

# Structure of numerals and classifiers in Chinese

## Historical and typological perspectives and cross-linguistic implications

One-Soon Her

National Chengchi University

It is controversial whether a classifier (C) or measure word (M) in Chinese forms a constituent first with Num (numeral) or N in a [Num C/M N] phrase. This paper reviews evidence for the [Num C/M] constituency from modern Chinese and then provides evidence from historical and typological perspectives. Under the [Num C/M] constituency, not only the C/M word orders attested in Chinese history, but also all those attested elsewhere, can be straightforwardly accounted for by the head parameter, while such simplicity is unattainable under the [C/M N] constituency. In addition, fresh evidence is obtained from the internal word order within a complex numeral; e.g. *san-shi* '30' is base-final, with *n* (3) and base (10) entering into a multiplicative function,  $3 \times 10$ . The same multiplicative function exists between Num and C/M, e.g. *san-duo hua* '3 C flower' =  $3 \times 1$  flower, and *san-da hua* '3 dozen flower' =  $3 \times 12$  flower. C/M and bases are thus unified as multiplicands, an insight further supported by the consistent correlation between the base-final order and the C/M-final order throughout the history of Chinese. A closer examination of the 103 classifier languages in Greenberg (1990[1978]) further reveals that, among the 52 languages whose numeral systems and C/M word orders can be obtained, the synchronization between the numeral base and C/M is nearly universal. The base-C/M unification as multiplicands and base-C/M synchronization in word order strongly suggest that Num and C/M form a single constituent.

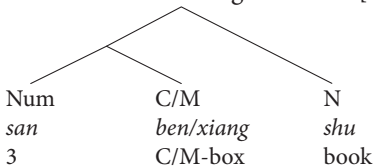
**Keywords:** classifier, measure word, constituency, numerals, multiplication, head-parameter

## 1. Introduction

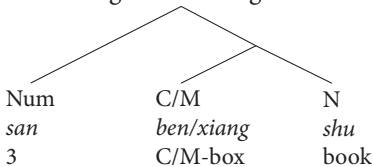
In a Chinese phrase formed by a numeral (Num), a classifier (C) or measure word (M), and a noun (N), i.e. [Num C/M N], C/M occupy the same position, as in (1a) and (1b), respectively. Whether this phrase involves a left-branching or right-branching structure, as shown in (2a) and (2b) respectively, or both structures are in fact required, has been a rather contentious issue.

- (1) a. 三 本 書 (Classifier, C)  
*san ben shu*  
 3 c book  
 '3 books'
- b. 三 箱 書 (Measure word, M)  
*san xiang shu*  
 3 M-box book  
 '3 boxes of books'

- (2) a. Unified left-branching structure: [Num C/M] constituency



- b. Unified right-branching structure: [C/M N] constituency



Note that (2a) and (2b) are only meant to be schematic, with phrasal labels and many other details left out intentionally. Greenberg (1990[1975]:227), having examined a large number of classifier languages, claims that a numeral universally forms a unit with a classifier first, as in (2a). Within Chinese linguistics, while some works assume Greenberg's approach, most of the recent formalist accounts adopt (2b) instead. While actual accounts in the literature have more elaborated structures than (2a) and (2b), the aim of this paper is to demonstrate that the left-branching approach is on the right track, and we shall thus avoid being tangled up in the structural details of specific accounts.

The paper is organized into seven sections. Section 2 first reviews the similarities and differences between C and M from semantic, structural, and mathematical perspectives. Section 3 then presents the three approaches to the constituency

of C/M phrase [Num C/M N], i.e. left-branching, right-branching, and split accounts. Section 4 discusses how C/M's common properties favor the left-branching approach. Section 5 turns to C/M's word order typology and demonstrates how C/M's left-branching structure offers the simplest explanation. We shall first focus on C/M's word order variation in the history of Chinese and then on C/M's word order typology in general. Section 6 offers fresh evidence favoring the [Num C/M] constituency obtained from the significant synchronization between the word order of C/M (e.g. *san-da* '3 dozen' is C/M-final, with Num (3) and C/M (*da*) entering into a multiplicative function,  $3 \times 12$ ), and that of the *base* in a complex numeral (e.g. *san-shi* '30' is base-final, with *n* (3) and base (10) likewise entering into a multiplicative function,  $3 \times 10$ ). This synchronization is consistent throughout the history of Chinese. Further research on 52 classifier languages in Greenberg (1990[1978]) reveals that the base-final-to-C/M-final and base-initial-to-C/M-initial synchronization is nearly universal with only one possible exception. Section 7 concludes the paper.

## 2. Differences and similarities between C and M

This section demonstrates how C/M behave similarly and also differently. §2.1 first pinpoints C/M's semantic differences; §2.2 then presents how mathematically C/M can be unified under the concept of multiplicand but differ in their respective value. §2.3 discusses the consequent differences in scope phenomena. C/M's common properties are then summarized in §2.4.

