɕu / appears in both systems, but with different values. No confusion exists: in the decimal system / ɲiɕu /, meaning ‘twenty’, is used alone, while in the vigesimal system, where / ɲiɕu / means 400, it is always accompanied by a multiplier, including the number 1.

Shifts in the names of bases and powers of bases are not uncommon. For instance in American English the word *billion* means ‘a thousand millions’ 10^9, whereas in England it means ‘a million million’ 10^{12}. In Chepang, the word /hale/, which is the exact etymological correspondent of Dz /khe/ and WT *khal*, designates not twenty, but twelve. As we will see in a moment, this is because the Chepang system is duodecimal: ‘12’ is its base.

3.1 The Dzongkha decimal system

In the Dzongkha decimal system each multiple of 10 is expressed by a multiplier followed by the root for 10. Each power of ten receives a new name. Thus when we reach 100 we encounter the word /ɟa/ as /ɟa-thampa/ 'hundred-full' in counting, insisting on the fact that a level has been reached, or as /cik-ɟa/ 'one hundred', as the first in the sequence of counted hundreds, which are expressed by multiplication just as the tens are. This is parallel to the construction in Written Tibetan. Note that the multiplier /cik/ is not aspirated in Dzongkha, while it is in Tibetan.

Multiples of the base

The names for powers of 10 are expressed by simple names up to 10^8. 'thousand' is /toŋ/ used as /cik toŋ/ '1-1000' or in a compound with the suffix /ʈha/ (WT *phrag* 'interval'), which is then counted according to the normal Dzongkha syntax: noun + quantifier, /tõ-ʈha ciː/ '10^3-group one'. The names of the higher powers of ten are borrowed from Tibetan as evidenced by the aspirated form of the multiplier 'one' in /chikʈhi/ '10^4' and by the treatment of WT *by* as /ɟh/ instead of the regular Dzongkha treatment /pɟh/ in /ɟhewa/ '10^7'.

Intermediate numbers

Intermediate numbers are formed by addition. In these words we recognize a reduced form of the name of the decade followed by the name of the unit, for example gu-p-cu '90' (9.10) (compare WT *dgu-bcu*) vs. gho-ci '91' (90.1). These two morphemes form a single word, and thus undergo (or rather have historically undergone) some phonological modification: the second syllable loses its tone, its length, and some contrasts on its initial consonant. The reduced form of the name of the decade corresponds historically to an unprefixed form of the free number. The names of numbers have interesting etymological, morphological and phonological peculiarities on which more details and explanations can be found in Mazaudon (1982, 1985). From the point of view of their mathematical structure however they use multiplication and addition in a very straightforward manner.

3.2 The Dzongkha vigesimal system

In the vigesimal system the name of the base, twenty, is the unanalyzable morpheme /khe/. It is always followed by a multiplier, including one, as we already noted for its next level of rank, or power, /ɲiɕu/ '400'. This applies as well to the higher powers of 20. The form is phonologically two words, each one keeping its full form including its tone: compare /khe ˈɲiː/ '40', or '2 scores' with /cu-ɲi/ '12' (10 + 2) where the second morpheme has lost its length and its tone. Syntactically the construction respects the usual structure of noun plus quantifier as in /ra ˈɲiː/ '2 goats'.

The multiples of twenty are expressed by the regular succession of the multipliers until 20^2. There is no break at the hundreds: '100' is simply 20 x 5, '200' = 20 x 10, '300' = 20 x 15 (see table 3).

Intermediate numbers: Fractional count and overcounting

From 1 to 9 in each score (or 20-group), the unit is added to the multiple of the score, and the words are joined by the all-purpose connector /da/. The same applies to numbers equal to a multiple of 20 plus a number from 11 to 14 and from 16 to 19 (see table 4 for examples of numbers between 21 and 39).

At the half step between two exact multiples of the base, a new construction appears. '30', '50' etc. are not expressed as 'one score plus ten', 'two scores plus ten', but as a half-count with the help of the word /pɟhe/ 'half', the same as in /phop pɟhe/ 'half a cup'. They are also expressed by reference to the higher limit of the interval concerned. So '30' is /khe pɟhe-da ˈɲiː/ (20 half-*da* two) or 'half in the second twenty-group'.

We have seen this construction in several European languages. What is much more unexpected is the expression of the numbers ‘35’, ‘55’ etc., the numbers situated three-quarters of the way in the interval between two steps in the gradation (see figure 1). These are formed with the help of the word /ko/ ‘three-quarters’, the same as in /phop ko/ ‘three-quarters of a cup’.

Table 4: Intermediate numbers in the Dzongkha vigesimal system

21	khe ci: da ci:	20 1 *da* 1
22	khe ci: da 'ɲi:	20 1 *da* 2
23	khe ci: da sum	20 1 *da* 3
24	khe ci: da ʑi	20 1 *da* 4
25	khe ci: da 'ŋa	20 1 *da* 5
etc.		
30	khe pɟhe-da 'ɲi:	20 ½ *da* 2
31	khe ci: da cu-ci	20 1 *da* 11
32	khe ci: da cu-ɲi	20 1 *da* 12
33	khe ci: da cu-sum	20 1 *da* 13
34	khe ci: da cy-ʑi	20 1 *da* 14
35	khe ko da 'ɲi	20 ¾ *da* 2
36	khe ci: da cu-ɖu	20 1 *da* 16

Note that ‘5’ added to a multiple of the base, as in ‘25’, ‘45’, ‘65’ and so on, is expressed by the word ‘five’ and not by a word meaning ‘a quarter’. In fact Dzongkha does not have a word for ‘a quarter’. St Quintin Byrne already noted the fact in 1909, and quoted *phye gi phye* ‘half of a half’ as the translation for ‘one quarter’[7].

This may sound awkward, but we should reflect that fractions of larger numbers are groupings, like the bases or their multiples. If groupings are a way to help the mind envision a quantity larger than is easy to conceive by enumeration, it can be understood that there is no incentive to use a fraction for ‘5’, but much more to use it for ‘10’ or ‘15’. We will see later that in Ao Naga, which does not use fractions but resorts to overcounting, it is used only when more than half the value of the interval between multiples of the base is reached. This seems to follow the same logic.

A figurative representation of the use of fractions in overcounting is presented in figure 1. Like the Maya languages of Central America, the Dzongkha fractional construction builds the numbers by reference to the interval between two multiples of the base, rather than by reference to the lower multiple, as in English, or to the higher multiple of the base as in subtractive constructions (see the Latin example above).

