

# Numeral Classifiers and Numeral Bases in Kuki-Chin Languages: A Case of Word Order Harmonization

**Abstract:** This study examines the numeral classifiers and numeral systems in South-Central Tibeto-Burman, aka Kuki-Chin (KC) languages, and first assesses their alignment with Greenberg’s (1990a[1972]) Universal 20A (U-20A) regarding the typology of classifier word orders, i.e., the noun (N) never intervenes between the numeral (Num) and the classifier (Cl). A comprehensive survey was conducted on KC languages, with data collected from the literature and fieldwork. The final analysis reveals that all KC languages with numeral classifiers employ an atypical CLASSIFIER-INITIAL order, where Cl precedes Num, thus [Cl Num], and the noun consistently takes the lead, i.e., [N Cl Num], confirming U-20A. Furthermore, multiplicative numerals in all the KC languages surveyed employ a BASE-INITIAL order, i.e., [base × multiplier]. Thus, KC classifier languages also obey the BASE-CL HARMONIZATION in word order, an implicational universal first proposed by Greenberg (1990b[1978]), where the base order dictates the classifier order. Such a harmonization thus solves the mystery posed by Evans (2022) regarding the atypical order of [Cl Num]. Future research will explore the historical and sociolinguistic factors shaping these patterns in the larger context of the Tibeto-Burman branch of the Sino-Tibetan family.

**Keywords:** Classifier Word Order, Greenberg’s Universal 20A, Kuki-Chin Languages, Numeral Systems, South-Central Tibeto-Burman

## 1. Introduction

The study of numeral systems and numeral classifiers is essential for understanding the broader linguistic typology of numeral formation, particularly in languages that utilize classifiers as an integral part of their grammatical structure (Her et al. 2022). Numerals as linguistic expressions of natural numbers provide information about the cognitive and cultural practices of a speech community (Hurford 2011). In this paper we are concerned with multiplicative numerals. For example, in Mandarin, the numeral *sanbai* ‘300’ is constructed by a base-final *bai* ‘100’ multiplied by a preceding multiplier *san* ‘3’, while in Ibibio, the numeral *ikie ita* ‘300’ is constructed by a base-initial *ikie* ‘100’ multiplied by a following multiplier *ita* ‘3’. If a language has both orders, it is called base-split.

Meanwhile, numeral classifiers, which are used to categorize nouns based on inherent characteristics such as shape, size, animacy, or function, are fundamental in the syntax of quantification in many languages around the world, particularly, though not exclusively, within the East and Southeast Asian linguistic areas. (Bale et al. 2019, Her and Tang 2020, Vittrant et al. 2021). Numeral classifiers are so called because they occur with a numeral, and we are concerned with order between the numeral and the classifiers, i.e., whether the classifier precedes or follows the numeral. More importantly, is there any correlation between the base order and the classifier order.

Since Aikhenvald’s (2000) typological investigation into classifier systems, research on numeral classifiers has continued to advance both theoretically and empirically. Surveys in the World Atlas of Language Structures Online (WALS) indicate that 140 of 400 languages exhibit classifier systems (Gil 2013), illustrating their wide typological distribution. Recently, the World Atlas of Classifier Languages (WACL) examined 3338 languages and identified 723 as having numeral classifier systems (Her et al. 2022), thereby contributing to a more comprehensive understanding of the global distribution of numeral classifier languages.

Numerals (Nums) exhibit a multiplicative relation with classifiers (Cls), as illustrated in (1). Similarly, simple numerals (n's) also display a multiplicative relation with numeral bases (b's), as shown in (2). These patterns suggest a systematic connection among Num, base, and classifier within the nominal phrase (Her 2017).

(1) Mandarin:

- |    |                    |                     |         |                      |
|----|--------------------|---------------------|---------|----------------------|
| a. | wu                 | ke                  | pingguo | (5 × <u>1</u> apple) |
|    | five               | C                   | apple   |                      |
|    | 'five apples'      |                     |         |                      |
| b. | wu                 | da                  | meigui  | (5 × <u>12</u> rose) |
|    | five               | C <sub>-dozen</sub> | rose    |                      |
|    | 'five dozen roses' |                     |         |                      |

(2) Mandarin:

- |    |                 |                 |                        |
|----|-----------------|-----------------|------------------------|
| a. | wu              | qian            | (5 × 10 <sup>3</sup> ) |
|    | five            | 10 <sup>3</sup> |                        |
|    | 'five thousand' |                 |                        |
| b. | wu              | bai             | (5 × 10 <sup>2</sup> ) |
|    | five            | 10 <sup>2</sup> |                        |
|    | 'five hundred'  |                 |                        |

While the multiplication in multiplicative numerals, e.g., (2a-b), is rather transparent, the insight that a multiplicative function likewise exists between the numeral and the classifier, e.g., (1a-b), is first observed by Greenberg (1990 [1972]:172) and later refined by Her (2012): sortal classifiers encode the numerical value 1, e.g., (1a); mensural classifiers encode a value that is not 1. A host of subsequent studies supported this view, e.g., Her (2017), Her and Tang (2020), Her et al. (2019), Her et Tang (2022), Her et al. (2022), Her and Li (2023), and Her et al. (2024), which in turn inspired this present study on KC languages.

In terms of word order, classifier languages across the world seldom display a [Cl Num] order (Evans 2022). According to Evans (2022), the Sino-Tibetan languages with the [Cl Num] pattern are primarily found in Northeast India and surrounding areas. We propose that this atypical word order is linked to the presence of base-initial structure in numeral expressions. Our study focuses on examining and explaining the relationship between [Cl Num] order and base-initial patterns in numeral phrases within Kuki-Chin (KC) languages, a group of Sino-Tibetan languages spoken in the targeted area where base-initial languages are notably concentrated (Her et al. 2024).

The paper is organized as follows. Section 2 explains the classifier word order typology, Greenberg's Universal 20A, and the word order hypothesis of base-Cl harmonization, based on a functionalist interpretation of the [Num Cl] unit as a multiplicative unit, where Num is a multiplier, and Cl, a multiplicand. Section 3 provides an overview of numeral systems and numeral classifiers in Tibeto-Burman languages. Section 4 addresses the numeral systems and numeral classifiers in Kuki-Chin languages. Section 5 concludes the paper with a discussion of the findings and suggestions for future research.