### 2.1 Semantic distinction between C/M

It has been widely observed that C/M differ in their semantic contribution to the nominal phrase. Greenberg (1990[1974]:201) states that C's are redundant in translation into a non-classifier language like English, but M's are not. W. Li (2000:1117) offers a more accurate description that C is semantically redundant in [Num C/M N], as it does not contribute any semantic content that the noun does not already have. Thus, crucially, a C does have its semantic content, but it is semantically redundant in the context of [Num C/M N]. M is not (e.g. Tai & Wang 1990; Croft 1994; Peyraube 1998; Cheng & Sybesma 1999; and many others). Her (2012b) proposes a simple formulation in set-theoretic terms to capture the semantic distinction between C/M: i.e. properties denoted by C constitute a subset of those denoted by N, which is not true for M. Thus, in (1a), the C 本 *ben* and the N 書 *shu* 'book' are compatible in that being a bound volume, which is the denotation of *ben*, is an inherent property of a book. In contrast, the measure word

箱 *xiang* ‘box’ in (1b) adds information to the noun, i.e. the books in question are contained by a box and/or their quantity fills a box.

## 2.2 C/M’s mathematical properties: convergence and divergence

In (3) below, there are four different C’s, all with the same numeral and noun. Crucially, in spite of the different C’s, the denotations of these four nominal phrases are the same, i.e. *three eggs*. This fact confirms that, semantically, each C is redundant in [Num C/M N] and thus the denotation of each of the four phrases can be composed from Num and N, with or without C. From a mathematical perspective, Au Yeung (2005, 2007), Yi (2009, 2011), and Her (2012a, 2012b) propose that this fact can only be derived by viewing each C in (3) as a multiplicand with the value of 1. In other words, the implicit mathematical operation between Num and C is multiplication, [*multiplier*  $\times$  *multiplicand*], where the multiplicand C is 1.<sup>1</sup>

- (3) a. 三 個 雞 蛋 ( $3 \times 1$  egg = 3 eggs; *ge* = 1)  
       *san ge jidan*  
       3   c   egg  
       ‘3 eggs’
- b. 三 粒 雞 蛋 ( $3 \times 1$  egg = 3 eggs; *li* = 1)  
       *san li jidan*  
       3   c   egg  
       ‘3 eggs’
- c. 三 顆 雞 蛋 ( $3 \times 1$  egg = 3 eggs; *ke* = 1)  
       *san ke jidan*  
       3   c   egg  
       ‘3 eggs’

This idea of C as ‘times one’ is stated explicitly by Greenberg (1990[1972]:172): “all the classifiers are ... merely so many ways of saying ‘one’ or, more accurately, ‘times one.’” Her (2012a, 2012b) takes this idea further and suggests that [Num C/M] in general can be seen as ‘times *x*’, where all Cs assign *x* the same numerical value of 1, but each M assigns *x* a unique value, which can be numerical or non-numerical, indeed anything except the value of 1. To put it simply, Cs are necessarily 1, Ms are not. Here are three examples of M.

1. Using *san da meiguihua* or *three dozen roses* as an example, in the equation  $3 \times 12$  roses = 36 roses, 12 is the multiplicand, representing the number in a group, group being *dozen* in this case, and 3 is thus the multiplier, referring to the number of groups. Even though the order between the two is irrelevant mathematically, as  $3 \times 12$  roses = 12 roses  $\times$  3, we shall see in §4 that it does matter linguistically for C/M and numerals.

- (4) a. 三打雞蛋 (3 × dozen egg ≠ 3 eggs; dozen ≠ 1)  
       san da jidan  
       3 M-dozen egg  
       ‘3 dozens of eggs’
- b. 三磅雞蛋 (3 × pound egg ≠ 3 eggs; pound ≠ 1)  
       san bang jidan  
       3 M-pound egg  
       ‘3 pounds of eggs’
- c. 三箱雞蛋 (3 × box egg ≠ 3 eggs; box ≠ 1)  
       san xiang jidan  
       3 M-box egg  
       ‘3 boxes of eggs’

Unlike C’s uniform numerical value of 1, each M has its own unique mathematical value; consequently, the three phrases have drastically different denotations, due solely to their different M’s. Crucially, even though each M can be represented *redundantly* as 1M, 1M is not 1. For example, in (4a), [3 × dozen egg] (= [3 × 1 dozen egg]), is not the same as [3 × 1 egg]. Note also that, even though the actual value of an M may *accidentally* be 1, C’s value is *necessarily* 1. For example, in (4b), *three pounds of eggs* may accidentally be three eggs with each one weighing exactly one pound. This does not make *pound* a C. Therefore, however precise or imprecise the value of an M, be it number, weight, volume, size, time, height, length, or monetary value, the crucial point is that mathematically it is *not necessarily* 1. In contrast, a C – regardless of the particular semantic feature or features it may highlight, be it humanness, animacy, shape, or function – is *necessarily* 1 mathematically.

In short, mathematically C/M converge in being the multiplicand with Num as the multiplier; [Num C/M] thus shares exactly the same internal structure of a high number round figure, e.g. 三十 *sanshi* ‘30’ can be seen as [3×10], 三百 *sanbai* ‘300’ as [3×100]. However, C/M diverge in their precise mathematical value: C is necessarily 1, but M is not.

