Classifiers before numerals

Origin of an atypical morpheme order

Jonathan P. Evans

Institute of Linguistics, Academia Sinica

Languages with sortal classifiers (CLF) are distributed across the world, with a large concentration in East and South (east) Asia. Across the world's classifier languages, few are attested which order classifiers before numerals in counted noun phrases (CLF NUM). The Sino-Tibetan language family includes languages without sortal classifiers, languages with the more typical NUM CLF order, as well as languages with CLF NUM order. The latter group of languages are concentrated in Northeast India and neighboring regions. The Sino-Tibetan languages with CLF NUM order do not fit under one genealogical node. The present hypothesis is that the classifiers arose through repetition of the counted noun, following which the order spread via contact. The CLF NUM order appears to have arisen at least as early as the time of Proto-Bodo-Garo. However, we do not yet find cognate classifiers at a time depth beyond Proto-Bodo-Garo. A remaining mystery is why such a cross-linguistically rare morpheme order spread so easily within the Northeast India linguistic area.

1. Introduction

In a survey of 400 languages, Gil (2013) identified sixty-two in which numeral classifiers (also sortal classifiers, CLF)[1] are optional and seventy-eight languages in which numeral classifiers are obligatory. Both types of languages are more concentrated in South, Southeast, and East Asia than in other locations around the world (Gil 2013; Her et al. 2019). Mandarin Chinese presents a well-documented case of an obligatory classifier language; every counted noun is preceded by either a sortal or a mensural classifier (1).

---

1. Where possible, Leipzig glossing conventions have been followed. A list of other abbreviations occurs before the references.

(1)   Sortal (1a) and mensural (1b) classifiers in Mandarin Chinese
   a.   **san   li   mi**
        three   CLF:kernel   raw.rice
        'three grains of rice'
   b.   **si   jin   mi**
        four   CLF:catty   raw.rice
        'four catties of rice'

Of the six logically possible orderings of the three classes of morphemes/words, the two combinations in which the numeral immediately precedes the classifier (NUM CLF N, N NUM CLF) are crosslinguistically common. Meanwhile, the two combinations in which the classifier immediately precedes the numeral (CLF NUM N, N CLF NUM) have been documented in only a few languages. Outside of Sino-Tibetan, the CLF NUM order is reported in Ibibio (Niger-Congo: Greenberg 1990 [1972]) and Warekena (North Arawak, Brazil, and Venezuela: Aikhenvald 2000: 109). Sequences that interpose the noun between the numeral and the classifier are not attested (*CLF N NUM, *NUM N CLF); cf. Aikhenvald 2000, Greenberg 1990 [1972], Her et al. 2019). Table 1 lists groups of Sino-Tibetan languages according to the structure of sortal classifiers within their languages.

Table 1. Presence, absence, ordering of classifiers and numerals across Sino-Tibetan languages

CLF NUM order	NUM CLF order	No evidence for CLF
Bodo-Garo group	Dhimal	Central Naga group
Northern Naga group	Newar (minimal)	Angami-Pochuri group
Jinghpo-Luish group	Kiranti group (Camling, Belhare; Hayu low numbers)	Zeme group
Kuki-Chin group	Puroik/Sulong	Tangkhulic group
Karbi/Mikir	Tangut-Qiang group (not all rGyalrongic)	Meithei
Tani group	Nungic group	Mru
Tawrã-Idu group	Tujia	Tibeto-Kanauri group (except Dhimal)
Kman	Lolo-Burmese group	Kham-Magar-Chepang group
	Naxi	
	Karenic group	
	Bai	
	Sinitic group	

Table 1 shows that Sino-Tibetan languages with CLF NUM order are concentrated in Northeast India and its environs. Languages with NUM CLF order occur both to the west and east of the CLF-NUM languages. Languages with no evidence for sortal classifiers occur to the west of the area where the order CLF NUM occurs. Various

higher-level groupings of the language groups in Table 1 have been proposed. Of these, the Sal grouping (next section) seems to be the least controversial. The groups and group names given here are not completely without controversy, but seem to represent widely agreed-upon classifications nonetheless (Bradley 2002, 2012; DeLancey 2015, in press; Matisoff 2003; among others).

Section 2 documents the presence of CLF NUM in the languages and groups found in the first column of Table 1. Section 2 also notes that among the languages of the Bodo-Garo branch the CLF-NUM complex can occur either before or after its noun, with a concomitant change in definiteness. Section 3 explores possible pathways of origination of this sequence. Section 4 considers possible spread westward into medieval Newar of the CLF NUM ordering. Conclusions are presented in Section 5.

2. Documentation of classifier numeral order

In this section, I document the CLF NUM order in the languages and groups identified in the first column of Table 1. The first three language groups in Table 1 (Bodo-Garo, Northern Naga, Jingpho-Luish) have been grouped together by Benedict (1972: 6–7), Burling (1983) and Thurgood (2003), into a subgroup that Burling termed 'Sal'. The group name is based on a proposed shared innovation of the protoform *sal 'sun'; other shared innovative lexemes have been proposed. Bradley (1997: 20–27) maintains a Bodo-Garo – Northern Naga subgroup within Sal. Similarly, Matisoff (2013) supports a close relationship between Bodo-Garo and Northern Naga, and also gives data to show that Jingpho-Luish may be closer to Northern Naga than to Bodo-Garo. A simplified Sal family grouping is presented in Figure 1, with connecting lines omitted, due to conflicting claims about subgrouping. Furthermore, Bradley (1997) proposes that Kuki-Chin and Pyu could be closely related to Sal.