## 2. Numeral Classifiers as Multiplicands

Observation of examples (1) and (2) suggests that the structure [Num Cl] or [Cl Num] functions as a multiplicative unit, with Num acting as the multiplier and Cl as the multiplicand, analogous to the numeral phrase (n × b) or (b × n), where n is the

multiplier and b the multiplicand (Her 2017). Since [Num Cl] or [Cl Num] form a phrase with a multiplicative relation, the noun (N) cannot intervene between the numeral and the classifier (Her 2017). Her (2017) synthesized Greenberg's (1990a[1972]) typological generalizations on classifier word order, as summarized in (3), and formulated Greenberg's Universal 20A (hereafter U-20A), shown in (4). According to U-20A, the noun does not appear between the numeral and the classifier, and the numeral tends to precede the classifier.

(3) Classifier word order typology

- a. ✓ [Num Cl N] (many languages, e.g., Mandarin (Sinitic))
- b. ✓ [N Num Cl] (many languages e.g., Thai (Tai))
- c. ✓ [Cl Num N] (few languages e.g., Kilivila (Oceanic))
- d. ✓ [N Cl Num] (few languages e.g., Jingpho (Tibeto-Burman))
- e. \* [Cl N Num] (no languages)<sup>1</sup>
- f. \* [Num N Cl] (no languages)

(4) Greenberg's Universal 20A (U-20A)

- Part 1: Of the three elements Num, Cl, and N, any order is possible as long as N does not come between Num and Cl.
- Part 2: There are many more languages with Num > Cl orders than languages with Cl > Num orders.

In quantifying phrases, all classifiers act as multiplicands in relation to numerals, which serve as multipliers, and together they form a syntactically coherent category as shown in (5) and (6). Classifiers can be categorized based on their mathematical values, as shown in Table 1.

(5) Mandarin:

yi	zhang	zhuozi
one	Cl	table
‘a table’		
(1 × <u>1</u> table)		

(6) Mandarin:

yi	ke	pingguo
one	Cl	apple
‘an apple’		
(1 × <u>1</u> apple)		

---

<sup>1</sup> There are alleged cases of violations to U-20A in the form of [Cl N Num] in some Tai-Kadai languages, where the alleged Num must be an indigenous numeral one, and several African languages. Her (2017) has demonstrated that in the former cases, the numeral “1” is, in fact, in fact, an indefinite article like the English a/an, and the alleged numeral classifiers in the African cases turned out to be nouns/an indefinite article.

**Table 1. Types of mathematical values denoted by C/Ms**

Numerical	Fixed	$n = 1$ e.g., <i>ben</i> (本), <i>ke</i> (顆), <i>tiao</i> (條), <i>zhi</i> (隻)	C
		$n = 2$ e.g., <i>dueli</i> (pair 對); $n = 12$ e.g., <i>da</i> (dozen 打)	M <sub>1</sub>
	Variable	$n > 1$ e.g., <i>pai</i> (row 排), <i>bang</i> (gang 幫), <i>die</i> (stack 疊)	M <sub>2</sub>
Non-numerical	Fixed	e.g., <i>gongjin</i> (kilogram 公斤), <i>gongli</i> (kilometer 公里)	M <sub>3</sub>
	Variable	e.g., <i>di</i> (drop 滴), <i>dai</i> (bag 袋), <i>bei</i> (cup 杯)	M <sub>4</sub>

Numeral classifiers thus come in two varieties: sortal classifiers (C), whose mathematical value as a multiplicand is fixed at 1, and mensural classifiers (M), whose mathematical values are not 1. Ms can be classified into four types: while M<sub>1</sub> and M<sub>2</sub> both encode numerical values, M<sub>1</sub> has fixed values and M<sub>2</sub> does not. While M<sub>3</sub> and M<sub>4</sub> both encode non-numerical values, M<sub>3</sub> has fixed values and M<sub>4</sub> does not.

The shared function as multiplicands between Cl and the numeral base provides a cognitive motivation for the base-Cl harmonization. Her et al. (2019) examined the role of sortal and mensural classifiers in classifier languages when quantifying nouns with numerals. The study investigated the harmonization of word order between numeral classifiers and numeral bases, suggesting that this harmonization is linked to their shared function as multiplicands. Two potential probabilistic universals were proposed: the presence of classifiers implies the presence of multiplicative numerals, and the word order of base and classifiers should be harmonized, either both final or both initial, as shown in example (7) and (8).

(7) Bawm:

som- khat  
ten one  
'ten'

(8) Bawm:

mai pum khat  
pumpkin Cl one  
'one pumpkin'

Her et al. (2019) systematically surveyed six language families, Sinitic, Miao-Yao, Austro-Asiatic, Tai-Kadai, Tibeto-Burman, and Indo-Aryan (dubbed SMATTI), to verify these universals. Their analysis revealed that 100% of the 219 SMATTI classifier languages use multiplicative numerals. Additionally, 97.26% of the languages show harmonization between numeral bases and classifiers. The four attested word orders can be classified based on two binary parameters: [base-final vs. base-initial] and [Cl-final vs. Cl-initial], leading to the taxonomy in Table 2.

**Table 2. Harmonization between numeral bases and classifiers (Her et al. 2019)**

	Cl-final	Cl-initial	No. of Languages
base-final	187 (85.39%)	5 (2.28%)	192
base-initial	1 (0.46%)	26 (11.87%)	27
Languages	188	31	219

U-20A thus provides an insightful framework for studying the variation of classifier word orders in classifier languages. For example, it can be employed to elucidate the harmonization of word order between Cl and the numeral bases in Num,

first proposed by Greenberg (1990b[1978]: 292) and further developed by Her et al. (2019). This hypothesis suggests that the word order between Cl and Num, whether Cl-final [Num Cl], e.g., Mandarin, or Cl-initial [Cl Num], e.g., Kilivila, should correlate with the internal word order between a numeral base and its multiplier in a multiplicative numeral, whether base-final, e.g., *sān-bǎi* ‘three hundred’ in Mandarin or base-initial, e.g., *akatu-tolu* ‘hundred three’ in Kilivila. This hypothesis is summarized in (9).