### 2.3 C/M's differences in scope phenomena

The semantic and mathematical distinctions have consequences in scope phenomena, as noted in Her & Hsieh (2010).<sup>2</sup> As shown in (5b), the pre-C adjective scopes over C as well as N.<sup>3</sup> In contrast, the pre-M adjective in (5a) does not scope over N.

- (5) a. 一大箱 蘋果 (Her & Hsieh 2010 (13a))  
*yi da xiang pingguo*  
 1 big M-box apple  
 '1 big box of apples'  
 ≠  
 一箱 大蘋果  
*yi xiang da pingguo*  
 1 M-box big apple  
 '1 box of big apples'
- b. 一大顆 蘋果 (Her & Hsieh 2010 (13b))<sup>4</sup>  
*yi da ke pingguo*  
 1 big C apple  
 '1 big apple'  
 =

2. An anonymous reviewer asks how the consequences are derived from the mathematical distinction but not just from the semantic distinction. The two distinctions are interrelated as the two sides of a coin. As a multiplicand 1, C is semantically redundant and transparent; M, with its mathematical value being anything except 1, cannot be semantically redundant and transparent.

3. Note that traditionally it has been claimed that only M, not C, can be modified by a bare adjective (e.g. Chao 1968; Tai & Wang 1990; Cheng & Sybesma 1998, 1999). This is quite simply incorrect, as convincingly demonstrated in many recent works, e.g. Hsieh 2008, Her & Hsieh 2010, Zhang 2011, and Li 2011. Note also that Cheng & Sybesma's (1999) two-structure account for C/M is based on this incorrect observation and another, also incorrect, observation that only M, not C, allows *-de* insertion, which will be discussed momentarily.

4. X. Li (2009: 118) disputes this fact with the following (apparent) counter-example:

- (i) Wo chi-le yi da tiao xiao huanggua.  
 I eat-PERF one big CL small cucumber  
 'I ate a small cucumber, which is big (for my stomach).'

X. Li argues that the pre-C adjective *da* 'big' does not modify the noun cucumber in (i); otherwise, (i) one would be self-contradictory, as the cucumber cannot be big and small at the same time. However, crucially, *xiao huanggua* is a particular kind of cucumber and the proper translation is not 'small cucumber', but 'gherkin' or 'Japanese cucumber'. Thus, to me and my informants, the only possible interpretation of (i) is 'I ate a big Japanese cucumber.'

一顆大蘋果  
*yi ke da pingguo*  
 1 C big apple  
 ‘1 big apple’

Such scope phenomena are consequences of C/M’s different semantic and mathematical properties: C, being semantically and mathematically redundant in [Num C/M N], does not constitute a barrier for adjectives to scope over N; M, on the other hand, being semantically and mathematically substantive, does constitute a barrier.

The same is true for numeral quantification. Again, C, being semantically and mathematically redundant in [Num C/M N], does not constitute a barrier for Num to scope over N; thus, in (6a), all C’s can be omitted for stylistic purposes and the truth value remains the same. In the Beijing dialect of Mandarin Chinese especially, Cs are much more freely omitted than in other dialects (e.g. Chu 1994; Wang 2004; Ma 2011; Her 2012a), and in many Tibeto-Burman languages with a less developed classifier system, e.g. Tibetan, Jingpho, and Cuona Monpa, the use of C is often optional (e.g. Jiang 2006:18). M, on the other hand, being semantically and mathematically substantive, does constitute a barrier. In (6b), numeral quantification only scopes over M, which thus cannot be omitted without changing the meaning of the sentence.

- (6) a. 五(張) 餅 二(條) 魚 餵飽 五千 (個) 人  
*wu (zhang) bing er (tiao) yu weibao wuqian (ge) ren*  
 5 C loaf 2 C fish feed-full 5000 C person  
 ‘5000 people were fed with 5 loaves and 2 fish.’
- b. 五\*(箱) 餅 二\*(籃) 魚 餵飽 五千 \*(組) 人  
*wu\*(xiang) bing er \*(lan) yu weibao wuqian \*(zu) ren*  
 5 M-box loaf 2 M-basket fish feed-full 5000 M-group person  
 ‘5000 groups of people were fed with 5 boxes of loaves and 2 baskets of fish.’

## 2.4 C/M’s common properties

As noted in Her (2012b), other than the differences discussed above, C/M behave the same and six properties are given. An additional property that involves the use of *ban* ‘half’ and *duo* ‘more’ is also noted in Hsieh (2008:45–46).