The same construction is used for higher numbers formed on the names of the powers of the base: see ‘600’ and ‘1.100’ in figure 1.

---

[7] G. van Driem incorrectly analyzes this construction as a subtraction: « fourth less than fortyscore », « half less than fortyscore ». If a doubt is (mathematically) possible with the half count, for the 3/4 count the evidence that the meaning of /ko/ is ‘three-quarters’ and not ‘one quarter’ is compelling. Besides the modern value quoted above for ‘3/4 of a cup’, we can refer to St Quintin Byrne again. The word *ko* appears in his work for a coin whose value is ‘12 *annas*’, which is 3/4 of a rupee. In Jirel, we find the same word used with a measure word as /padi koo/ ‘3/4 of a *pathi*’ (one pathi is roughly one gallon) (Hari, 2004: 716).

Figure 1: Fractions and overcounting in Dzongkha

• In the expression for **30:** khe pɟhe-da ˈɲiː
20 ½-*da* 2

are expressed: the value of the interval = 20^1 (a power of the base)
the fraction of the interval reached = ½
the upper limit of the reckoning interval (expressed as a multiplier of the value of the interval) = 2 (scores)

```
 0                          20            30          40
 |--------------------------|========>*--------------|----------------------------|--
                      khe ciː          pɟhe         khe ˈɲiː
                                   fraction ½
```

• In the expression for **55**: khe ko-da sum
20 ¾-*da* 3

are expressed: the value of the interval = 20^1
the fraction of the interval reached = ¾
the upper limit of the interval = 3 (scores)

```
20                          40                   55     60                         80
--|-------------------------|==============>*--------|----------------------------|--
                     khe ˈɲiː                    ko     khe sum
                                        fraction ¾
```

• In the expression for **600**: ɲiɕu pɟhe –da ˈɲiː
400 ½-*da* 2

are expressed: the value of the interval = 20^2 = 400
the fraction of the interval reached = ½
the upper limit of the interval = 2 (i.e. the 2nd interval of value 400)

```
 0                          400           600         800                          1200
 |--------------------------|=========>*--------------|----------------------------|--
                     ɲiɕu ciː          pɟhe         ɲiɕu ˈɲiː
                                  fraction ½
```

• In the expression for **1100**: ɲiɕu ko-da sum
400 ¾-*da* 3

are expressed: the value of the interval = 20^2 = 400
the fraction of the interval reached ¾
the upper limit of the interval = 3 (i.e. the 3rd interval of value 400 = the interval from 800 to 1200)

```
400                         800                  1100   1200                       1600
--|-------------------------|==============>*-------|----------------------------|--
                      ɲiɕu ˈɲiː                  ko    ɲiɕu sum
                                        fraction ¾
```

In Dzongkha number names, fractional expressions can occur only at the end of the number; they are not stepping stones for building other numbers. So ‘31’ cannot be built by adding ‘1’ to ‘30’ (as 21 is built by adding 1 to 20; see table 4). The system reverts to the vigesimal progression, and ‘31’ is expressed as 20 + 11.

In the Bumthang language, spoken in the East of Bhutan, another Bodish language, very close to, but outside of, the group of Tibetan dialects, we found a very similar numerical system. There, it is possible to use the fractional expression inside a more complex number name:

Bumthang: 31 =	khe	phedaŋ	zon	niŋ	the
	20	½-*dang*	2	and	1

In that language /daŋ/, which corresponds to Dz /da/, has become fused with the word ‘half’, and a different connector /niŋ/ is used for addition.

It should be noted that fractions in overcounting are not used for numbers smaller than the base: /phop ko/ is ‘three quarters of a cup’, but you cannot say in Dzongkha ‘three-quarters of twenty’ to express ‘15’ nor ‘three-quarters of 400 (a score of scores)’ to mean ‘300’ (see /khe ceŋa/ (20.15) in table 3).

4. Overcounting in other Tibeto-Burman languages

4.1 Other vigesimal systems with fractions and overcounting

Number constructions using fractions and overcounting are found in several other Bodish[8] languages, although not in such a well preserved state as in Dzongkha.

Bodish languages

In Kaike (Bodish, non-Tibetan) the word list collected by Fisher indicates a similar system.

40	nghe thal	2 (x) 20	50	pheraang sumthal	½-*rang* 3x20
60	sum thal	2 (x) 20	70	pheraang lithal	½-*rang* 4x20

The questionnaire did not include any of the numbers which could have revealed a ¾ fractional form (35, 55 etc) or the number ‘400’ (Fisher 1971; Weera 1999).

Similarly, in Gongar, a Bodish language spoken in Eastern Bhutan and in the Kamrup district of Assam, we find the form *khay phedang nga*, meaning ‘90’, which is analyzable as ‘twenty.½-*dang*.five’, i.e. halfway through the fifth interval of twenty (Hofrenning 1959: 5). Judging from the very scanty information on the language, Gongar seems very close or identical to Tshangla.

In a description of Sharchokpa, also called Tshangla, spoken a little further North in East Bhutan, in the area of Tashigang, S. Egli-Roduner quotes a simple name for the number ‘400’: *nyzhu* or *nyzh-thur*, where *thur* means ‘one’. ‘hundred’ is expressed as *khe-nga* (20.5), and there is no competing decimal (monomorphemic) name for it. So the vigesimal system seems still alive. Overcounting with fractions is also reported: *khe-phedang-nigtsing* (20. ½-*dang*. 2) ‘thirty’, *nyshu-phedang-nigtsing* (400. ½-*dang*. 2) ‘six-hundred’ (Egli-Roduner 1987: 203 sqq).

---

[8] This term belongs to Shafer’s classical classification of Sino-Tibetan languages (Shafer 1955). It is outdated in many points, but its Bodish section, after removing Gyarung, holds rather well. The Bodish section is one of the sections of the large Bodic division. More discussions of classification can be found in Thurgood and LaPolla (2003).

Lepcha

Outside the Bodic division, but in the same geographical area, Lepcha, an unclassified Tibeto-Burman language of Sikkim, still had a rather rich vigesimal system when its descriptor, Mainwaring, undertook to amend it. He describes how the numbers are formed after 'twenty' proceeding on by scores, e.g. *khá nyat sa kati* (20.2 and 10) for 'fifty', and adds a footnote.

> "This form was of course very cumbersome and awkward, entirely preventing the teaching of ordinary arithmetic. A decimal mode was necessary; which the Lepchas, on being taught, at once saw the advantage of, and learned with avidity. In a school which I established, during the short period I had the opportunity of conducting it, the Lepchas made great progress in arithmetic; and now, on all occasions when counting, they adopt the decimal mode taught them" (Mainwaring 1876: 115).