<div align="center">

– Bodo-Garo. (All attested languages)

Sal – Northern Naga. (Chang, Thangsa)

– Jingpho-Luish. (Jingpo, Turung, Kadu)

</div>

Figure 1. Sal node within Sino-Tibetan. Named languages demonstrate CLF-NUM order

The discussion begins with Bodo-Garo languages because the phenomenon has been more widely documented within this group. Furthermore, the semantic changes that occur when the CLF-NUM is located before vs after the counted noun have also been documented, although unevenly throughout the group.

2.1   CLF NUM order in Bodo-Garo languages

Among languages of the Bodo-Garo branch the sequence CLF-NUM is the only grammatical arrangement of classifiers and numerals, as exemplified in Table 2 and Example (2). This order has been reconstructed for Proto-Bodo-Garo (Wood 2011).

Table 2. Classifier Numeral combinations across Bodo-Garo languages

Dimasa	Bodo	Kokborok	Garo	Rabha	
sau-si	sa-se	kʰoɾok-sa	sak-sa	sak-sa	CLF:person-one
pʰaŋ-si	pʰaŋ-se	pʰaŋ-sa	pʰaŋ-sa	pʰaŋ-sa	CLF:plant-one

(2) Classifier-Numeral as word in Dimasa[2]
<u>homao</u> **pher-shi**=ha   **tho-dan**[2]
rice.cake CLF:flat-one=LOC CLF:rupees-fifty
'One rice cake (costs) 50 rupees.'

Within the Bodo-Garo literature it has been noted that the compound CLF-NUM can occur either before or after the noun that is counted, as in Rabha (3).

(3) Rabha examples of N CLF-NUM and CLF-NUM N   (Joseph 2007)
   a. aŋ-i   <u>másu</u> **maŋ-aníŋ** to-a
       1SG-GEN cattle CLF-two   there.be-PRS
       'I have two cattle.'
   b. <u>caŋ-ba</u>   **sak-sa**   <u>kai</u>   riba-eta
       who-INDF CLF-one person   come-CONT
       'Someone (some one person) is coming.'

In at least some Bodo-Garo languages, the difference in order signals a change in definiteness, in which the order N CLF-NUM is definite (4) and CLF-NUM N is indefinite (5). Unless noted otherwise, all Dimasa examples come from Dhrubajit Langthasa (p.c.).

(4) N CLF-NUM order in Dimasa (definite)
   <u>subung</u> **sau-shi** *phai-yaba-de*   *cha ling-ya*
   man   CLF-one come-NOMZ-TOP tea drink-NEG
   'One of the people who came won't have tea.'

---

(5) CLF-NUM N order in Dimasa (indefinite)
sau-shi   _subung_ *phai-ba*
CLF-one   man   come-PST
'A man came.'

A similar property is found in some Tai languages in which the classifier can be positioned either before or after the noun, as in Standard Thai (6) (Noss 1964: 107).

(6) CLF-N order in Standard Thai
a. **sǎam saan:** _níi_
three CLF   dish
'these three dishes (of food)'
b. <u>kɛɛŋ</u> **sǎam saan:**
curry three CLF
'three dishes of curry'

The word order N CLF-NUM is reported to be the default order in counted noun phrases across Bodo-Garo, occurring more frequently than the order CLF-NUM N. This distributional asymmetry has been noted in Rabha (Joseph 2007), Bodo (Basumatary 2015), Garo (Burling 1961), Kokborok (Debbarma 2015) Atong (van Brueghel 2014), and Dimasa (Evans & Langthasa ms.). The Bodo-Garo pattern of N CLF-NUM may be exemplified with examples from Dimasa, in which the sequence N CLF-NUM can occur with agents (7), patients (8), and oblique arguments (9).

(7) N CLF-NUM as agent noun phrase
<u>*Asampa-tu*</u> **g-tham**   *jang malai ni   no   ha   jinis*
PN-PL   CLF:human-three COM others GEN house LOC things
khao-hi-ma   *thang-ba*
steal-TRNS.LOC-PRSP go-NF:IPFV
'Three people, Asampa and them, together went to someone's place to steal things.'   (Evans & Langthasa ms.)

In (8), /gni/ {gini} 'two:human' is a fusion of the human classifier /g-/ and /gni/ 'two'.

(8) N CLF-NUM as topicalized agent and as patient noun phrase
*Asampa jang bu ni   <u>lugu</u> **gini**   bo <u>bosta</u>   **ma-shi**
PN   COM 3SG GEN friend two:human TOP sack (IE) CLF-one
ma-shi *la   hi   bu bising ha   thim-kha*
CLF-one take SUCC DEM inside LOC hide-PRF
'Asampa and his two friends each took one sack and hid inside it.'
                                        (Evans & Langthasa ms.)

In (8) the sequence N CLF-NUM occurs inside the topicalized agent 'two friends' as well as within the patient 'one sack each'. Although *lugu gini* specifies 'two friends', there is no plural marking on *lugu*. It is not yet clear why in some cases, human plural nouns are marked for number, but in other instances they are not.

(9)   N CLF-NUM as oblique noun phrase (locative)
    ibu   *jila*   **ma-gin** *ha* *hajing* *mai* *sibling*   *ode* *khun*
    PRX.DEM district (IE) CLF-two LOC ginger rice black sesame and thread
    bo   *mai-du*
    TOP get-PIPFV
    'Ginger, rice, black sesame and thread are also obtained in these two districts.'
                                                            (Evans & Langthasa ms.)