### (7) C/M Common Properties

1. [Num C/M N]: C/M mutually exclusive but [Num C/M] allows conjunction
2. [Num C/MN]: allow N ellipsis and N extraction
3. [Num ~~C/M~~N]: allow C/M ellipsis but [C/M N] allows no extraction

4. [Num C/M-*de* N]: allow *-de* insertion and [Num C/M] allows conjunction
5. [~~Num~~C/M N]: allow Num ellipsis, if Num = 1
6. [Num C/M-*ban/duo* N]: allow *-ban/duo* insertion and [C/M-*ban/duo*] allow conjunction

The first property is that C/M occupy the same position; they are mutually exclusive and allow no conjunction, as in (8a) (e.g. Jiang 2012:11–12). This indicates that C/M are of the same category and have the same structure in [Num C/M N]. Furthermore, [Num C/M] can be conjoined, as in (8b)–(8c).

- (8) a. \*一本 (和/或) 箱 書  
       *yi ben (han/huo) xiang shu*  
       1 c and/or box book
- b. 一箱 或 兩 箱 (的) 書  
       *yi xiang huo liang xiang (de) shu*  
       1 m-box or 2 m-box DE book  
       ‘one or two boxes of books’
- c. 一百 本 或 兩百 本 (的) 書  
       *yi-bai ben huo liang-bai ben (de) shu*  
       100 c or 200 c DE book  
       ‘one hundred or two hundred books’

The second property is that C/M both allow N ellipsis, when N is recoverable from discourse, as in (9a)–(9b), as well as N extraction, but not [C/M N], as in (9c) and (9d), respectively.

- (9) a. 他有 三 箱 書，我也 有 三 \*(箱) *e*  
       *ta you san xiang shu, wo ye you san xiang e*  
       he have 3 m-box book I also have 3 m-box  
       ‘He has 3 boxes of books, I also have 3 boxes (of books).’
- b. 他有 三 本 書，我有 三 \*(本) *e*  
       *ta you san ben shu, wo you san ben e*  
       he have 3 c book I have 3 c  
       ‘He has 3 books, I have 3 (books).’
- c. 書，他有 三 本 *e*  
       *shu ta you san ben e*  
       book he have 3 c  
       ‘Books, he has 3.’
- d. \*本 書，他有 三 *e*  
       *ben shu ta you san e*  
       c book he have 3



The third property is that C/M also allow themselves to be omitted, but only when Num is a high number round figure, as in (10a)–(10b). However, though it looks like [C/M N] is deleted, it cannot be extracted at the same time, as in (10c).

- (10) a. 他有 一百本 書，我有 三百/\*三 你 有  
*ta you yibai ben shu, wo you sanbai/\*san ni you*  
 he have 100 c book I have 300/3 you have  
 幾百/\*幾？  
*ji-bai/\*ji?*  
 how-many-hundred/how-many  
 ‘He has 100 books, I have 300/3, how many (hundred) do you have?’
- b. 他有 一百箱 書，我有 三百/\*三 你 有  
*ta you yibai xiang shu, wo you sanbai/\*san ni you*  
 he have 100 m-box book I have 300/3 you have  
 幾百/\*幾？<sup>5</sup>  
*ji-bai/\*ji?*  
 how-many-hundred/how-many  
 ‘He has 100 boxes of books, I have 300/3 (boxes of books), how many (hundred) (boxes of books) do you have?’
- c. \*箱 書，我有 三百  
*xiang shu, wo you sanbai*  
 m-box book I have 300

Note that a high round figure like *sanbai* ‘300’ is [ $3 \times bai$ ], and the [Num C/M] sequence, e.g. *san ben* [ $3 \times 1$ ], involves the same multiplication relation. The base *bai* (a multiplicand) bears a resemblance to C/M (also a multiplicand) and can thus be interpreted as a C/M cognitively. But if the numeral is not analyzable as having the internal structure of [multiplier  $\times$  multiplicand], e.g. *san* ‘3’ and *ji* ‘how many’, C/M cannot be omitted.

The right-branching analysis misses this generalization between numeral bases and C/M and must stipulate this property of high round numbers. Furthermore, the fact that C/M and N are both deleted does not suggest the [C/M N] constituency because it cannot be extracted when the Num is a high round number, as shown in (10c).

The fourth property is that C/M both allow *-de* insertion. (11a) is from the Sinica Corpus, cited in Hsieh (2008), and (11b) is equally good. Here the [Num C/M] sequence can be conjoined, as in (11c).

5. Note that if *ji* is replaced with *duoshao*, the question is then well-formed. That is because unlike *ji*, which requires a C/M, *duoshao* does not. The former is numerical, while the latter refers to quantity.