In spite of his efforts the Lepcha vigesimal system survives to this day as the only system up to 100. It has two parallel versions: one using the additive principle, as in *khá kát sá kati nyet thap* (score.1 and 12) 'thirty-two', where *kati nyet thap* 'twelve' is (10.2-added). The alternative construction uses the fractional half-count: *khá báng nyet* 'thirty' (score half two), allowing the addition of units to the fractional forms: *khá báng nyet sá kát* 'thirty-one' (score half two and one). This second construction, which was not mentioned by Mainwaring, is used by Lepchas of Sikkim and is less familiar to the Lepchas of Kalimpong (Plaisier 2006: 91).

4.2 Overcounting without fractions: Ao and Sema Naga

In 1893 Mrs. E. W. Clark published a grammar of Ao (Clark 1893), in which she included the number system. Table 5 presents the data.

Table 5. Ao numbers: decimal, overcounting without fractions

1	ka	11	teri ka
2	ana	12	teri ana
3	asy̥m	13	teri asy̥m
4	pezy̥	14	teri pezy̥
5	pungu	15	teri pungu
6	trok	16	metsy̥ maben trok
7	tenet	17	metsy̥ maben tenet
8	ti	18	metsy̥ maben ti
9	ty̥ko	19	metsy̥ maben ty̥ko
10	ter	20	metsy̥
20	metsy̥	26	semy̥r maben trok
21	metsy̥ri ka	27	semy̥r maben tenet
22	metsy̥ri ana		
30	semy̥r	36	lir maben trok
40	lir	46	tenem maben trok
50	tenem	56	roky̥r maben trok

100	telang *or* noklang	90	telang tyko

In her commentary on the sequence of number names Mrs. Clark points out the principle of overcounting.

> "As will be observed from the above, the Aos have distinct names for the digits, and the compounds are regularly formed up to sixteen, as, ten and one are eleven 'teri ka', ten and two are twelve 'teri ana', &c.; also from twenty to twenty-six, twenty and one 'metsyri ka', twenty and two 'metsyri ana', &c. The same with thirty, forty, &c. But when the six is reached in the compounds, the succeeding ten seems to be anticipated, and we have for sixteen 'metsy maben trok' twenty not brought six, equivalent to the sixteen before twenty; […] . In the same manner from twenty-six to thirty, on reaching the six, thirty is anticipated, thus 'semyr maben trok' the six before thirty, twenty-six […] &c." (Clark 1893: 45).

In several dialects of Sema, another Naga language, Hutton pointed out that the overcounting expression was used with a different frequency depending on the number: it was always used with numbers ending in 9, often for numbers ending in 7 or 8, where an alternative formation by addition was also possible. Hutton draws attention to the logic of the shift of construction at 6; in this way speakers systematically refer to the closest round number: they add to the lower multiple of the base up to 5 steps in the interval, then they shift to referring to the next higher multiple of the base, the higher limit of the interval (Hutton 1916: 4). This can be seen as the cognitive motivation for a mixed system like that of Ao or Sema as compared to fully regular systems, whether of the additive type like English, or of a fully regular overcounting structure like Maya: reference is always made to the closest grouping.

Table 6 Some Sema numbers

1	laki, khe	11	chüghi-khaki
2	kini	12	chüghi kini
3	küthu	13	*(not listed)*
4	bidhi, bidi	14	*(not listed)*
5	pongu	15	*(not listed)*
6	tsogha, soghoh	16	*(not listed)*
7	tsini	17	mukuma tsini *or* chüghi tsini
8	thache, tache	18	mukuma tache
9	thuku, tuku	19	mukuma tuku
10	chüghi	20	muku

As we can see, '17' has two forms: one overcounting (20-*ma* 7), one additive: 10 (+) 7. The numbers ending in 8 and 9 only have the overcounting form.

Modern Ao and Sema have apparently lost all trace of their old systems. Neither Sreedhar nor Gowda mentions anything else than a regular additive decimal system (Sreedhar 1980; Gowda 1975). The automatic influence of increasing bilingualism was certainly at work, but the conscious efforts of educators cannot be ignored. Thus Mrs. Clark, although she had well recognized the logic of the Ao system, saw it as an obstacle to progress:

> "This method of counting is very objectionable to children learning the use of figures, as in adding up a column if the amount is seventeen – 'metsy maben tenet' – the mind catches the twenty, and *two* is very likely carried instead of *one* to the next column. In the schools an effort is being made to discard the above irregularities, and count regularly thus, 'teri trok' sixteen, 'metsyri trok' twenty-six, &c." (Clark, 1893: 45).

5. More non-decimal systems in Tibeto-Burman

5.1 Vigesimal systems

Bodic languages

Most Tibeto-Burman vigesimal systems are found in the Bodic division of Shafer's classification, which comprises languages spoken all along the Himalayas. We have already examined the most extensive system, that of Dzongkha, and three other languages of the Bodish section of the Bodic division, which, like Dzongkha, have the rare characteristics of using fractions in number building, and overcounting. These two features are in no way linked to vigesimal systems, as we already saw with the use of overcounting in the Ao decimal, non-fractional system.

Vigesimal without fractional count or overcounting

Complete lists of numbers are not available for most languages, but from word lists and dictionaries we can see that the Tibetan dialect of Balti spoken in north-eastern Pakistan can build numbers on the base 20 at least up to '90': *ngis-bcu bzhi* '80' (20.4), *ngis-bcu bzhi-na bcu* '90' (20.4-and 10) or *rgu-bcu* (9.10) (Sprigg 2002: 122). The Tibetan dialect of Jirel, spoken in Nepal, builds vigesimal numbers up to '400' but does not have a name for '400', which is called 'a score of scores' /khal³ khal-jyik³/ (20.20.1) (Strahm and Maibaum 2005: 789). At least two more outlying dialects of Tibetan inside Tibet itself, Thewo and Shigar, count in twenties up to 100, e.g. Shigar /niɕu nis na cuksum/ (20.2 and 13) '53' (Nicolas Tournadre, collection 2006).