A less frequent pattern puts the CLF-NUM word before the noun that is counted. This kind of NP can, like the aforementioned N CLF-NUM, occur as an agent (10), patient (11) or oblique (12).

(10)   CLF-NUM N as agent noun phrase
     odehe *bu* *gorai* *jang* *mar*   *dada* *ar* **ma-shni** *gorai* *phai*
     and   DEM horse COM take.a.liking NF   and CLF-seven horse come
     pa-ba-kho   *ne*
     APPL-NF:IPFV-DPROG QUOT
     'And it is said that seven more horses came along with that horse.'
                                                          (Evans & Langthasa ms.)

(11)   CLF-NUM N as patient noun phrase
     ma-shi *samosh*   *ling-kha*   *de*   *hang* *la-ba*   *ha*
     CLF-one spoon (IE) drink-PRF TOP breath take-NMLZ LOC
     sa-ya-dao-du
     hurt-NEG-CMP-PIPFV
     '…, then drinking one spoon of lemon juice can make the pain less while breathing.'   (Evans & Langthasa ms.)

(12)   CLF-NUM N as oblique noun phrase (locative)
     ma-shi *station* *ha*   *train* *mathao-ba*   *ha*   *Asampa* **sao-shi**
     CLF-one station LOC train stop-NF:IPFV LOC PN   CLF:human-one
     Hadisa *khe* *Hindi* *grao* *jang* *sengkha* …
     Bengali ACC Hindi word INS ask-PRF
     'When the train stopped at one of the stations, Asampa asked one Bengali person in Hindi…'   (Evans & Langthasa ms.)

Counted nouns can occur in noun phrases with modification, as in (13).

(13) Modified counted noun (Dimasa)
      ibu **ma-gin** *gisim* <u>*daono*</u> *ani.*
      DEM CLF-two black chicken 1SG.GEN
      'These two black chickens are mine.'

In Dimasa, classifiers do not occur within non-counted deictic expressions, unlike
Mandarin Chinese and other classifier languages (14), (15).

(14) DEM CLF N in Mandarin Chinese
      zhe **zhi** <u>gou</u> shi wo de
      DEM CLF dog be 1st GEN
      'This dog is mine.'

(15) DEM N in Dimasa
      ibu   <u>*sisha*</u> *jini*   *she*
      PRX.DEM dog 1PL.GEN TOP
      'This dog is ours.'

Although numeral classifiers are bound morphemes, and never occur without a
numeral, there are various scenarios in which a numeral can occur without a clas-
sifier, such as round numerals as estimates, as well as numerals below ten, especially
when counting humans. In (16), we find the NP structure NUM NUM N, while in
(17) NUM NUM followed by the comitative marker conveys the adverbial sense of 'by
the (round) NUM'. In sentences with this structure, identity of the noun is specified
outside of the NUM NUM noun phrase. In (18), the counting of a few humans does
not require a classifier.

(16) Use of NUM NUM N (*CLF) for round number approximation
      bu **raja**   **raja**   <u>*meshep*</u> *khrang-ba*
      3SG hundred hundred buffalo raise-NF:IPFV
      'He reared hundreds and hundreds of buffalo.'   (16-Kalachand ni Jarmin)

(17) Use of NUM NUM COM for 'by the (round) NUM'
      mel *ha* **rzing**   **rzing**   *jang* <u>*subung*</u> *phaiba*
      meet LOC thousand thousand COM person come-NF:IPFV
      'People came to the meeting by the thousands.'

(18) Use of N NUM (*CLF) in counting a few humans
      ibu   *no*   *ha* **bonwa** <u>*subung*</u> *dong-phu-du*
      PRX.DEM house LOC five   person exist-POT-IPFV
      'Five people can live in this house.'

Dimasa speakers also employ CLF-NUM words with no overt noun. These CLF-NUM
words can occur as a noun phrase when the speaker perceives that the listener can

recover the referent of the noun without overt marking, as seen in (19), (20). In (19), the CLF-NUM compound *mashni* functions semantically as 'seven (horses)', with the 'horse' meaning recoverable from the previous clause.

(19)   Sentence-internal referent indicated by CLF-NUM (*N)
   *odehe bu   hadzi   dang-ya-sa   khe thi-kha ne   **ma-shi**
   and   DEM paddy.field work-AGT-person ACC say-PRF QUOT CLF-one
   *gorai gma-jao-ba-ha   **ma-shni**   mai sing-kha   na ning de?*
   horse disappear-MAL-NMLZ-LOC CLF-seven get inward-PRF Q 2SG TOP
   And told the farmer, 'Losing one horse you got seven more (horses)?'
                                                          (Evans & Langthasa ms.)

In (20), the reference to lemon inferred by *khepshi* 'one slice of' is recoverable from the earlier portion of the text. Likewise, there is no noun 'times' overtly expressed with *phintham* 'thrice'.

(20)   Sentence-external referent encoded by CLF-NUM (*N) (mark both CLF-NUM)
   ***khep-shi*   *jang ham-ya   thikha bo   phrong   sainjer sainbli*
   CLF:slice-one INS   good-NEG if   TOP morning noon   evening
   ***phin-tham**   *ling-kha   ning ham-du*
   CLF:repeat-three drink-PRF TOP good-PIPFV
   'If after once it doesn't improve, drinking it three times in morning, noon and evening, it will be good.'   (Evans & Langthasa ms.)