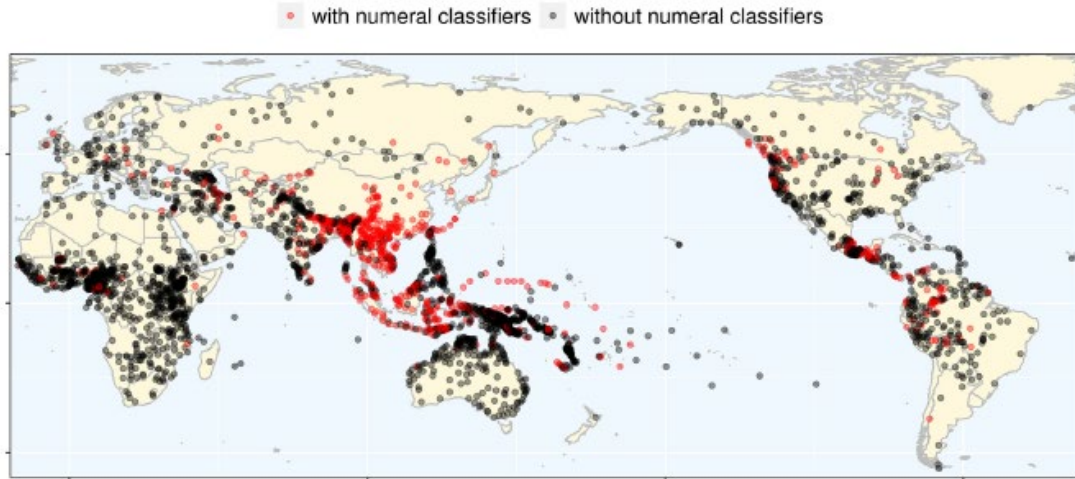
- (9) The word order hypothesis of base-Cl harmonization
- a. Cl-final order, or [Num Cl] → base-final numerals, or [n × base]
  - b. Cl-initial order, or [Cl Num] → base-initial numerals, or [base × n]

Greenberg’s Generalization 28 proposes that in languages with numeral classifiers, it is [numeral + classifier] that forms a syntactic unit, not [classifier + noun] (Greenberg 1990b[1978]: 292). Building on this generalization, along with multiplicative relations and the assumption that the noun (N) is the head of the nominal construction, Allasonnière-Tang and Her (2020:513) argue that there should be harmonization between base and N, base and Cl, and Cl and N, as stated in (10).

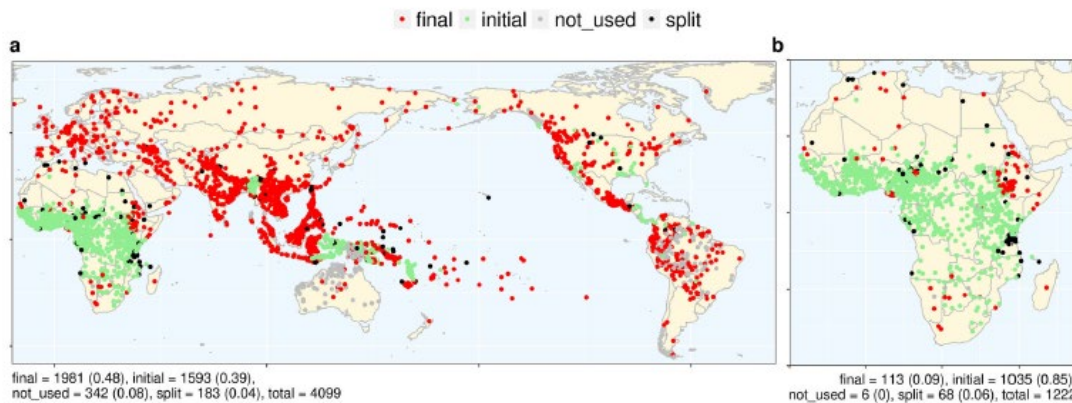
- (10) a. Base-N harmonization:  
If Num is base-initial, i.e., [base n], then an N-initial order obtains between N and Num; otherwise, an N-final order obtains.
- b. Base-Cl harmonization:  
If Num is base-initial, i.e., [base n], then a Cl-initial order obtains between Cl and Num; otherwise, a Cl-final order obtains.
- c. Cl-N harmonization:  
If a Cl-initial order obtains between Cl and Num, then an N-initial order obtains between N and Cl; otherwise, an N-final order obtains.

Allasonnière-Tang and Her (2020) examined word order patterns in a geographically and phylogenetically balanced sample of 400 languages. Their findings reveal that the harmonization between numeral bases, classifiers, and nouns as illustrated in (10) is statistically significant. Deviations from this pattern are largely found at contact zones between head-final and head-initial languages, suggesting that language contact is a key factor in exceptions to these probabilistic universals.

Her et al. (2022) and Her et al. (2024) conducted extensive investigations into numeral classifier languages and numeral base orders. The geographical distribution of numeral classifier systems across 723 of the 3338 languages surveyed (Her et al. 2022) is shown in Figure 1. Figure 2 shows the base order of the 4009 languages examined in Her et al. (2024).