Tamang, a non-Tibetan Bodish language of Nepal, has a vigesimal system without overcounting or fractions. In the Risiangku dialect 'fifty' is /⁴pokal ⁴niː-se ²ciui/ '20.2-from 10', 'fifty-one' /⁴pokal ⁴niː-se ²cukkhrik/ '20.2-from 11', etc. The system is not as extended as that of Dzongkha, having no name for '400' or higher bases, but it is the only system used up to 400: there is no competing decimal structure (Mazaudon 2002, 2003; Hale 1973). At least not yet – some well-meaning promoters of the language are trying to build such neologisms but speakers are deriding their efforts and resisting. The future may unfortunately be global borrowing of the Nepali or English number names. Note that, although overcounting is not presently used in Tamang general counting, an old lady gave her age in 2003 by saying: /⁴kjarca-ri ⁴niː-tiŋ ²joːsi ²som-tiŋ ²waŋci/ |hundred-in two-years completed three-years entered| 'I am eighty-two years old'.

The "Monpa" languages of Dirang and of Tawang in Arunachal, according to a very brief survey conducted in 2007, have vigesimal constructions up to 100: ex. Tawang /khei bli nei/ (20.4.2) 'eighty-two'.

The Magar language of Nepal (tentatively classified in one of the Himalayish sections of Bodic) has relexified a vigesimal system with Nepali loan roots for all numbers over 5. Thus, with loan words in italics, /nis *bis*/ (2.20) 'forty', /nis *bise das*/ (2.20 + 10) 'fifty', /som *bis*/ (3.20) 'sixty', /nis maː-hol-na *say*/ (2 not reaching 100) 'ninety-eight' (with a subtraction), /*pãːc bise* nis/ (5.20 + 2) 'hundred and two'(Hale 1973).

Vigesimal with additive fractional count: Bahing

A vigesimal system with a fractional structure was reported for Bahing, a Kiranti language of East Nepal, in the nineteenth century (Hodgson 1857, 1880, vol. 1: 331). The base '20' has a simple name *asim* used with a multiplier as *kwong ásim*, and a secondary base 10 is used for numbers from 11 to 19: *kwaddyum* '10' *kwaddyum kwong* (10.1) '11'. 'forty' is *niksi ásim*, 'forty-one' is *niksi ásim kwong*. The odd decades, thirty, fifty, etc., are expressed by fractional expressions, which are constructed by addition to the lower multiple of the base: 'thirty' is *kwong ásim*, *kwong áphlo* which Hodgson translates word for word as 'one score, one its half'; 'fifty' is *niksi ásim áphlo* (2.20 its-half), and so on for seventy and ninety. The numbers cited stop at 102, *gnó ásim niksi* (5.20.2). From a few intermediate numbers quoted by Hodgson we can see that a fractional number could, as in Bumthang, be followed by a smaller digit, as in ‘fifty-one’ *niksi ásim áphlo kwong* (2.20 its-half 1).

Non-Bodic languages

Manipuri (or Meithei) and Karbi (or Mikir), spoken along the India-Burma border, show traces of a 20-based system (Chelliah 1997: 85-86; Jeyapaul 1987: 91). These languages have been placed in the Kuki-Chin-Naga group, but Burling prefers to leave them unclassified (Burling 2003: 187).

Bodo-Garo languages (Shafer’s Baric division of TB) are spoken in several pockets in the Western and Southern areas of Northeast India. They may be the most vigesimal languages of Northeast India. Remnants of a vigesimal system are found for instance in Kokborok (or Tripura or Tipra) with forms like: /khol-pe/ '20', /khol-pe ci/ '30' (20+10), /khol-nui/ '40' (20 x 2), /khol-nui ci/ '50' (20x2+10), /khol-brui ci/ '90' (20x4+10). But ‘100’ is already decimal (Chakraborti 1971).

In Garo, the modern system is globally decimal, but vigesimal from 20 to 39: /kol-grik/ ‘20’, /kol-grik-sa/ ‘21’, /kol-a-chi/ ‘30 (20-*a*-10)’, /kol-a-chi-sa/ ‘31 (20-*a*-11)’. Older people still remember a more complete vigesimal system, with /kol-chang/ as the name of the score in counting: /kol-chang-gini/ (20-group 2) ‘40’, etc. /kolchang-bri ge chi-sku/ (score-4 classifier 19) ‘ninety-nine’ (Burling 2004: 245).

Bodo, which we will consider in more detail later, has several competing ways of counting. First the ‘theoretical academic’ usage provides a way to count regularly in a decimal system up to 100. But, over 5, common usage either borrows Assamese numbers or uses a base-4 system which we will see later (Bhattacharya 1977: 135). Traces of a special status for ‘20’ are also found in the alternate name ²khu¹ri, which Bhattacharya says is borrowed from Assamese *kudi*, itself borrowed from “Austric”, according to Chakraborti.

One of the words for ‘eighty’ in Bodo is ²phə²nay¹se (40.2.1) or ‘one double-forty’, where the word ²phə, meaning apparently ‘forty’ – since a ‘double-forty’ is ‘eighty’ – seems to be cognate to another common Tibeto-Burman root for ‘twenty’, PTB *bo, (WT *'bo*, Tamang ⁴po in ⁴pokal, Meithei phu etc.), originally a measure word like the other root for ‘twenty’ which we have encountered, PTB *kal.

The Lolo-Burmese languages (the Burmish section, sister branch of Kuki-Chin-Naga in Shafer’s Burmic division), spoken further east, in Burma, Thailand and South China, seem to show no example at all of vigesimal systems.

No vigesimal system is reported in the Karen languages either. This leaves us with a geographical distribution in or close to the Himalayas. Whether this reflects the influence of a common substratum, maybe Austro-Asiatic, or whether these mountainous areas have been able to retain a previously more widely spread structure remains an open question for the moment.

5.2 Other bases in Tibeto-Burman: five, twelve, four

5.2.1 Quinary systems

Except for the etymological relationship between PTB roots *g-nyis ‘2’ and *s-nyis ‘7’, still visible in many TB languages (Matisoff 1997: 84), traces of quinary systems are rare in Tibeto-Burman languages.