This type of structure often occurs in texts when the referent is recoverable. In some cases, CLF-NUM combinations have become lexicalized, as in *baw-si* CLF:thought-one 'the same' (21) and *ma-si* CLF:generic-one 'exactly the same' (22).

(21)   Use of lexicalized *bawsi*
   *Rahul=zaŋ Arup=ni daŋ-taj   **baw-si***
   PN=COM₁   PN=GEN work-NMLZ CLF:thought-one
   'Rahul and Arup's working (behaviors) are the same.' (Evans & Langthasa ms.)

(22)   Use of lexicalized *masi*
   *Rahul=zaŋ Arup=ni daŋ-taj   **ma-si***
   PN=COM1   PN=GEN work-NMLZ CLF:generic-one
   'Rahul and Arup's working (behaviors) are exactly the same.'
                                                          (Evans & Langthasa ms.)

In (21) and (22), lexicalized CLF-NUM functions as a predicate, or as the complement of a verbless clause.[3] Non-lexicalized CLF-NUM can also function as a predicate (23).

---

3.   I would like to thank an anonymous reviewer for this observation.

(23)   Predicate use of CLF-NUM (*N)
       homao   ***pher-shi**=ha*   ***tho-dan***
       rice.cake   CLF:flat-one=LOC   CLF:rupees-fifty
       'One rice cake (costs) 50 rupees.'   (Evans & Langthasa ms.)

Given that the number of rupees varies across transactions, the predicate *thodan* (23) is considered not to be lexicalized.

To summarize, in Bodo-Garo languages, counted nouns typically contain a word in which a classifier is prefixed to a numeral. The CLF-NUM word can occur either before or after its head noun, depending on language-specific properties, especially definiteness.

2.2   CLF NUM order in Northern Naga languages

We look next at the presence of classifier-numeral units within Northern Naga languages, as exemplified by Chang Naga, Hakhun Tangsa, and Muklom Tangsa. In all three langauges, the CLF NUM compound follows the counted noun. Chor's (2011) analysis of Chang Naga posits CLF and NUM as free morphemes, while Hakhun and Muklom Tangsa are described with CLF-NUM forming a single word. The enumerating of nouns in Chang Naga is exemplified in (24).

(24)   Noun counting in Chang Naga   (Chor 2011: 20)
       púhèk **pém sə̀m**
       fruit   CLF   three
       'three pieces of fruit'

The Hakhun Tangsa N CLF-NUM pattern is shown in (25).

(25)   Hakhun Tangsa N CLF-NUM examples   (Boro 2017: 119, 120)
       a.   miʔ   **và-ni**   kə́mə́ kûmpô nə́m   t-aʔ
            person   CLF-two   ERG   money borrow PST-3
            'Two people borrowed money (from her).'   [MOV-2-1.86]
       b.   cûŋpù   **pù-tʰeʔ**
            flower   CLF-one
            'one flower'

The Hakhun Tangsa classifier /và/ in (24a) is a morpheme meaning 'male human' and /pù/ is the last syllable of 'flower' (24b).

A similar pattern may be observed in the Northern Naga language Muklom Tangsa, where the classifier-numeral complex follows the noun (26).

(26) Noun counting in Muklom Tangsa   (Mulder 2020: 191–192)
      ŋaʔ⁰ liŋ¹-təm²
      fish CLF-three
      'three lengthy fishes'
       βaʔ⁰   kʰɔ¹-ʃɛ¹
      bamboo CLF-one
      'one piece of bamboo, cut lengthwise'

Although there is not a lot of grammatical information on Northern Naga languages, nevertheless, we find that at least three languages in the group demonstrate the order of placing the classifier before the numeral, and also placing the classifier-numeral word/phrase after the modified noun.

2.3   CLF NUM order in Jingpho-Luish languages

The third branch of the Sal subfamily is Jingpho-Luish. The morpheme order N CLF NUM is attested in languages of both branches: Jingpho (e.g. Jingpho, Singpho) and Luish (e.g. Kadu). For Jingpho, Cheung (2007) shows that sortal classifiers are optional, while mensural classifiers are not (27). The analyses of Cheung (2007), Dai (2012) and Dai & Xu (1992) present classifiers as independent morphemes (cf. also Kurabe 2016).

(27) Jingpho examples of N CLF NUM order   (Cheung 2007: 34)
      a.   Sortal classifier (optional)
           sara   **(marai) masum**
           teacher CLF   three
           'three teachers'
      b.   Mensural classifier (mandatory)
           n-gu   *(sing) shi kru
           raw.rice CLF:liter ten six
           'sixteen liters of raw rice'

In Turung, a variety of Singpho, classifiers are somewhat marginal, there being only two sortal classifiers, and the application of these being optional (28) (Morey 2010: 325–331). As in Bodo-Garo, two orders are attested: N CLF NUM and CLF NUM N.

(28) Optional application of classifier in Turung   (Morey 2010: 325)
      a   *guila ngkhong numnang* … *numnang ngkhong nga*
      EXCL CLF   two   friend   … friend   two   have
      'There were two friends, … two friends.'

In the Luish language Kadu, classifiers, which are bound, precede the numeral one, but follow higher numerals (29).

(29)   Variable CLF NUM order in Kadu   (Sangdong 2012: 286)
       hú-à   **kalìng-hú**   **sóm-hú**   an-yeù-sà   nāng
       CLF.person-one   two-CLF.person   three-CLF.person   DEM-manner-DIM   go
       =mā   =ták
       =RLS   =HS
       'A couple of people like this went (there), it is said.'