**Figure 1.** Geographical distribution of numerical classifier languages (Her et al. 2022)

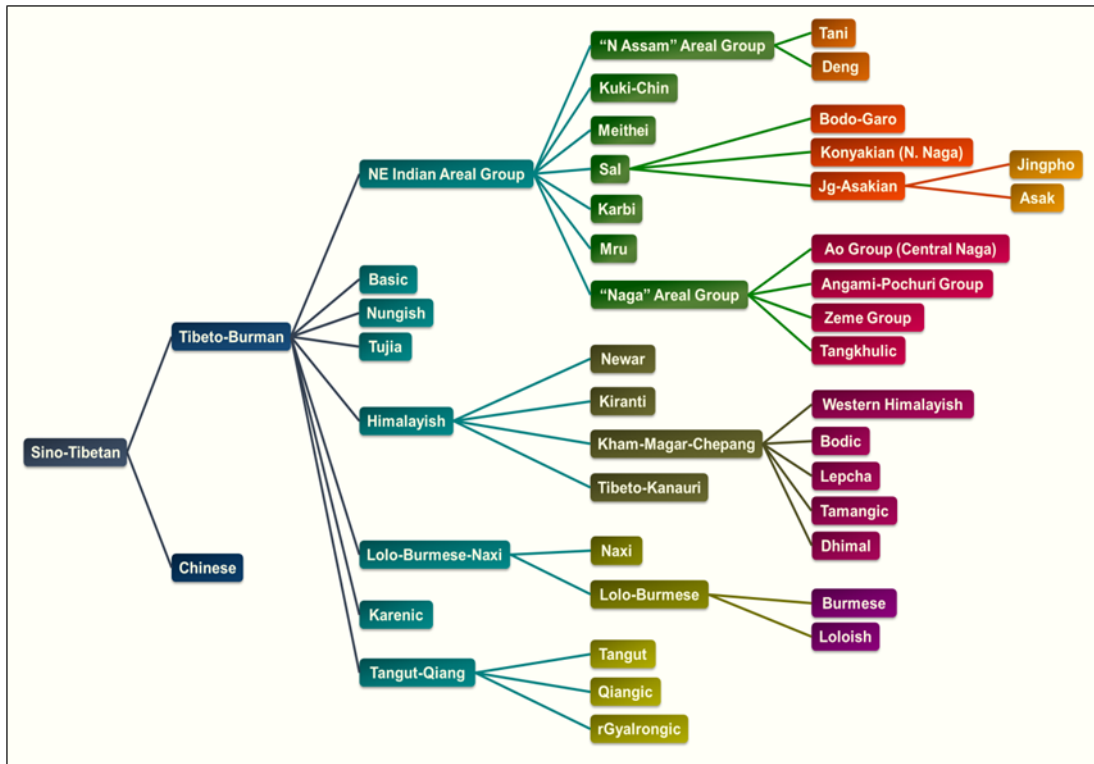


**Figure 2.** Geographical distribution of 4009 languages by base order (Her et al. 2024).

### 3. An overview of numeral systems and numeral classifiers in TB

Numeral classifiers are a grammatical feature in many languages around the world, particularly those in East and Southeast Asia (e.g., Aikhenvald 2022). These classifiers function as linguistic tools that categorize nouns based on their inherent properties, such as shape, size, animacy, or function, when these nouns are quantified by numerals. In languages with classifier systems, a numeral alone is often insufficient to quantify a noun; the numeral must be accompanied by a classifier that is semantically appropriate for the noun being counted. For example, in a language with a classifier system, one might say “three sheet paper” rather than simply “three paper,” where “sheet” serves as the sortal classifier that categorizes “paper” according to its physical properties. The choice of classifier provides additional information about the noun.

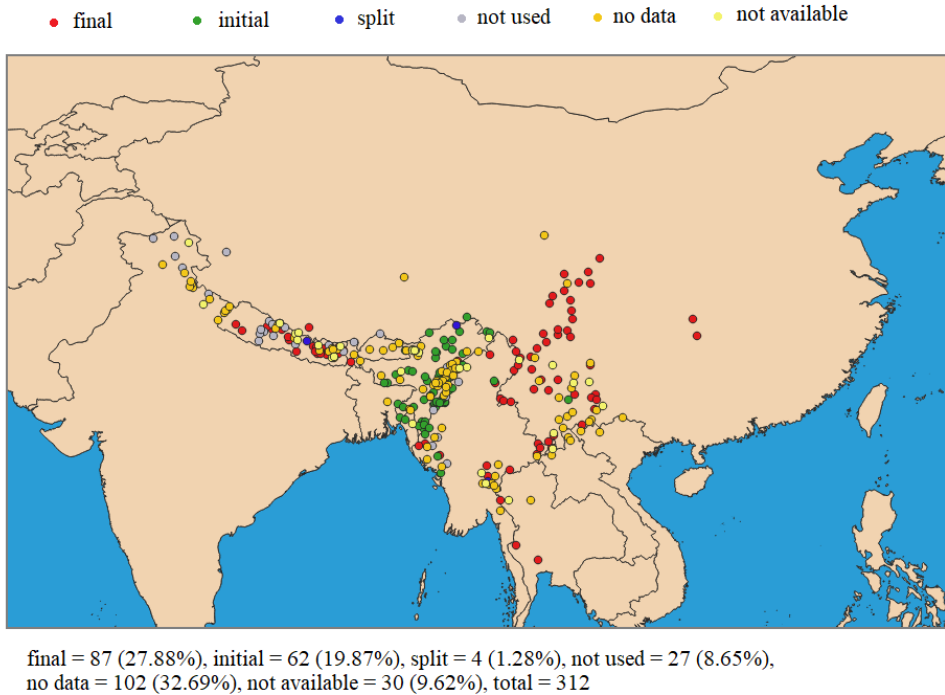
The Tibeto-Burman (TB) language family, a major branch of the Sino-Tibetan language family, is one of the most diverse and complex language families in the world (LaPolla 2013). Figure 3 is a linguistic family tree showing the classification of the Sino-Tibetan language family (Matisoff 2015:xxxii), which includes a diverse group of languages spoken in East Asia and Southeast Asia.



**Figure 3.** The Sino-Tibetan language family tree (Matisoff 2015: xxxii)

The Sino-Tibetan family is primarily divided into the Chinese and the Tibeto-Burman languages. The Tibeto-Burman languages form a vast and complex subgroup within the Sino-Tibetan family, spoken across a wide geographic area. It includes over 400 languages spoken by millions of people in regions as varied as northeastern India, Nepal, Bhutan, Myanmar, Tibet, and parts of China, including Yunnan, Sichuan, and Tibet Autonomous Region (Thurgood & LaPolla 2017). This language family is characterized by significant linguistic diversity, both in terms of phonological systems and grammatical structures.

Syntactically, word order varies considerably across languages. While many Tibeto-Burman (TB) languages tend to follow a Subject-Object-Verb (SOV) pattern, others display word order patterns that resist straightforward typological classification (DeLancey 2022). This structural diversity extends to numeral phrases, where both base-initial and base-final orders, as well as classifier-initial and classifier-final configurations, are attested across TB languages (Her et al. 2019). Figure 4 shows the distribution of classifier languages and classifier word orders in TB.