Hayu

Modern Hayus (Himalayish section, Nepal) only count up to four in their own language. Upwards, they use borrowed Nepali numerals (Michailovsky 1988). At the end of the 19th century, Brian Hodgson recorded numbers up to ‘6’, and also listed a series of what he called “numeral collectives” which constitute a quinary-vigesimal system, using the names for 'hand' and 'foot' to figure groupings of 5 and 10. Intermediate numbers in that system are not mentioned. We cannot know if they ever existed or not. Table 7 lists the names of numbers after Hodgson (Hodgson 1857, 1880, vol. 1: 232)

Table 7. The quinary-vigesimal system of Hayu

5	kolu got' khulup	1 hand entire
10	nayung got' khulup	2 hands entire
15	nayung got' khulup-ha kolu got' khulup	2 hands entire-and 1 hand entire
or	nayung got' khulup-ha ba khulup	2 hands entire-and ½ (*ba*) entire
20	le got' khulup	feet hands entire
or	cholok, kolu cholok	20, 1 score
40	nayung cholok	2 score
60	chhuyung cholok	3 score
80	blining cholok	4 score
100	uning cholok	5 score
or	kolu got' cholok	1 hand of scores

We may notice that cholok 'twenty' is a simple unanalyzable word, whereas the words for 'five' and 'ten' are built with the modern forms of the words for 'hand' and 'foot'. This suggests that the names for 5, 10 and 15 may possibly be refections. 'fifteen' has two alternate names, one expressing a plain addition (2 hands entire + 1 hand entire), and the other making use of a fraction (2 hands entire + half of 2 hands entire). This, and the Bahing construction (§5.1) are the only examples we have found in TB of a fraction used in an additive construction. This is also the only fractional expression used below the base (below 20). We would like to know how 'thirty' was said, but Hodgson does not tell us.

At the time of Hodgson, Hayu numbers were used in all functions, e.g. in adverbial expressions with the word *phi* 'times' as in *nayung got khulup phi* 'ten times'.

The first collection of Hayu since Hodgson was done in the late 1970's in the village of Murajor (Michailovsky 1988). A visit to another village, Adhamara, in 2005-6, revealed the survival among older speakers of the following forms (Michailovsky, personal communication):

Table 7 b. Adhamara Hayu numbers

5	kolu got khumluk	1 hand entire
10	naʾung got khumluk	2 hand entire
15	naʾung got kolu le	2 hand 1 foot
20	lewo got khumluk	foot-*wo* hand entire
or	kolu khɔsin	1 *muri* (a volume measure worth 20 *pathis*). This measure word can be used for '20'

Yakkha

Gvozdanovic mentions several number systems in Kiranti languages (Himalayish, East Nepal) which have partially resisted the influence of Nepali. Some of these languages, like old Hayu, seem to have re-built number names after a quinary-vigesimal pattern from the modern name of the hand, sometimes the foot. Yakkha is one of them (Gvozdanovic 1985: 137).

Table 8. Yakkha numbers

1	kolok	1
2	hitci	2
3	sumci	3
4	sumcibi usongbi kolok	3 and 1
5	muktapi	hand
6	muktapi usongbi kolok	hand and 1
7	muktapi usongbi hitci	hand and 2
8	muktapi usongbi sumci	hand and 3
9	mukcurukbi kolok hongbi	hand-s 1 less
10	muktapi hita	hand 2
20	lang-curuk-muk-curuk	foot-s hand-s = 20

We see in table 8 number names which are built on numbers which are not bases in the system, like ‘4’ as ‘3+1’. Such formations are found in other very damaged systems. They seem to be repairs rather than relics of older structures. Only comparative evidence would help show their etymological status. The system stops at ‘twenty’ and ‘twenty’ does not have a simple name, as one would expect of a base.

5.2.2 A lone duodecimal system: Chepang

Although groups of twelve are used in Sino-Tibetan languages and cultures to express time in hours, months and years, there do not seem to exist 12-based, or duodecimal, numeral systems, except for an interesting trace in the Chepang language of Nepal. Until very recently the thirty thousand Chepangs of central South-West Nepal led a semi-nomadic life, close to the forest, which they still now use for hunting and gathering. They used to supplement this activity with slash and burn agriculture, which is progressively replaced by sedentary agriculture. Their way of life, different from that of their neighbours, has contributed to restrict their contacts outside their group (Caughley 2000).

Ross Caughley, who has surveyed practically all the places where Chepangs live, says that there is no Chepang nowadays who can count over 5 in his own language. Nepali numbers, along with their decimal system are used above five. But in certain circumstances, especially when counting game like bats or birds, a twelve-based system is used. Using earlier publications by Caughley (Caughley 1972) I have proposed to identify the Chepang word for twelve /hale/ as the regular cognate of the root /khal/~/kal/ meaning ‘twenty’ in Bodish languages (also cognate to /khol/ in Bodo-Garo) (Mazaudon 1982, 1985).

We have already encountered examples of the shifting of the name of a base to another base, as in Dzongkha where /ɲiɕu/, etymologically ‘twenty’, shifted in the vigesimal system to be the name of the base of second rank ‘four-hundred (20x20)’, or in English where *billion* is either 10^9 or 10^{12}, depending on the dialect. The measure word *khal, when it integrates the numeral system, means: ‘base of the system’ that is ‘smallest grouping on which higher numbers are constructed by raising to a power’. Its numerical value can thus vary from one language to the next. Duodecimal numbers exist in Chepang up to ‘fifty’ (Table 9).

Table 9. Chepang numbers

1	yat, ya.jyo⁷	13	yat hale yat (jyo⁷) |1.12(+)1 cls|
2	nis	14	
3	sum	15	yat hale sum.jyo⁷ |1.12(+)3 cls|
4	pləy, ləy		
5	poŋa	19	yat hale *sat.gota* |1.12(+)7.cls|
6 to 11	(*Nepali*)	20	yat hale *at.gota* |1.12(+)8.cls|
12	yat hale |1.12|	29	nis hale poŋa.jyo⁷ |2.12(+)5.cls|
60	poŋa hale |5.12|		

jyo[7] is a classifier, like its Nepali equivalent *gota*. Caughley agreed with my interpretation, and brought forth two arguments to support it. First, the Chepangs, as is common all over the Himalayas, still use a method of tallying on the phalanges (finger segments) – of which there are 3 per finger – when counting. Tallying on four fingers brings us to twelve. More intriguing is Caughley's other discovery. The Chepang recite a series of numbers which they consider as "a mythological spirit system of counting" (Table 10).

Table 10. The Chepang spirit numbers

1	ya	5	poŋa	9	trak
2	gi	6	prek	10	--
3	sum	7	taguji	11	--
4	kləy	8	hlukum	12	--

The list ends at nine and contains a strangely long form for '7'. When comparing these numbers with the ordinary numbers quoted by Hodgson a century before, Caughley was able to reconstruct the mystery: /taguji/ is a conflation in recitation of the words for 'nine' (Hodgson *tagu*) and 'ten' (Hodgson *gyib*). He thus reconstructed an older version of the spirit system as the likely old Chepang system, without 'six' and 'seven' which have been lost, and where the rest of the list falls into place with the generally admitted reconstruction of PTB numbers up to 10 (Table 11). The etymology of the words for 'eleven' and 'twelve' remains mysterious.