The Kadu classifier-numeral word follows its head noun: /kasà nú-à/ tiger CLF-one
'one tiger' (Sangdong 2012: 288).

To summarize, within the Jingpho-Luish branch of Tibeto-Burman, the order
N CLF NUM is attested in both Jingpho and Luish languages. Thus, Jingpho-Luish,
Northern Naga and Bodo-Garo languages all provide support for the existence of
this unusual morpheme within the Sal family, although I am not aware of recon-
structions of classifiers that date back to proto-Sal.

2.4   CLF NUM order in Kuki-Chin languages

The same morpheme ordering has been documented in at least two Kuki-Chin
languages, a group of languages that Bradley (1997: 20) proposes may be related to
the Sal group. The Lutuv, or Lautu, language of the Maraic group exemplifies this
order with bound classifiers (30).

(30)   N CLF-NUM order in Lutuv   (Berkson ms.)
       a.   <u>cawsaa</u> **pa-hning**
             human   CLF-two
             'two humans'
       b.   <u>thuo</u>   **thluo-thung**
             mountain   CLF-three
             'three mountains'

VanBik (2009) identifies Maraic as an independent branch within Kuki-Chin, while
Peterson (2017) groups Maraic within Central Kuki-Chin. Hakha Lai, of VanBik's
(2009) Central group and Peterson's (2017) Core Central group, also demonstrates
N CLF-NUM order (31).

(31)   N CLF-NUM order in Hakha Lai   (James Wamsley p.c.)
       <u>chizawh</u> **pa-hnih**   nih   nga   an-ei
       cat   CL-two   ERG   fish   3PL-eat
       'Two cats are eating fish.'

2.5   CLF NUM order in Karbi/Mikir

The same order of morphemes is found in the Karbi (Mikir) language (32), which is geographically close to Kuki-Chin, but whose affiliation within Tibeto-Burman is unclear. The (pre-) history of the language indicates the effect of contact with Tibeto-Burman languages, as well as with the Austroasiatic Khasi language.

(32) CLF-NUM order in Karbi   (Konnerth 2014: 140)

… Bey a-tum   <u>korte</u>   **bang-kethom**   …
title POSS-PL brother CLF:HUM:PL-three
'… three Bey brothers…'

2.6   CLF-NUM order in Tani languages

CLF-NUM order has also been documented in the Tani languages, which are mainly spoken in central Arunachal Pradesh (Post & Sun 2017). The Tani-speaking area is contiguous with Tawrã and Kman, in which the CLF NUM order is exemplified in subsequent sub-sections. Galo places the CLF-NUM complex after the counted noun (33).

(33) CLF-NUM order in Galo
<u>hɨɨdàa</u> **dáa-ɲì=go…**   kozzúu=ogò…   **adáa=go**   ɲûm
stick   CLF.stick-two=IND awhile.ago=ANAP.LOC CLF.stick=IND DLMT
túr-tó
support.from.below-PFV
'Two sticks (were propped against the door)/…a minute ago, only one stick had been propped against (the door).'   (Post 2007: 388)

Mising (Miri) is another Tani language in which the order N CLF-NUM has been attested. The phenomenon can be exemplified by Sayang Mising (34), the largest Mising dialect (Prasad 1991: 4).

(34) N CLF-NUM order in Mising
bui <u>po'rok</u> **pir-keng'-ka** mô-kê tô
he   fowls   CLF-six-INDF killed has
'He killed six fowls.'   (Needham 1886: 20)

2.7   CLF-NUM order in Tawrã-Idu languages

This cluster of (at least) two languages, spoken in the border areas of Arunachal Pradesh and Tibet, appears to be the most Northeastern extension of the CLF NUM order within Sino-Tibetan. These languages seem to be closely related to Tani (Prasad 1991: 2; Jackson Sun p.c.). In these languages, the sequence CLF NUM follows nouns (35), (36).

(35) N CLF NUM in Tawrã
<u>santara</u> **bra kasaŋ** sizana!
orange CLF three bring
'Bring three oranges!'   (Johakso Manyu p.c.)

The same pattern is found in Idu (36).

(36) N CLF NUM in Idu
*Atu   mra'a   <u>naru</u> **ka-ni**,   <u>naru</u> **ka-sō-ma**
up.there high.mountains range CLF-two, range CLF-three-loc
ayi-do-ho-ne   ichi-ge   echa api   tō=wa.
like.this-where-when-NF little.bit-one this bamboo.sp rod=top
'Far away in the high mountains, past two or three ranges, only then do you find this *api* bamboo.'   (Naomi Peck p.c.)

2.8   N CLF NUM order in Kman

Kman (Miju) is a Tibeto-Burman language of Arunachal Pradesh and Tibet. Together with Tawrã it forms an ethnic unit (Mishmi), but not a linguistic unit. Example (37) shows that sortal classifiers (/bɑt^{35}/ 'CLF.segment') follow the counted noun and precede the numeral.

(37) N CLF NUM in Kman
ɑn^{35} ɑn^{31}dzɑʊ^{35} <u>bloŋ^{35}</u> **bɑt^{35}**   **mu^{53}** ʊi^{53} ŋin^{55} kɯ^{31}sʌm^{53}
this mountain   road   CLF.segment one 3SG day   three
thɑi^{53} kɑ^{35}
walk   TAM
'He walked for three days on this one section of mountain road.' (Li 2002: 98)

The preceding sections demonstrate that classifiers precede numerals in a geographically focused group of Tibeto-Burman languages, all of which are spoken in Northeast India and its environs. The CLF NUM order does not appear to be attested in Sino-Tibetan languages that are spoken outside of this area.