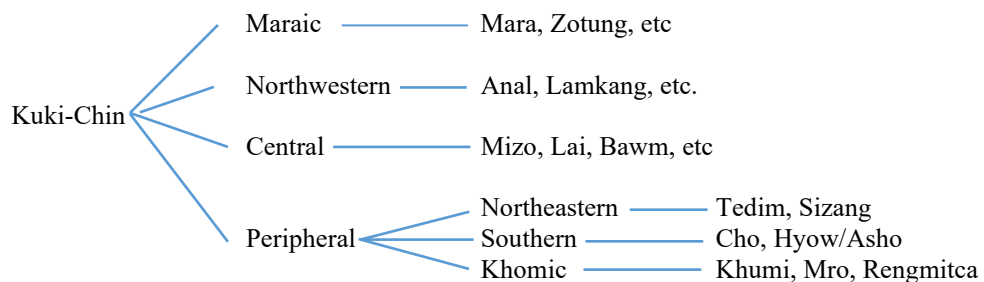


**Figure 4.** Distribution of classifier languages and classifier word orders in TB

Out of the 312 TB languages in the WACL database (Her et al. 2022), 180 has information regarding the classifier feature, where 87 use the Cl-final order, 62 use the Cl-initial order, 4 use both orders, and 27 have no classifiers. All other classifier languages in the Sino-Tibetan family have only the Cl-final order.

#### 4. Numeral Systems and Classifiers in Kuki-Chin Languages

Kuki-Chin (KC) languages, spoken primarily in northeastern India as well as in parts of Myanmar and Bangladesh, constitute a linguistically and culturally rich subgroup of the Tibeto-Burman family (Post & Burling 2017). With over a million speakers across a diverse area including the Indian states of Manipur, Assam, Mizoram, Nagaland, and Tripura in India and the Chittagong Hill Tracts in Bangladesh, and the Chin and Rakhine states and Sagaing division in Myanmar, KC languages display a wide range of dialectal and typological variation. This geographic and social diversity has contributed to the emergence of complex linguistic features, including tonal (Mortensen 2023) and morphological systems. Loanwords and the adoption of phonological and syntactic features reflect the region’s intricate social and linguistic history (Zakaria 2024). Figure 5 is the Stammbaum of Kuki-Chin language family according to Peterson (2017:206).



**Figure 5.** Stammbaum of Kuki-Chin languages (Peterson 2017:206)

The data on the numeral bases in KC languages is mostly derived from the raw data in Eugene Chan’s (2024) website *Numeral Systems of the World’s Languages* (<https://lingweb.eva.mpg.de/channumerals/>), which covers some 4,000 languages. A substantial number of KC languages employ sortal classifier systems to categorize nouns. The data on the numeral classifiers in KC languages is mostly obtained from the *World Atlas of Classifier Languages* (WACL), which has information on 3338 languages and recognizes 723 as having numeral classifier systems (Her et al. 2022). However, KC languages remain relatively understudied with respect to their numeral systems and classifier systems.

In this section, we present additional new data collected through fieldwork with native speakers, supplemented by secondary sources such as existing typological databases and linguistic surveys (e.g., Aimol 2009, Suantak 2013, Singh 2014, Kim 2016, Awan 2017, Wamsley 2019, Saikia and Mamta 2020, Singha 2022). Fieldwork was conducted in regions where specific KC languages are spoken, including the Indo-Burmese border, Manipur, Mizoram, Assam, Tripura, and the Chittagong Hill Tracts of Bangladesh. Primary data was obtained through elicitation sessions with native speakers.

For each language, the numeral systems and classifier constructions were documented based on fieldwork data, with particular attention to word order patterns. The collected data were synthesized to identify major typological patterns across KC languages. This involved analyzing the relationship between numeral bases and classifiers, comparing these patterns with existing typological databases, and referencing broader Tibeto-Burman typological trends. Based on the observed data, KC numeral systems were classified into three categories: base-initial, base-final, and base-split (i.e., with both base-initial and base-final orders). The classifier systems were categorized as Cl-initial, Cl-final, Cl-split (i.e., with both Cl-initial and Cl-final orders), and no-Cl.

#### 4.1 Base and classifier features in Kuki-Chin languages

According to Suantak et al. (2026), 45 KC languages have been documented with available information regarding their base features, while an additional 9 languages lack reliable data. The distribution of base features in KC languages is summarized in Table 3. It indicates a predominant base-initial structure, with 41 out of the 54 KC languages in *Glottolog* (75.93%) categorized as base-initial. This strong preference for a base-initial structure suggests a syntactic inclination towards placing the base element at the beginning of the noun phrase.

No attested KC languages exhibit a base-final numeral structure. Only a small number of languages (approximately 7%) show base-split configurations, suggesting that such structures are relatively uncommon in the KC subgroup. Notably, 16.67% of the surveyed languages are without relevant data. This reflects both the region’s limited accessibility and the current state of documentation and highlights the importance of continued fieldwork.

**Table 3.** Distribution of the base feature in KC languages

Base feature	No. of languages	Percentage
base-final	0	0%
base-initial	41	75.93%
base-split	4	7.41%
No data	9	16.67
Total	54	100.01%

The distribution of CI features in KC languages is summarized in Table 4. In terms of classifier features, no attested KC languages exhibit a CI-initial structure. 59.26% of KC languages utilize a CI-initial structure, and 12.96% of KC languages do not employ classifiers at all. Similarly, 27.78% of KC languages are listed as having no data for CI word order, pointing to the potential value of further field investigation.

**Table 4. The classifier feature in KC languages (Suantak et al. 2026)**

CI feature	No. of languages	Percentage
CI-final	0	0%
CI-initial	32	59.26%
CI-split	0	0%
no-CI	7	12.96%
No data	15	27.78%
Total	54	100%

#### 4.2 Analysis of fieldwork data

The study of numeral bases across the 28 KC languages from direct fieldwork reveals a predominant reliance on the decimal system. The features of the languages are summarized in Table 5. The table provides a detailed overview of the numeral systems and word orders used in various KC languages spoken primarily in the northeastern regions of India and adjacent areas in Myanmar and Bangladesh. In all 28 languages, the numeral system follows a base-initial format, denoted as  $(b \times n)$ , where the numeral base precedes the numeral in the sequence. The word order for numeral classifiers across these languages is notably uniform. The classifier precedes the numeral, as shown in the [CI Num]  $\rightarrow$  base-initial pattern. This means that in these languages, classifiers are positioned before the numeral itself in numerical expressions. This uniformity in numeral systems and word order across the languages listed suggests a strong regional influence in the structure of numerical expressions, reflecting shared linguistic features among the languages spoken in the Indo-Burmese border and surrounding regions.