Table11. Restituted Chepang (spirit/old) numbers

1	ya	5	poŋa	9	tagu
2	gi	6	--	10	ji
3	sum	7	--	11	hlukum
4	kləy	8	prek	12	trak

5.2.3 Counting by fours: base or grouping ?

Bodo, Deuri

Boro, or Bodo, has several parallel numeral systems, and each number has many variants (Bhattacharya 1977: 134 sqq). As in the other Bodo-Garo languages, numbers are accompanied by classifiers, and, in the case of a number where units remain to be expressed after the multiples of the base, a classifier intervenes inside the numerical expression between the round number and the units[9].

[9] For an analysis of the insertion of 'unit-counters' in Dzongkha, a language devoid of classifiers, before the units inside a complex number, see Mazaudon (1985: 144-145)

Table 12. Bodo numbers: groupings of 4

	Decimal	*Counting by fours*	
1	-^{2}se		
2	-^{1}nəy		
3	-^{2}tham		
4	-^{1}brəy / -^{1}bri		
5	-^{1}ba		
6	-^{2}do / -^{2}ro	**^{2}zo^{2}khay**^{1}se *^{2}mən*^{2}nəy	(4.1 cls 2)
7	-^{1}sni / -^{2}si^{1}ni	^{2}zo^{2}khay^{1}se *^{2}mən*^{2}tham	(4.1 cls 3)
8	-^{1}zad/ -^{1}daŋ/ -^{1}dayn	^{2}zo^{2}khay^{2}nəy	(4.2)
9	-^{2}si^{1}kho/ -^{2}sugu/ -^{1}ne	^{2}zo^{2}khay^{2}nəy *^{2}mən*^{2}se	(4.2 cls 1)
10	^{1}zi / ^{1}zu/ ^{2}khaw^{1}se		
11	^{2}zi^{1}se / ^{2}khaw^{1}se *^{2}mən*^{1}se	^{2}zo^{2}khay^{2}nəy *^{2}mən*^{2}tham	(4.2 cls 3)
12	^{2}zi^{1}nəy / ^{2}khaw^{1}se *^{2}mən*^{1}nəy	^{2}zo^{2}khay^{2}tham	(4.3)
16	^{2}zi^{1}do/ ^{2}khaw^{1}se *^{2}mən*^{1}do	^{2}zo^{2}khay^{1}brəy	(4.4)
20	^{2}nəy^{1}zi / ^{2}khaw^{2}nəy	^{2}zo^{2}khay^{1}ba	(4.5)
30	^{2}tham^{1}zi / ^{2}khaw^{1}tham	^{2}zo^{2}khay^{1}sni *^{2}mən*^{2}nəy	(4.7 cls 2)
40	^{2}brəy^{1}zi / ^{2}khaw^{1}brəy	^{2}zo^{2}khay^{1}zi	(4.10)
80	^{2}dayn^{1}zi / ^{2}khaw^{1}daŋ	**^{2}phə^{2}nay**^{1}se	(80.1)

Numbers in the first column are decimal. From one to five they are commonly used. After five it is more common to use either borrowed Assamese numerals, or the word /^{2}zo^{2}khay/ meaning 'a quartet, or group of 4' followed by a multiplier, as in /^{2}zo^{2}khay^{2}nəy/ (quartet 2) '8'. Such forms are followed by the number of units to be added to the multiple of the base, obligatorily preceded by a classifier, as in /^{2}zo^{2}khay^{2}nəy ^{2}mən^{2}tham/ (4x2+3) '11', where /^{2}mən/ is a general classifier. The system seems to break down after 40. Also we notice that 16 (4^2) does not have a simple name: it is expressed as 'four quartets' not as 'a quartet of quartets'. So the group of four seems to be at most an auxiliary base, or a simple grouping. It is impossible to say if a complete base-4 numeral system ever existed in Boro.

Deuri, another Bodo-Garo language of Assam, also uses quartets from 4 to 19, after which it uses scores from 20 to 99 (Jacquesson 2005: 264). Complex expressions are built for intermediate numbers, calling upon the 5-group represented by the word 'hand'. So '30' is /kôa-ca utû(m) pe-kini/ (score-one hand classifier-two) i.e. 20 + 2 hands = 30.

Bai

The Bai language, spoken in South China, in Yunnan, is considered by some as a Lolo-Burmese language, and by others as forming an independent subgroup. Modern Bai uses a full decimal system. But a study of ancient shell-money reveals a different way of counting (Fu and Xu 2002). There are no old Bai texts, but the area has been dominated by the Chinese since 300 BC, and Chinese historians have recorded information on the local population since remote times. Thus in texts dating from the 12th and 14th century AD, we find remarks like the following:

> "Silk and shells are used in trading. Large shells are similar to fingers. Sixteen shells make one man"

> "a *ba* (shell) is called *zhuang*. Four *zhuang* make a *shou* ('hand'). Four hands make one man, and five men make a *suo*".

How does this work? Shells which are used for trading are counted as if they were fingers in a system where the thumb is considered differently from the other fingers. So the old Bai hand had 4 fingers and a thumb rather than five fingers[10]. In this way the four limbs of a man have 4 fingers/toes multiplied by 4 limbs = sixteen 'shells' to one 'man'.

This is the first time we encounter a counting system based on 4 where the next round level is 16 rather than 20 (4 limbs x 5 fingers = one score, as in Deuri). Thus we could believe that old Bai exemplified a true base-four numeral system. But notice that we leave the powers of 4 after 16, since the next level up is one suo = 5 x 16 = 80; the third rank of the base, 4^3 = 64 is not recognized as a level in the system. So it seems that we arrive at sixteen as a step (having its own name) in old Bai exactly as we arrive at twenty in Deuri, by a simple multiplication by 4 limbs, and not by raising the base to the second power (a process which then could be repeated indefinitely). We have not encountered in Tibeto-Burman a true 4-base system, or a true quinary system (no language has been found with a simple name for 125 = 5^5).

6. Rare constructions

We have up to now tried to discover principles for the regular construction of numbers in Tibeto-Burman languages. There are also oddities, or the limited use of constructions which are more frequent in other parts of the world. Addition and multiplication are the most common ways of forming number-words, in Tibeto-Burman and the world over. Subtraction, addition at unexpected points in the number sequence, and even division (not to be confused with half-counting) can be sporadically encountered. They do not constitute regular number building tools.