3.   Possible path of historical development

The data presented above suggest that sortal classifiers developed at least as early as Proto-Bodo-Garo (Table 2, cf. Wood 2011). The CLF NUM order exists within various higher-level nodes, such as Sal (subsuming Bodo-Garo, Northern Naga, and Jingpho-Luish), as well as lower taxonomic levels as shown in Table 1. However, other than Proto-Bodo-Garo, there do not appear to be reconstructions of forms that date back to protolanguages. Thus, in terms of shared innovation, there is not enough evidence to posit a greater time depth than Proto-Bodo-Garo.

   Among the languages in question, the process of development appears to have started with repetition of nominal elements, often following the head noun in a counted noun phrase. In many cases, the repeated element appears to have been the second syllable in a disyllabic noun. In many languages, the repeated bound element was bound to the subsequent numeral. The order N CLF-NUM is still the most frequent ordering across most of the languages documented above. We also note that numerals following nouns is a common pattern in languages of this area (Burmese, Thai, etc.).

   Aikhenvald (2000: 103) documents the creation of numeral classifiers through repetition of a noun or a nominal element. This process has been observed in various Sino-Tibetan languages, such as Burmese and Chinese (38).

(38)   Repeated element as classifier
   q̲a̲k̲h̲a̲n̲: **hcau'-hkan:**   (Burmese)
   room   six-CLF
   'six rooms'   (Cornyn & Roop 1968: 228, cited in Dryer 2008)
   liu jian f̲a̲n̲g̲j̲i̲a̲n̲   (Mandarin Chinese)
   six CLF   room
   'six rooms'   (Jackson Sun p.c.)

Numeral classifiers of the "repeater" sort are common across Bodo-Garo languages, and are demonstrated here by Dimasa (39).

(39)   Classifiers from repetition   (Dimasa)
   a.   _di-phong_   **phong-shi**   _labu._
       water-container CLF:container-one   bring
       'Bring a container.'
   b.   _bokhong_ **khong-shi**   _khaikho_
       Packet   CLF:packet-one   remove
       'Take out a packet.'

For some semantic groupings, such as names for fruit, a frequently repeated bound element occurs at the beginning or end of many of the words in the set. For example,

numerous Dimasa words for fruit have *thai-* as the first syllable (40). However, some 'fruit' words have *-thai* as the final element, as in 'gooseberry' and 'haritaki'; also, not all fruit words contain this element ('orange', etc.). Wood (2008: 79–96) identifies about thirty such "class terms" in Proto-Bodo-Garo, including *thái for both 'fruit' and 'round' (2008: 82, 88).

(40) Dimasa fruit words with and without *thai*

thaiju	mango	*hamlaithai*	gooseberry
thaiplung	jackfruit	*silikathai*	haritaki
thaisa	lemon	*hondra*	orange
thailik	banana	*thintri*	tamarind
thaisum	cucumber	*laimuri*	pineapple

Some forms ending with *-thai* invoke a looser semantic connection with fruit; e.g. none of the items in (41) are actually fruit, although stones and eyes are prototypically round.

(41) Dimasa non-fruit words with *-thai*

muthai	eye
longthai	stone
grauthai	word

All three aforementioned classes of noun (fruit names with *thai*, fruit names without *thai*, non-fruit words with *thai*) require the sortal classifier *thai-* (42).

(42) Dimasa counting with *thai-*

    a.   Fruit name beginning with *thai-*
thaiju   **thai-shi**
mango   CLF:fruit-one
'one mango'

    b.   Fruit name ending with *-thai*.
hamlaithai   **thai-bri**
gooseberry   CLF:fruit-four
'four gooseberries'

    c.   Fruit name without *thai*
hondra   **thai-gin**
orange   CLF:fruit-two
'two oranges'

    d.   Semantic extension of fruit
muthai   **thai-gin**
eyes   CLF:fruit-two
'two eyes'

It appears that the tendency to locate CLF after its head noun was affected by postnominal location of the numeral. As noted in Section 1, arrangements of N, CLF, NUM in which the N is interposed between CLF and NUM are not attested. Thus, for classifier languages with the numeral placed after the noun, the only possible sequences are N NUM CLF and N CLF NUM. In grammars of Bodo-Garo languages, a distinction has been drawn between noun phrases in which CLF-NUM occurs after or before the head noun. For example, in Dimasa, the order N CLF-NUM can convey a topic comment structure with a meaning like, 'Of the N, CLF-NUM of them….' (cf. (43) vs (44)). Examples (4) and (5) showed that the change in word order can also signal a change in definiteness.

(43)   N CLF-NUM   (Dimasa)
    bu   <u>alu</u> **ma-gin**   na   ji-ba
    DEM cat CLF-two   fish eat-NF:IPFV
    'Those cats, two are eating fish.' (Lit.: Out of all the cats present, those two are eating fish.)

On the other hand, CLF-NUM N functions like a simple NP, as in (44).