- (11) a. 五百萬 隻 的 鴨子  
*wubaiwan zhi de yazi*  
 5-million c DE duck  
 '5 million ducks'
- b. 五百萬 箱 的 鴨子  
*wubaiwan xiang de yazi*  
 5-million M-box DE duck  
 '5 million boxes of ducks'
- c. 五 箱 還是 五 籃 的 鴨子  
*wu xiang haishi wu lan de yazi*  
 5 M-box or 5 M-basket DE duck  
 '5 boxes (of ducks) or 5 baskets of ducks'

Note that traditionally it has been claimed that only M allow *-de* insertion (e.g. Chao 1968; Tai & Wang 1990; Cheng & Sybesma 1999). This is incorrect, as convincingly demonstrated in many recent works, e.g. Hsieh (2008), Her & Hsieh (2010), Zhang (2011), and Li (2011), with solid corpus data as well as judgment by native speakers. It is true, as noted by Her & Hsieh (2010), that while *-de* is generally acceptable in [Num M-*de*], numerals with greater computational complexity are preferred with [Num C-*de*]. However, it is crucial to our task at hand to note that all numerals, regardless of the internal computational complexity, must have the same structure in [Num C/M N].<sup>6</sup>

The fifth property is that C/M also allow Num to be elided, but only when Num's value is 1. Mathematically, the explanation is simple:  $(n \times m) = m$ , iff  $n = 1$ ; the multiplier 1 is thus redundant.

- (12) a. 這 (一) 本 書  
*zhe (yi) ben shu*  
 the 1 c book  
 'This book'

6. An anonymous reviewer suggests that M with and without *de*-insertion has a slight difference in meaning; e.g. [3 cup *de* water] is about the volume, while [3 cup water] is about the individual units. However, Li (2014) demonstrates convincingly that C and M do not differ in this regard and thus argues for a unified account for C/M. The reviewer further suggests that such a meaning difference is ideally derived structurally. However, the reviewer's conclusion that this difference cannot be accounted for structurally if [Num C/M] is a constituent is a bit hasty. In fact, *de*-insertion is more a problem for the right-branching account, where [C/M N] is a constituent and *de*-insertion must interrupt this constituency. In her right-branching account, Li (2014) thus resorts to post-syntax phonological insertion of *de*. In a left-branching account, this meaning difference is easily accounted for by the presence and absence of the complementizer *de* (e.g. Hsieh 2008; Her 2012b).

- b. 這 (一) 箱 書  
*zhe (yi) xiang shu*  
 the 1 M-box book  
 'This box of books'

The sixth property is that C/M both allow *ban* 'half' or *duo* 'more' to be inserted following C/M, as the examples in (13) and (14) from Hsieh (2008) show.

- (13) a. 兩 個 半 鐘頭 (Hsieh 2008:46 (53a))  
*liang ge ban zhongtou*  
 2 C half hour  
 '2 and a half hours'
- b. 一個 多 鐘頭 (Hsieh 2008:46 (54a))  
*yi ge duo zhongtou*  
 1 C more hour  
 'more than 1 hour'
- (14) a. 兩 瓶 半 可樂 (Hsieh 2008:46 (53b))  
*liang ping ban kele*  
 2 M-bottle half Coke  
 '2 and a half bottles of Coke'
- b. 一瓶 多 可樂 (Hsieh 2008:46 (54b))  
*yi ping duo kele*  
 1 M-bottle more Coke  
 'more than 1 bottle of Coke'

To summarize, C/M converge and diverge at the same time, as observed by many researchers before (e.g. Chao 1968; Tai & Wang 1990; Cheng & Sybesma 1999; Jiang 2012). The multiplicative theory of C/M nicely captures this fact; i.e. C/M converge as a multiplicand but diverge in their respective mathematical value. Consequently, C/M differ semantically in that C is semantically redundant in [Num C/M N], due to its precise but redundant value of 1 as the multiplicand. This in turn has the consequence of C allowing adjectival and numeral quantification to scope to cover N. M, on the other hand, is semantically substantive, due to its mathematical value as the multiplicand that is not necessarily 1, and thus does block its adjectives and numerals from scoping over N. C/M otherwise behave the same syntactically. C/M thus constitutes a single syntactic category formally; however, semantically, Cs and Ms form two different subcategories.

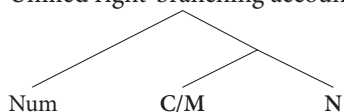
### 3. Controversy over C/M constituency

Previous syntactic accounts of the C/M construction can be divided along two factors: first, whether C/M in all languages have a uniform structure or not, and second, if there is indeed a uniform structure, whether C/M form an immediate constituent with Num or with N (Zhang 2011; Her 2012b). The former has been dubbed a ‘**unified**’ account, and the latter, a ‘**split**’ account. The three possibilities are shown schematically in (15)–(17).

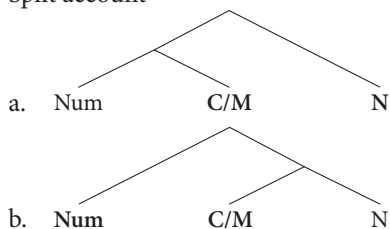
(15) Unified left-branching account



(16) Unified right-branching account



(17) Split account



The more notable unified left-branching accounts include Chao (1968) for Chinese and Greenberg (1990[1975]), where the [Num C/M] constituency is proposed to be universal. Liu (2006) put forth a derivational account based on [Num C/M] constituency for Chinese, Jingpho (Tibeto-Burman), and Yi (Tibeto-Burman). Many more support this position for Chinese, e.g. Li & Thompson (1981:105), Paris (1981:105–117), Huang (1982), Tang (1990a), Croft (1994:151), Lin (1997:419), Hsieh (2008), and Tsai (2011).