**Table 5. Numeral systems and word orders in the 28 Kuki-Chin languages**

Language	Region	Numeral System	Word Order
Anal	Indo-Burmese border of Southeastern Manipur, India	Base-initial ( $b \times n$ )	CI-initial order, [CI Num] $\rightarrow$ base-initial
Bawm	Chittagong Hill Tracts of Bangladesh, Chhimituipui, Lunglei, Aizawl Districts, Tripura, Assam	Base-initial ( $b \times n$ )	CI-initial order, [CI Num] $\rightarrow$ base-initial
Chiru	Manipur, Cachar district of Assam	Base-initial ( $b \times n$ )	CI-initial order, [CI Num] $\rightarrow$ base-initial
Darlong	Unakoti District, Tripura	Base-initial ( $b \times n$ )	CI-initial order, [CI Num] $\rightarrow$ base-initial
Gangte	Manipur, adjacent areas of Meghalaya and Assam	Base-initial ( $b \times n$ )	CI-initial order, [CI Num] $\rightarrow$ base-initial

<b>Language</b>	<b>Region</b>	<b>Numeral System</b>	<b>Word Order</b>
Hmar	Manipur, Mizoram, Assam, Western Myanmar, Eastern Bangladesh	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Hrangkhawl	Dhalai District, parts of West District in Tripura	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Kom	Manipur	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Paite	Manipur, Mizoram, Assam, Chin state (Burma)	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Paang	Lawngtlai District, Mizoram	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Sakachep	KarbiAnglong, Dima Hasao, Cachar districts of Assam; Tripura, Mizoram, Manipur	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Simte	Manipur	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Thadou	Six districts of Manipur; KarbiAnglong, North Cachar district of Assam; Myanmar	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Zou	Mizoram, Manipur, Western Burma	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Biate	Dima Hasao, Assam; Jaintia Hills, Meghalaya	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Faihriem/ Saihriem	Cachar district, Assam	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Koireng	Manipur, India	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Lamkang	Manipur	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Chothe	Southeastern Manipur and Nagaland near Myanmar border	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Moyon	Chandel District, Manipur, India	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Tarao	Manipur	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Falam-Chin	Falam township, Myanmar; Lawngtlai district, Mizoram	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Hakha-Chin	Myanmar	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Halam	Dhalai district, Tripura	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial
Zyphe/Zophei	Thantlang Township, Chin state, Myanmar; Champhai, Siaha Districts, Mizoram	Base-initial (b × n)	Cl-initial order, [Cl Num] → base-initial

Language	Region	Numeral System	Word Order
Ralte	Aizawl North, Chhimtuipui District, Northern parts of Mizoram	Base-initial ( $b \times n$ )	Cl-initial order, [Cl Num] → base-initial
Mara	Chhimtuipui Districts, Southern Mizoram, India; adjacent villages in Burma	Base-initial ( $b \times n$ )	Cl-initial order, [Cl Num] → base-initial
Vaiphei	Manipur, Assam, Mizoram, Nagaland, Meghalaya, and in the Chin State of Myanmar	Base-initial ( $b \times n$ )	Cl-initial order, [Cl Num] → base-initial

The numeral systems across the analyzed languages exhibit a remarkable consistency in structure. These languages universally employ a base-initial numeral system, characterized by the formula ( $b \times n$ ), where  $b$  denotes the base (typically 10, 100, etc.), and  $n$  represents the multiplier.<sup>2</sup> This base-initial structure is a defining feature of these languages and demonstrates a shared cognitive and linguistic framework. For example, the numeral systems in three KC languages are shown below:

(11) Bawm:

<i>som-khat</i>	$(10 \times 1) = 10$
<i>za-khat</i>	$(100 \times 1) = 100$
<i>som-hnih</i>	$(10 \times 2) = 20$
<i>za-hnih</i>	$(100 \times 2) = 200$

(12) Falam-Chin:

<i>pahraa</i>	$(10 \times 1) = 10$
<i>za</i>	$(100 \times 1) = 100$
<i>pahraa-ni</i>	$(10 \times 2) = 20$
<i>za-ni</i>	$(100 \times 2) = 200$

(13) Hakha-Chin:

<i>pahraa</i>	$(10 \times 1) = 10$
<i>za-khat</i>	$(100 \times 1) = 100$
<i>pahraa-hnih</i>	$(10 \times 2) = 20$
<i>za-hnih</i>	$(100 \times 2) = 200$

This pattern of base-initial numerals is consistent across the examined languages, suggesting a deeply rooted linguistic tradition within the KC subgroup.

The classifier systems in these languages are intricately linked to their numeral structures. A dominant pattern observed across the languages is the [N Cl Num] word order, which reflects a harmonization between the base-initial numeral systems and the Cl-initial word order. This order is consistent across the languages studied and is crucial for understanding the linguistic typology of the region. Examples of [N Cl Num] Word Order in some KC languages are shown below:

<sup>2</sup> We thank an anonymous reviewer for pointing out that some KC languages have vigesimal numeral bases; for example, Dimasa has a mixed system, with some vigesimal numerals. However, the focus of this is on the word order of the numeral bases in general and thus whether the bases are decimal or vigesimal is beyond the scope of our investigation.

- (14) Bawm:
- |               |     |      |
|---------------|-----|------|
| mai           | pum | khat |
| pumpkin       | Cl  | one  |
| ‘one pumpkin’ |     |      |
| banla         | pum | nih  |
| banana        | Cl  | two  |
| ‘two bananas’ |     |      |
- (15) Faihriem/Saihriem:
- |                      |       |      |
|----------------------|-------|------|
| lekha                | pek   | khat |
| paper                | Cl    | one  |
| ‘one sheet of paper’ |       |      |
| ui                   | khuah | nih  |
| dog                  | Cl    | two  |
| ‘two dogs’           |       |      |
- (16) Kom:
- |                                    |     |        |
|------------------------------------|-----|--------|
| mai                                | pum | inkhat |
| pumpkin                            | Cl  | one    |
| ‘one pumpkin’                      |     |        |
| lekha                              | pek | inni   |
| paper                              | Cl  | two    |
| ‘two papers (two pieces of paper)’ |     |        |

No instance of N intervening between Num and Cl has been found in KC languages. Her et al. (2019), based on multiplicative relations, propose a hypothesis of word order harmonization between Cl and numeral bases. It can be observed that the word order of all languages in Table 5 is Cl-initial and base-initial. The Cl and base orders in the structures [Cl Num] and [base × n] are accurately predicted by the hypothesis of base-Cl harmonization.