(44)   CLF-NUM N   (Dimasa)
    bu   **ma-gin**   <u>alu</u> na   ji-ba
    DEM CLF-two   cat fish eat-NF:IPFV
    'The two cats (there are no other cats) are eating fish.'[4]

Despite the apparent added complexity of a topic-comment structure in Dimasa, the word order N CLF-NUM is reported to be the default order in counted noun phrases across Bodo-Garo, occurring more frequently than the order CLF-NUM N. This ordering asymmetry has been observed in Rabha (Joseph 2007), Bodo (Basumatary 2015), Garo (Burling 1961), Kokborok (Debbarma 2015), Dimasa (Evans & Langthasa ms.) and Atong (van Brueghel 2014).

---

4.   Thanks to Jackson Sun for asking about this.

4.   Potential spread of CLF-NUM order beyond this linguistic area

This typologically unusual order is claimed to have spread to Newar through contact with the Bodo-Garo language Mech (Kiryu 2009). The Mech (or Meche) population is currently centered in West Bengal, with communities in both Eastern Nepal and Assam. The earliest Newar texts, from the eleventh to the fourteenth century display a pattern in which animate nouns take classifiers, and where indefinite noun phrases display the order CLF NUM (45) and counted definite noun phrases have the order NUM CLF (46). In texts that are more recent than the 14th century, only the NUM CLF order is observed.

(45) CLF-NUM order in Medieval Newar indefinites   (Kiryu 2009)
    a. śikva saṃkśā **hmaṃ 50**
        died ?   CLF   50
        'All together 50 persons were killed.'   (GV41b.04)
    b. Doya mvaṇḍa   ḍyaṅā **hma 7** tyaṃkhodvala **hha 3**
        Doya beheaded cut.CP CLF 7 Tyankhodvala CLF 3
        'Seven Doyas were slaughtered. Three were killed in Tyankhodvala.'
                                       (GV41b.02)

(46) CLF-NUM order in Medieval Newar definites   (Kiryu 2009)
    a. thva **ṅa-hmaṃ** mī-m
        this five-   CLF   people-ERG
        'these five people…'   (GV43a.05)
    b. gvāṭha   **ne-ma** bhvānʾa mesa   smasta lisyaṃ hayā
        cowherd 2-CLF ?   buffalo all   back   bring
        'the two herdsmen who brought back all the buffaloes…' (GV57a.02)

There is not enough historical data at present to prove that the typologically rare CLF-NUM order was borrowed from Mech into Newar. However, the geographic distribution of the two languages does overlap. Other than Newar and Meche, the languages spoken in the area lack numeral classifiers (Kansakar 2005; see also Weidert 1984).

5.   Conclusions and further implications

It has been noted since Greenberg 1990 [1972], that the morpheme orders N CLF-NUM and CLF-NUM N are rare across the world's languages. However, for the Tibeto-Burman languages of the Northeast India linguistic area, these are the only possible orders of noun, sortal classifier, and numeral. Repetition of bound nominal

elements as bound sortal prefixes to numerals has given rise to CLF-NUM words, which occur both before and after their head noun.

These languages are in contact with numerous languages that lack this order, so it is not yet clear what language-internal properties have led to the preservation of CLF-NUM in languages of this area. Perhaps psycholinguistic experiments can aid in uncovering factors that have contributed to its preservation.

Acknowledgements

Funding

Abbreviations

Abbreviations in cited examples are defined in their source materials.

CMP	comparative	PIPFV	present imperfective
CONT	continuous	PL	plural
DPROG	durative progressive	PN	personal name
EXCL	exclusive	PRS	present
INDF	indefinite	PRSP	prospective
IE	Indo-European (borrowed word)	PRX.DEM	proximal demonstrative
N	noun	SUCC	successive
N	nonfuture	TRNS.LOC	trans locative
NUM	numeral		

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices: A typology of noun categorization devices*. Oxford: Oxford University Press.

Basumatary, Guddu P. 2015. Numeral classifiers in Bodo. *Nepalese Linguistics* 30. 19–24. http://himalaya.socanth.cam.ac.uk/collections/journals/nepling/pdf/Nep_Ling_30.pdf#page=23

Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511753541

Berkson, Kelly. ms. Field report on classifiers in Lutuv. *Bloomington: Indiana Working Papers in South Asian Languages and Cultures*.

Boro, Krishna. 2017. A grammar of Hakhun Tangsa. Eugene: University of Oregon, Doctoral dissertation.

Bradley, David. 1997. Tibeto-Burman languages and classification. In Bradley, David (ed.), *Tibeto-Burman languages of the Himalayas, Papers in South East Asian linguistics* 14, 1–71. Canberra: Pacific Linguistics.

Bradley, David. 2002. The subgrouping of Tibeto-Burman. In Christopher Beckwith (ed.), *Medieval Tibeto-Burman languages*, 73–112. Leiden: Brill.

Bradley, David. 2012. Tibeto-Burman languages of China. In Rint Sybesma (ed.), *Encyclopedia of Chinese languages and linguistics*. Leiden: Brill.
https://doi.org/10.1163/2210-7363_ecll_COM_00000419

van Breugel, Seino. 2014. *A grammar of Atong*. Leiden: Brill. https://doi.org/10.1163/9789004258938

Burling, Robbins. 1961. *A Garo grammar* 25. Postgraduate and Research Institute, Deccan College: Pune.

Burling, Robbins. 1983. The Sal Languages. *Linguistics of the Tibeto-Burman Area* 7(2). 1–32.

Cheung, C. C. H. 2007. On the noun phrase structure of Jingpo. *Bloomington: Working papers in linguistics* 3. 32–56.

Chor, W. T. 2011. *A sketch grammar of Chang*. Final year project report, Nanyang Technological University, Singapore.