The more prominent formal derivational accounts based on [C/M N] constituency include Simpson (2005), Borer (2005), and Huang et al. (2009), all considering this structure universal. Likewise, Gu & Wu (2005) also propose a similar account for Chinese, Jingpho, and Yi. Many more advocate the same approach for Chinese languages, e.g. Au Yeung (2005), Tang (2005), Cheng & Sybesma (1998, 1999), Zhang (2009), and Li (1998). In fact, most of the more recent works within

the Chomskyan derivational framework assume that C/M are the head that take N as a complement.

Still, there are a number of accounts that do not accept a uniform structure. Watanabe (2006) contends that Cs and Ms occur in different structures in Japanese. Huang & Ochi (2011) claim that Chinese, with its dominant [Num C/M N] word order, has the base-generated [C/M N] constituent, but Japanese needs both [Num C/M] and [C/M N]. Jenks (2010) has a similar proposal, that different classifier languages may have different structures, some [Num C/M] and others [C/M N]. The issue is further complicated by Zhang's (2011, 2013) split account for Chinese C/M, where some types of C/M involve a right-branching structure, and others, left-branching, and Li's (2011) account, which likewise requires the left-branching structure for C/M's measuring reading and right-branching for the counting reading.

In a derivational account, e.g. Watanabe (2006), the left-branching structure can be easily derived from the right-branching structure. However, it is well-accepted in the current syntactic theorizing that while movement is costly, base-generation is not. Furthermore, a unified account is also simpler and thus favorable, as it allows all of C/M's common properties to be stated once in one uniform structure, while in a split account these generalizations are lost and rendered accidental.

Yet, given the semantic and mathematical distinctions between C/M and, more importantly, their consequent differences in scope phenomena, they receive a structural solution in Zhang's (2011, 2013) split account. However, a lexicalist solution is available for the unified approach; i.e. while M's are lexical elements, C's are 'semi-lexical', as argued in van Riemsdijk (1988), Kubo (1996), Vos (1999), and Löbel (2001), among others. C's thus exhibit certain 'semi-lexical' or functional characteristics, one of which is allowing modification and quantification to go beyond C and scope over N. Her (2012b) offers such a unified left-branching account within Lexical-Functional Grammar. In short, scope phenomena favor neither the unified approach nor the split approach, which means the former should be preferred for the sake of simplicity.

Next, I shall test the two options under the unified approach against each of C/M's six common properties and see which fares better.

#### 1. [Num C/M N]: C/M mutually exclusive but [Num C/M] allows conjunction

The fact of C/M's mutual exclusiveness can be easily accounted for under either approach. As for the conjunction facts, there are two accounts available, coordination between two [Num C/M] constituents or by right-node raising of N. Coordination is a simpler account, as movement is generally seen as more costly. The beauty of the left-branching approach is that it allows both accounts, i.e. co-

ordination and right-node raising, while under the right-branching approach only the more costly right-node raising is available.

## 2. [Num C/M N]: allow N ellipsis and N extraction

N ellipsis and extraction are legitimate in both structures. Yet, in the left-branching structure, the fact that [C/M N], unlike N, cannot be extracted can be accounted for straightforwardly, i.e. [C/M N] does not form a constituent.<sup>7</sup> Under the right-branching account, if Num is assumed to be in the Spec of ClP (e.g. Zhang 2009:21), [C/M N] is an intermediate projection and thus cannot be extracted. However, most right-branching accounts have Num heading its own projection, taking ClP as its complement; in such accounts, the immobility of [C/M N] is a problem. Thus, the facts here do not favor a right-branching structure where Num is a head (see Jenks 2011, §3.4 for further discussion).

## 3. [Num C/M N]: allow C/M ellipsis but [C/M N] allows no extraction

As demonstrated in §2.4, C/M ellipsis is most felicitous when Num is a high round figure, thus of the form [multiplier multiplicand]. The left-branching structure is thus favored here, where the [Num C/M] constituent is compatible with a [multiplier multiplicand] constituent. This insight would be lost in a right-branching structure.

In a formal treatment of complex numerals involving multiplication and numeral classifiers, Ionin & Matushansky (2006) deliberate between precisely two structures, the left-branching (18a) and the right-branching (18b). (Note, in (18a) Num refers to number morphology, such as singular and plural, and thus not numerals.) While they remain somewhat uncommitted, they do demonstrate that for languages with plural morphology in cardinal-containing NPs, the structure in (18a) is to be preferred.<sup>8</sup> The point to be made here is that the left-branching

7. An anonymous reviewer suggests that the fact that [Num C/M] cannot be dislocated may be evidence against its constituency, as shown in (i). However, while dislocation may be evidence for constituency, its absence is not evidence for non-constituency, because other known modifier phrases also do not allow dislocation, as shown in (ii).