## 5. Discussions and Conclusions

The persistence of the base-initial numeral system in KC languages could be traced back to historical linguistic practices. Proto-Tibeto-Burman languages might have had a base-initial numeral system, which was inherited and maintained by descendant languages.<sup>3</sup> The historical stability of this numeral system across different languages and regions suggests that it was a significant feature in the ancestral language, influencing the numeral systems of contemporary KC languages. From a cognitive perspective, the base-initial numeral system might be advantageous for processing and understanding numerical information. This structure could be easier for speakers to use in everyday contexts, potentially making it a preferred system due to its cognitive simplicity and efficiency (Her et al. 2024).

---

<sup>3</sup> Classifiers are likely not present in either Proto-Sino-Tibetan or Proto-Tibeto-Burman. Numeral classifiers in Asia and the Pacific may have originated in Sinitic as suggested by Her and Li (2023).

The higher frequency of base-initial word order in TB languages relative to neighboring languages has been observed. For example, among the 219 languages examined by Her et al. (2019), including Sinitic, Miao-Yao, Austro-Asiatic, Tai-Kadai, Tibeto-Burman, and Indo-Aryan (SMATTI) languages, 27 out of 100 TB languages are base-initial. Furthermore, they are the only base-initial languages that appear in the study.

In the Northeast India linguistic area, there are groups of TB languages beyond the KC group. Base-initial languages are not limited to the KC group within the TB languages in this region. As a supplementary example, Galo, a language of the Tani subgroup within the TB branch, also exhibits base-initial word order, as shown in (17).

(17) Galo:

*cam-ji*            (10 × 2) = 20  
*cam-múm*        (10 × 3) = 30  
 (Post 2007: 381)

As predicted by the hypothesis of base-Cl harmonization, Galo is a Cl-initial language, as shown in (18).

(18) Galo:

*hiidàa*                      *dáa-*            *ji*  
 stick                        Cl-                two  
 ‘two sticks’  
 (Post 2007: 387)

Figure 2 shows that base-initial languages appear to be geographically concentrated in Africa and the Northeast India linguistic area. By comparing Figures 1 and 2, we can observe that Northeast India is the region where numeral classifier languages and base-initial languages overlap. This may account for what Evans (2022) referred to as a ‘mystery’, namely, why many languages with [Cl Num] structures are found in Northeast India linguistic area. The reason is that, although base-initial languages are also concentrated in Africa, this region is not a center for numeral classifier languages and thus does not exhibit base-Cl harmonization.

This study provides a comprehensive examination of numeral systems and sortal classifiers within the Kuki-Chin (KC) languages. The investigation reveals that KC languages predominantly employ a base-initial numeral system which is consistent across the analyzed languages. This pattern appears consistent with observations from some other Tibeto-Burman languages, suggesting a possible shared structural preference within the KC subgroup. The study also reveals a predominant use of Cl-initial structures and a general absence of Cl-final or Cl-split structures. This contrasts with patterns observed in other languages of the Tibeto-Burman branch, where a wider range of classifier structures is attested. The [N Cl Num] word order, which is prevalent across KC languages, aligns with Greenberg’s Universal 20A, in that KC languages consistently avoid placing nouns between numerals and classifiers. The atypical [N Cl Num] order in KC languages can be explained by the word order hypothesis of base-Cl harmonization proposed by Her et al. (2019).<sup>4</sup>

---

<sup>4</sup> An anonymous reviewer points out that other TB languages may have similar word order properties to KC languages but lack the order [N Cl Num]. It is our hypothesis that in other TB languages, and elsewhere as well, the base-Cl harmonization always holds. Thus, the Cl-initial order in a language requires base-initial numerals. This hypothesis is being explored in a separate study.

A number of KC languages, including Aimol, Anal, and Bawm Chin, exhibit base-initial and classifier-initial structures. These patterns are also attested in some other languages of the Tibeto-Burman branch, and their consistency may point to a shared linguistic ancestry or similar sociolinguistic influences. In contrast, Suantak et al. (2026) stated that languages such as A sho Chin and Daai Chin display base-split structures and variations in classifier usage. The presence of base-split structures in A sho Chin and Daai Chin suggests a divergence from the dominant base-initial pattern, indicating a level of internal diversity that may be attributed to distinct historical, geographical, or cultural influences.

Future research should focus on comparative analysis across different language families to determine whether the base-initial numeral systems and CI-initial order found in KC languages represent a rare configuration or appear in other unrelated language families as well. Exploring historical, cultural, and contact-related influences, as well as diachronic changes and cognitive aspects of numeral systems, can offer deeper insights. Enhancing documentation of under-studied languages and examining the syntactic and semantic roles of numeral classifiers will further contribute to a comprehensive understanding of these linguistic features and their evolution.

## Acknowledgments

To be added.

## References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2022. Classifiers: Setting the scene: An introduction to the special issue on classifiers in the languages of Asia. *Asian Languages and Linguistics* 3.2:141–152.
- Aimol, Chongom Damrengthang. 2009. A Comparative Study of Kom and Aimol Numerals. *North East Indian Linguistics* 2:155–65.
- Allasonnière-Tang, Marc, and One-Soon Her. 2020. Numeral base, numeral classifier, and noun: Word order harmonization. *Language and Linguistics* 21.4:511–556.
- Awan, Mechek Sampar. 2017. Numerals in Chiru. *Languages in India* 17.7:37–45.
- Bale, Alan, Jessica Coon, and Nicolás Arcos López. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa: A Journal of General Linguistics* 4.1:1–30.
- Chan, Eugene. 2024. *Numeral Systems of the World's Languages*. <https://lingweb.eva.mpg.de/channumerals>.
- DeLancey, Scott. 2022. Sino-Tibetan Languages. *The Major Languages of East and South-East Asia*, ed. by Bernard Comrie, 69–82. New York: Routledge.
- Evans, Jonathan P. 2022. Classifiers before numerals: Origin of an atypical morpheme order. *Linguistics of the Tibeto-Burman Area* 45.1:1–21. <https://doi.org/10.1075/ltba.21013.eva>
- Gil. 2013. Numeral Classifiers. WALS Online (v2020.4), ed. by Matthew S. Dryer and Haspelmath. Retrieved July 10, 2025, from <https://doi.org/10.5281/zenodo.13950591>
- Greenberg, Joseph H. 1990a[1972]. Numerical classifiers and substantival number: Problems in the genesis of a linguistic type. *On language. Selected writings of*