Cornyn, William Stewart & D. Haigh Roop. 1968. *Beginning Burmese*. New Haven: Yale University Press.

Dai, Qingxia & Xijian Xu. 1992. *Jingpoyu yufa* [A Jingpo grammar]. Beijing: Minzu University of China.

Dai, Qingxia. 2012. *Jingpoyu cankao yufa* [A Jingpo reference grammar]. Beijing: China Social Sciences Press.

Debbarma, Biman. 2015. Word order in Kokborok. Mysore. *Language in India* 15(5). 51–58.

DeLancey, Scott. 2015. Morphological evidence for a Central branch of Trans-Himalayan (Sino-Tibetan). *Cahiers de Linguistique–Asie Orientale* 44(2). 122–149.
https://doi.org/10.1163/19606028-00442p02

DeLancey, Scott. In press. Classifying Trans-Himalayan (Sino-Tibetan) languages. In Paul Sidwell & Mattias Jenny (eds.), *The languages and Linguistics of Mainland Southeast Asia*, 207–224. Berlin: De Gruyter. https://doi.org/10.1515/9783110558142-012

Evans, Jonathan P. & Dhrubajit Langthasa. ms. *Dimasa language: Structure and texts*.

Gil, David. 2013. Numeral classifiers. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/55, Accessed on 2021-03-18.)

Greenberg, Joseph. 1990. *On language: Selected writings of Joseph H. Greenberg*. Palo Alto: Stanford University Press.

Greenberg, Joseph. 1990 [1972]. Numeral classifiers and substantival number: Problems in the Genesis type. Working Papers in Language Universals, reprinted in Greenberg, 1990: 16–93.

Her, O. S., M. Tang & B. T. Li. 2019. Word order of numeral classifiers and numeral bases: Harmonization by multiplication. *STUF-Language Typology and Universals 72*(3). 421–452. https://doi.org/10.1515/stuf-2019-0017

Joseph, Umbavu V. 2007. *Rabha*. Leiden: Brill.

Kansakar, Tej Ratna. 2005. A historical development of the numeral classifier system in Newar. In Y. Yadav (ed.), *Contemporal issues in Nepalese linguistics*, 101–116. Kirtipur: Tribhuvan University.

Kiryu, K. 2009. *On the rise of the classifier system in Newar*. Osaka: *Senri ethnological studies 75*. 51–69.

Konnerth, Linda. 2014. A grammar of Karbi. Eugene, OR: University of Oregon PhD dissertation.

Kurabe, Keita. 2016. A grammar of Jinghpaw, from Northern Burma. Kyoto: Kyoto University, Ph.D. dissertation.

Li, Daqin. 2002. *Gemanyu Yanjiu* [Geman language research]. Beijing: Nationalities Press.

Longmailai, Monali. 2014. The morphosyntax of Dimasa. Shillong, India: Department of Linguistics, North-Eastern Hill University, Ph.D. thesis.

Matisoff, James A. 2013. Re-examining the genetic position of Jingpho: putting flesh on the bones of the Jingpho/Luish relationship. *Linguistics of the Tibeto-Burman Area* 36(2). 1–106.

Morey, Stephen. 2010. *Turung: A variety of Singpho language spoken in Assam*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Mulder, Miijke. 2020. A descriptive grammar of Muklom Tangsa. LaTrobe: LaTrobe University Doctoral dissertation.

Needham, Jack Francis. 1886. *Outline grammar of the Shaiyāng Miri language*. Shillong: Assam Secretariat Press.

Peterson, David. 2017. On Kuki-Chin subgrouping. In Picus Sizhi Ding & Jamin Pelkey (eds.), *Sociohistorical linguistics in Southeast Asia: New horizons for Tibeto-Burman studies in honor of David Bradley*, 189–209. Leiden: Brill.

Post, Mark W. 2007. A grammar of Galo. Bundoora: La Trobe University PhD dissertation.

Post, Mark W. & Jackson T-S. Sun. 2017. Tani languages. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages (2nd edition)*, 322–337. London: Routledge.

Prasad, Bal Ram. 1991. *Mising grammar* 17. Mysore: Central Institute of Indian Languages.

Sangdong, D. 2012. A grammar of the Kadu (Asak) language. Melbourne: La Trobe University Doctoral dissertation.

Thurgood, Graham. 2003. A subgrouping of the Sino-Tibetan languages: The interaction between language contact, change, and inheritance. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 1–21. Abingdon: Routledge.

VanBik, Kenneth. 2009. *Proto-Kuki-Chin: A reconstructed ancestor of the Kuki-Chin languages*. Berkeley, CA: STEDT Monograph 8.

Weidert, Alfonso K. 1984. The classifier construction of Newari and its Southeast Asian background. *Kailash* 11(3–4). 185–210.

Wood, Daniel C. 2008. An initial reconstruction of Proto-Boro-Garo. Eugene: University of Oregon Master's thesis.

Wood, Daniel. 2011. An initial reconstruction of the Proto-Boro-Garo Noun Phrase. In Gwendolyn Hyslop, Stephen Morey & Mark Post (eds.), *North East Indian linguistics* 3, 46–60. New Delhi, India: Cambridge University Press. https://doi.org/10.1017/UPO9788175968875.006

Address for correspondence

Jonathan P. Evans
Institute of Linguistics
Academia Sinica
Academia Road Section 2 No.128
Nangang District Taipei 11529
Taiwan
jonathan@sinica.edu.tw