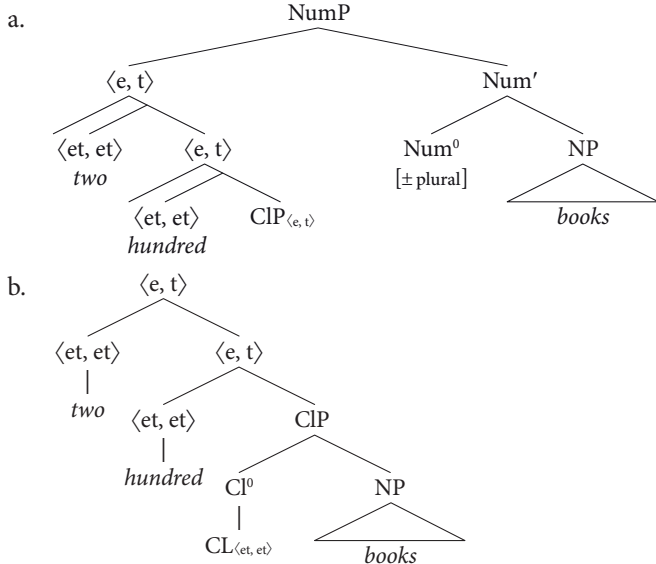
(i) \**Shi yi ben, Zhangsan kan t shu.*  
be one c Zhangsan read book  
Intended: 'It is a book that Zhangsan read.'

(ii) \**Shi hen da de, Zhangsan xihuan t fangzi.*  
be very big DE Zhangsan like house  
Intended: 'It is big houses that Zhangsan likes.'

8. As convincingly demonstrated in Her (2012a) with ample corpus data, the plural marker *-men* does co-occur with numerals, though optionally. The example below is taken from a

(18a), where [Num C/M] is a constituent, better accounts for the fact that C/M can be deleted when the preceding element is numeral base, thus a multiplicand like C/M itself.

- (18) Two structures for Classifier Phrases entertained by Ionin & Matushansky (2006)



In the right-branching (18b), though one might argue that the [C/M N] sequence can be deleted precisely because it is a constituent, such an account fails to explain why this deletion can happen only precisely when Num is a high number round figure.

#### 4. [Num C/M-*de* N]: allow *-de* insertion and [Num C/M] allows conjunction

Left-branching is favored, as it is uncontroversial that pre-*de* modifiers are constituents (e.g. C. Huang 1989; Tang 1990b:420; Zhang 2012). This is also supported by phonological evidence that [Num C/M] forms both a phonological and grammatical word (e.g. Duanmu 2000, 2005).

---

website of National Chiayi University in Taiwan. The URL is: [http://www.ncyu.edu.tw/ctedu/print.aspx?table\\_name=site\\_content&sn=4653&site\\_content\\_sn=4653](http://www.ncyu.edu.tw/ctedu/print.aspx?table_name=site_content&sn=4653&site_content_sn=4653)

- (i) 三 位 老師們 聽了， 無不 高興 萬分。  
*san wei laoshi-men ting-le wubu gaoxing wanfen*  
 3 c teacher-PL hear-ASP all happy extremely  
 'The three teachers were all extremely happy, having heard of this.'

#### 5. [Num C/M N]: allow Num ellipsis, if Num = 1

Num cannot be omitted unless it is 1. While there are grammatical and phonological restrictions, what concerns us here is that this fact can be accounted for under either approach.

#### 6. [Num C/M-*ban/duo* N]: allow -*ban/duo* insertion, thus [C/M-*ban/duo*]

This also favors left-branching, where [Num C/M] as a single constituent merges with *ban/duo* to form a coherent quantifier, which in turn receives a natural analysis, i.e. [[Num  $\times$  C/M] + *ban/duo*], one that resembles that of complex numerals; e.g.  $[[2 \times 10] + 5]$  for 二十五 *er shi wu* '25'. Yet, [Num C/M-*ban/duo*] does not form a constituent in a right-branching structure.

An anonymous reviewer suggests that semantic and categorial selections between C/M and N can be accounted for in a right-branching structure where C/M and N form a local head-complement configuration, much like the relation between a verb and its object. However, the semantic and categorial selections between C/M and N can also be easily captured in a local modifier-head configuration in a left-branching account, similar to the relation between an adjective and the noun, e.g. *female nurses/#brothers* and *colorful/#colorless colors*. Considerations of semantic and categorial selections between C/M and N thus do not favor either approach.

To summarize, the [Num C/M] constituency thus seems to enjoy some advantages. In the following two sections, further evidence will be provided to support C/M's unified left-branching structure. §4 offers evidence from the historical data on the internal structure of numerals. §5 then does the same based on historical data on word order variations.