Subtraction

Subtraction or back-counting is not as common as addition and multiplication world-wide, but is still much more frequent than over-counting. We have seen an example from Latin. Sanskrit also uses it: 19 is *(ek-) una-vimsati*, '(1) lacking (from) 20'; 39 is *una-čatvarimsat* 'lacking 40', where the number which is subtracted (one) is understood.

In Tibeto-Burman subtraction is extremely rare; when expressing a number by reference to an upper step in the gradation, over-counting is preferred. We find subtraction in Meithei (alias Manipuri), a language of the India-Burma border which Shafer classified in his mixed group of Kukish. There 8 is /ni-pan/ '2 lacks', 9 is /məpan/ '1 lacks' (understood 'from 10'). These forms are included in 18 /təra ni-pan/ '10.2-lacks' and 19 /təra mə-pan/ '10.1-lacks' (Chelliah 1997: 85).

Karbi (Mikir) also has complex expressions for the numbers from 7 to 9. Seven is expressed by an addition /throk-chi/ '6 (+) 1'; eight and nine by a subtraction from ten: 8 is /ne-r-kep/ (2-r-10), 9 is /chi-r-kep/ (1-r-10) (Jeyapaul 1987: 93).

Multiplication of non-base numbers

In the general case, a multiplier is applied to the base and its powers to build higher numbers. Sometimes multiplication applies to auxiliary bases or hierarchized groupings, as with the quartets of Bodo.

Multiplication applied to numbers which in no way qualify as bases is found in Kayah Li, a Karen language. Solnit quotes the following sequence (table 13).

[10] In modern Bai these units do not survive as such, but a form of the word for 'hand', under a different tone, is used as a classifier for objects that go by four (the boards of a coffin, or lines of some types of song).

Table 13. Numbers in Kayah Li (Karen)

1	tə-			
2	nʌ̄			
3	sō	6	sō swá	(3 twice)
		7	sō swá tə-	(3 twice 1)
4	lwī	8	lwī swá	(4 twice)
		9	lwī swá tə-	(4 twice 1)
5	ŋɛ̄	10	chʌ̄	

6 and 8 are formed by doubling 3 and 4; 7 and 9 by adding 1 to 6 and 8 respectively. Solnit interprets this construction as a refection for forgotten numbers (Solnit 1996).

Addition at odd points in the sequence of numbers

We have just seen in Kayah Li how 7 and 9 are built by the addition of 1 to an already awkward number name. In Yakkha too, we saw ‘four’ expressed as ‘three plus one’.

We find such unexpected structures in several languages when an old number name has been dropped for some reason or other. The number ‘seven’, etymologically PTB *s-nyis, which is often at risk of becoming confused with ‘two’, PTB *g-nyis, through phonological evolution, is a good candidate for refection. In Mikir (alias Karbi) for instance, ‘seven’ is /throk-chi/ ‘6 (+) 1’ (Jeyapaul 1987: 87).

Division

Division of a number is an extremely rare way to build a number name. It is found in Sanskrit *śat-ardha* (100-half) ‘fifty’, ardha-pančasat (half-50) ‘twenty-five’ (Menninger 1969: 78). In Meithei, we find a form /yaŋkhə́y/ ‘fifty’, in which Chelliah recognizes the root /kháy/ ‘to divide’, and which she interprets as ‘half of 100’. Whereas the half-count, as we have seen, is a well established, regular way of building numbers, division, when attested at all, concerns only a single number, and never seems to be a regular way of building numbers in a system.

7. Conclusion

The Tibeto-Burman languages we have examined have allowed us to broaden the known areas of the world where less common forms of number-building can be found. Thus overcounting, deemed to have once been prevalent in two areas of the world, the Germanic North of Europe, and ancient Mexico, plus the lone Ainu, an isolated language of Japan, has been found to be relatively common in Tibeto-Burman.

We have been able to enrich the types of attested unsusual mechanisms. Menninger mentions the half-count (e.g. German *anderthalb* ‘one and a half, lit. ‘half of the other’); in Dzongkha we have found a three-quarter-count.

It has seemed to us that old structures still recorded in the 19th century have disappeared. But is this always the case? After all, at the end of the 20th century, H. Plaisier found a fractional count in Lepcha, which Mainwaring had missed a century earlier. Michailovsky discovered in 2005-6 a village where Hayu speakers used forms that we had thought extinct thirty years before in a different village. All hope of recording rare structures before they disappear may not be completely lost.

It should be remembered that number systems are most often very poorly reported. Linguistic questionnaires, when they include numbers at all, are partisan to the decimal system. They always include the numbers up to ten, and often up to twenty. For higher numbers they list the expected steps of the decimal system: 10, 100, 1000. But to detect other possible bases we need to record at least the first next step of all possible systems: 16, just in case the system had some hidden base-4 principle, 25, for possible quinary systems, 400, for the vigesimal, 144, for the duodecimal. But these round numbers are not enough; before reaching 30 and 35 in Dzongkha who could have imagined the presence of a fractional expression?

When a regular formation of the familiar multiplication and addition type is detected, the researcher tends to assume he has understood the whole system. As we have seen, even if a fully regular decimal system is used in a language, there may exist parallel whole systems, as in Dzongkha, or parallel partial systems.

A proper typology of number systems needs more data. Time is running against us. Concerning Tibeto-Burman languages for which Northeast India is sometimes considered as a center of dispersion, it may be hoped that the new linguistic survey of India which is getting under way will provide the detailed observations which are so cruelly lacking on this fascinating aspect of human cognition.

References

Anonymous, 1977, *An introduction to Dzongkha*, New Delhi, printed at Jayyed Press, Delhi, 102 p.

BENEDICT, Paul K., 1942, Thai, Kadai and Indonesian: a new alignment in Southeastern Asia, *American Anthropologist* 44: 576-601; repr. In *Austro-Thai*, New Haven: HRAF Press, 1971, pp. 438-463.

BHATTACHARYA Pramod Chandra, 1977, *A descriptive analysis of the Boro language*, Gauhati (India), Gauhati University, 380 p.

BURLING, Robbins, 2003,. The Tibeto-Burman languages of Northeastern India. In *The Sino-Tibetan languages*, Thurgood, Graham & Randy LaPolla (eds). 169-191. London and New York: Routledge (Routledge Language Family Series).

—, 2004. *The language of the Modhupur Mandi (Garo). Vol 1: Grammar*. New Delhi: Bibliophile South Asia & Promilla, 406 p.

BYRNE St Quintin, 1909, *A colloquial grammar of the Bhutanese language*, Allahabad, India, The Pioneer Press, 72 p.