- Joseph H. Greenberg, ed. by Keith Denning and Suzanne Kemmer, *On language. Selected writings of Joseph H. Greenberg*, 166–193. Stanford, CA: Stanford University Press. (Original work published 1972)
- Greenberg, Joseph H. 1990b[1978]. Generalizations about numeral systems. *On language: Selected writings of Joseph H. Greenberg*, ed. by Keith Denning & Suzanne Kemmer, 271–309. Stanford, CA: Stanford University Press. (Original work published 1978)
- Her, One-Soon. 2012. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122.14: 1668–1691.
- Her, One-Soon. 2017. Deriving classifier word order typology, or Greenberg’s Universal 20A and Universal 20. *Linguistics* 55.2:265–303.
- Her, One-Soon and Marc Tang. 2020. A statistical explanation of the distribution of sortal classifiers in languages of the world via computational classifiers. *Journal of Quantitative Linguistics* 27.2:93–113.
- Her, One-Soon, Marc Allasonnière-Tang, and Bing-Tsiung Li. 2019. Word order of numeral classifiers and numeral bases: Harmonization by multiplication. *STUF-Language Typology and Universals* 72.3:421–452.
- Her, One-Soon, Hammarström Harald, and Marc Allasonnière-Tang. 2022. Defining numeral classifiers and identifying classifier languages of the world. *Linguistics Vanguard* 8.1:151–64.
- Her, One-Soon, and Bing-Tsiung Li. 2023. A single origin of numeral classifiers in Asia and the Pacific. *Nominal classifications in Asia and Oceania: Functional and Diachronic Perspectives*, ed. by Marc Allasonnière-Tang and Marcin Kilarski, 113–160. Amsterdam: John Benjamins.
- Her, One-Soon, Yung-Ping Liang, Eugene Chan, Hung-Hsin Hsu, Anthony Chi-Pin Hsu, and Marc Allasonnière-Tang. 2024. Early humans out of Africa had only base-initial numerals. *Humanities and Social Sciences Communications* 11.254: 1–7. Retrieved July 10, 2025, from <https://doi.org/10.1057/s41599-023-02506-z>
- Hurford, James R. 2011. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
- Kim, D. Mary Haopkip. 2016. Numeral System in Thadou. *Language in India* 16.8:68–78.
- LaPolla, Randy J. 2013. Eastern Asia: Sino-Tibetan linguistic history. *The Encyclopedia of Global Human Migration, Prehistory Volume, Chapter 25*, ed. Emmanuel Ness, 204–208. Chichester: John Wiley & Sons, Ltd.
- Matisoff, James A. 2015. *The Sino-Tibetan Etymological Dictionary and Thesaurus*. Berkeley: STEDT. Retrieved 20 June 2024 from <https://stedt.berkeley.edu/dissemination/STEDT.pdf>
- Mortensen, David R. 2023. Kuki-Chin Phonology: An Overview. *Himalayan Linguistics* 22.1: 5–35.
- Peterson, David. 2017. On Kuki-Chin subgrouping. *Sociohistorical linguistics in Southeast Asia: New horizons for Tibeto-Burman studies in honor of David Bradley*, ed. Picus Sizhi Ding and Jamin Pelkey, 189–209. Leiden: Brill. Retrieved July 10, 2025, from [https://doi.org/10.1163/9789004350519\\_012](https://doi.org/10.1163/9789004350519_012)
- Post, Mark William. 2007. A Grammar of Galo. Doctoral dissertation, La Trobe University, Bumdoora.
- Post, Mark W. and Robbins Burling. 2017. The Tibeto-Burman languages of Northeast India. *The Sino-Tibetan Languages*, ed. by Graham Thurgood and Randy J. LaPolla, 213–242. London: Routledge.

- Saikia, Anujeema and Kumari Mamta. 2020. A comparative analysis of Kuki-Chin numeral system. *Working Papers on Linguistics and Literature XIV*. 90–97.
- Singh, Elangbam Sharatkumar. 2014. Numeral System in Hmar. *International Journal of English and Education* 3.3:187–194.
- Singha, K. Amrita. 2022. Number in Khelma. *Journal of English Language and Literature* 9.1:40–47.
- Suantak, Khawlsongkim. 2013. Vaiphei phonetics phonology and morphology: a descriptive study. Shillong: Doctoral dissertation, North-Eastern Hill University, Sillong. Retrieved October 28, 2024. Retrieved from <https://shodhganga.inflibnet.ac.in/handle/10603/169680>
- Suantak, Khawlsongkim, Marc Allasonnière-Tang, Eugene Chan, Anthony Chi-Pin Hsu & One-Soon Her. 2026. A GIS View of the Word Orders of Numeral Bases and Numeral Classifiers in Kuki-Chin Languages. *Digital Scholarship in the Humanities*. <https://doi.org/10.1093/llc/fqaf133>
- Thurgood, Graham & Randy J. LaPolla. 2017. *The Sino-Tibetan languages*. London: Routledge.
- Vittrant, Alice and Marc Tang. 2021. 31 Classifiers in Southeast Asian languages. *The Languages and Linguistics of Mainland Southeast Asia: A comprehensive guide*, ed. by Paul Sidwell and Mathias Jenny, 733–772. Berlin, Boston: De Gruyter Mouton.
- Wamsley, James C. 2019. Demonstratives in the Noun Phrase Structure of Hakha Chin. *Indiana Working Papers in South Asian Languages and Cultures* 1.1: 1–17.
- Zakaria, Muhammad. 2024. Evidence of a contact-induced change: Relative-correlative clause in a South Central (Kuki-Chin) language of the Tibeto-Burman branch in Bangladesh. *Asian Languages and Linguistics* 5.1:132–77.