CAUGHLEY Ross C., 1972, *A Vocabulary of the Chepang Language*, Summer Institute of Linguistics

—, 1989, Chepang - a Sino-Tibetan language with a duodecimal numeral base?, *in* D. Bradley *et al.* (eds.), *Prosodic Analysis and Asian Linguistics: to honour R. K. Sprigg*, Canberra, Pacific Linguistics, pp. 197-199.

—, 2000, *Dictionary of Chepang, a Tibeto-Burman language of Nepal*, Canberra (Australia), Australian National University («Pacific Linguistics»), 540 p.

CHAKRABORTI S. K. [alias Chakravarti], 1971, Reflections on Tipra numerals, *Pines (Shillong, India)* 1/1, pp. 74-76.

CHELLIAH Shobhana L., 1997, *A Grammar of Meithei*, Berlin, New York, Mouton de Gruyter («Mouton grammar Library 17»), xxv-539 p.

CLARK Mrs. E. W. [Mary M.], 1893, *Ao Naga Grammar with illustrative phrases and vocabulary*, Shillong (India), Assam Secretariat Printing Office, 1 vol., 181 p.

EGLI-RODUNER Susanna, 1987, *Handbook of the Sharchhokpa-lo / Tshangla (language of the people of Eastern Bhutan)*, Thimphu, Bhutan, Helvetas (Swiss Association for Development and Cooperation), 353 p.

FISHER, James F., 1971, A vocabulary of the Kaike language, mimeograph, Kathmandu.

FU Jingqi & XU Lin, 2002, *"Coquillage" en tant que monnaie, mot emprunté au chinois ancien en langue bai* («Seizièmes Journées de Linguistique d'Asie Orientale», Paris).

GOWDA K. S. Gurubasave, 1975, *Ao Grammar*, Mysore (India), C.I.I.L.

GVOZDANOVIC Jadranka, 1985, *Language System and its Change. On Theory and Testability*, Berlin, New-York, Mouton de Gruyter («Trends in Linguistics, Studies and Monographs 30»), 221 p.

HALE Austin (ed.), 1973, *Clause, Sentence and Discourse Patterns in selected languages of Nepal*, Norman, Oklahoma, SIL («SIL publications in Linguistics and related fields»), vol. 4, 314 p.

HARI, Maria & Chhegu LAMA. 2004. *Dictionary Yohlmo-Nepali-English*. Kathmandu: Central Department of Linguistics, Tribhuvan University, 848 p.

HODGSON Brian Houghton, 1857, Comparative vocabulary of the broken tribes of Nepal, *JASB* 26/5, pp. 317-349.

—, 1880, *Miscellaneous essays relating to Indian subjects*, London, Trübner & Co, reprint 1992 Delhi, Asian Educational Services, 2 vol.

HOFRENNING Rev. Ralph W., 1959, First Bhutanese Grammar, mimeograph.

HUTTON John Henry, 1916, *Rudimentary Grammar of the Sema Naga Language, with vocabulary,* Shillong (India), 95 p.

JÄSCHKE H. A., 1881 [1968], *Tibetan-English Dictionary*, London, Routledge et Kegan Paul, 671 p.

JEYAPAUL, V. Y. 1987. *Karbi grammar*. (CIIL Grammar Series 15). Manasagangotri, Mysore (India): CIIL, 171 p.

MAINWARING Colonel G. B., 1876 [1971], *A Grammar of the Rong (Lepcha) Language*, New Delhi, Manjusri («Bibliotheca Himalayica»), 146 p.

MATISOFF James A., 1997, Sino-Tibetan Numeral Systems: Prefixes, Protoforms and Problems, *Pacific Linguistics, series B-114*.

MAZAUDON Martine, 1982, *Dzongkha Numerals* («15th Interbnational Conference on Sino-Tibetan Languages and Linguistics », Beijing, China).

—, 1985, Dzongkha number systems, *in* S. Ratanakul *et al.* (eds.), *Southeast Asian Linguistic Studies presented to André-G. Haudricourt*, Bangkok, Mahidol University, pp. 124-157.

—, 2002. Les principes de construction du nombre dans les langues tibéto-birmanes. In *La Pluralité*, François, Jacques (ed.). 91-119. Paris: SLP (Mémoires de la Société de Linguistique de Paris. Nouvelle série tome 12). [http://halshs.archives-ouvertes.fr/halshs-00166891/fr/]

—, 2003, Tamang, *in* G. Thurgood et R. LaPolla (eds.), *Sino-Tibetan Languages*, London and New York: Routledge: 291-314.

MENNINGER Karl, 1969, *Number words and number symbols. A cultural history of numbers*, Cambridge, Mass. and London, England, M.I.T. Press, 480 p. [translated by Paul Broneer from the revised German edition *Zahlwort und Ziffer*, 1958, Vandenhoeck and Ruprecht, Göttingen, Germany].

MICHAILOVSKY Boyd, 1988, *La langue hayu*, Paris, Editions du CNRS («Sciences du Langage»), 234 p.

—, 1994, Bhutan: Language Situation, *in* R. E. Asher (ed.) *The Encyclopedia of Language and Linguistics* 1, Oxford, Pergamon Press, pp. 339-340.

PLAISIER, Heleen. 2006. *A grammar of Lepcha*. PhD. University of Leiden, 226 p.

SHAFER Robert, 1955, Classification of the Sino-Tibetan languages, *Word* 11/1, pp. 94-111.

SOLNIT David, 1996, *Eastern Kayah Li*, Hawaii, University of Hawaii Press, 385 p.

SPRIGG Richard K., 2002, *Balti-English English-Balti Dictionary*, Londres, Routledge-Curzon, 259 p.

SREEDHAR M. V., 1980, *Sema Grammar*, Mysore (Inde), Central Institute of Indian Languages («CIIL Grammar Series N°7»), 195 p.

STRAHM, Esther & Anita Maibaum. 2005. *Jirel-Nepali-English Dictionary*. Kirtipur, Nepal: CNAS, Tribhuvan University, 869 p.

THURGOOD, Graham & Randy LaPolla eds). 2003. *The Sino-Tibetan Languages*. London: Routledge (Routledge Language Family Series). 727 p.

VAN DRIEM George, 1998, *Dzongkha*, Leiden, School of Asian, African and Amerindian Studies («Languages of the Greater Himalayan Region, 1»), 489 p.

WEERA Ostapirat, 1999, Half and plus is minus, *LTBA* 22/1, pp. 201-202.