### 4. Internal structure of numerals and C/M constituency

Given the underlying mathematics between Num and C/M as [multiplier multiplicand], it is the product of this multiplication that serves to quantify N. Thus, only the [Num C/M] constituency under the left-branching approach is compatible with its underlying mathematics (e.g. Au Yeung 2005, 2007; Her 2012a). It is certain that multiplicative numerals came before C/M in Chinese, because the decimal numeral system is already fully mature in the earliest written records, and yet C/M are rather poorly developed at the same time.<sup>9</sup> In this section, we shall

9. According to Dempsey (1995) and Matisoff (1997), numerals can be reconstructed to the Sino-Tibetan level, but LaPolla (2003:27) states explicitly that classifiers cannot.



see that, consistently throughout Chinese history, the internal structure of the [Num C/M] constituent faithfully corresponds to that of multiplicative numerals. Furthermore, relevant data from some Tibeto-Burman languages that employ the reverse [C/M Num] order, also confirm the generalizations found in Chinese.

#### 4.1 Internal word orders in multiplicative numerals and word orders of C/M in Chinese

In Comrie (2011), an extensive survey on the numeral systems of the world's languages, among the 196 languages covered, 172, or some 88%, employ a combination of multiplication and addition, as described in the quote below.<sup>10</sup> Also, in an earlier work, Comrie (2006) offers a more concise formulation, as in (19). I shall follow the terminology established in (19).

By the “base” of a numeral system we mean the value  $n$  such that numeral expressions are constructed according to the pattern  $\dots xn + y$ , i.e. some numeral  $x$  multiplied by the base plus some other numeral (The order of elements is irrelevant, as are the particular conventions used in individual languages to indicate multiplication and addition). (Comrie 2011, emphasis in original).

##### (19) General Pattern of Numeral Expressions

$(n \times \text{base}) + m$ , where  $m < \text{base}$  (Comrie 2006)

While Comrie (2011) is explicit that the linear order between  $n$  and *base* in the generalized mathematical pattern in (19) is irrelevant, he does recognize that this linear order does vary across languages. Though the majority of multiplication-based numeral systems indeed follow the order [multiplier > multiplicand], or [ $n \times \text{base}$ ], the reverse order [ $\text{base} \times n$ ] is by no means rare. Mathematically, numbers such as *three hundred* can of course be expressed either as *three hundred* or *hundred three*. I shall demonstrate that while the difference is purely notational mathematically, it is significant linguistically. Let us call this the **base-parameter**. The former type shall be referred to as **base-final**, and the latter **base-initial**. Chinese is well-known for its highly regular *base-final* decimal numeral system; e.g. 二十五 *er shi wu* ‘25’ conforms to [ $[2 \times 10] + 5$ ], and indeed often serves as a textbook example of a base-final decimal numeral system in languages.

##### (20) Base-parameter:

- a. base-final, thus [ $n \text{ base}$ ]
- b. base-initial, thus [ $\text{base } n$ ]

10. Incidentally, in Comrie's (2011) survey there are also a small number of languages that do not have the multiplication-based numeral system in (19) and use either an extended body-part system (4 languages) or have rather restricted numerals (20 languages).

Now, consider the two elements Num and C/M. Likewise, two possible orders obtain. Opposite to the familiar, consistent **C/M-final** order in Chinese and other East Asian languages, there are languages that employ a **C/M-initial** order, as we shall see momentarily. Let us call this the **C/M-parameter**.<sup>11</sup>

(21) C/M-parameter:

- a. C/M-final, thus [Num C/M]
- b. C/M-initial, thus [C/M Num]

To the best of my knowledge, Greenberg (1990[1978]:293) is the first to make the generalization that base and C/M tend to have the same word order, which he calls ‘harmonization’, and this generalization has since been totally overlooked by the research community. This generalization is restated as an implicational universal in more precise terms in (22).

(22) Synchronization between the C/M-parameter and the base-parameter:

- a. C/M-final  $\Rightarrow$  base-final
- b. C/M-initial  $\Rightarrow$  base-initial

A mathematical insight is thus available for this implicational universal. A central point this paper makes is that the base-C/M synchronization in Chinese and elsewhere is the natural manifestation of the underlying mathematics. Given that [*n base*] and [Num C/M] both function as [multiplier  $\times$  multiplicand], C/M and base, as multiplicands, should of course follow the same order.

The crucial fact, therefore, is that, although the order between a multiplier and a multiplicand is free, mathematically, it is grammatically encoded one way or the other in a language. In Chinese numerals, it has always been base-final, or [multiplier  $>$  multiplicand]. More important than word order is the issue of constituency. As an anonymous reviewer points out, mathematically, multiplication is subject to the law of associativity, i.e.  $(3 \times (12 \times \text{egg})) = ((3 \times 12) \times \text{egg})$ , thus accommodating both the left-branching structure and the right-branching structure. Contra Ionin & Matushansky (2006), He (2015) argues most convincingly that a multiplicative complex numeral in Chinese forms a constituent. In a language that encodes the multiplicative [*n base*], e.g. *san-shi* (three-ten) [ $3 \times 10$ ], as a syntactic constituent, it is only logical that it encodes the same multiplicative [Num C/M], e.g. *san-ge* (three-C) [ $3 \times 1$ ] and *san-da* (three-M<sub>-dozen</sub>) [ $3 \times 12$ ], also as a constituent.

However, in spite of the similarity between C/M and base, they belong to two different syntactic categories. C/M are not part of the numeral system and form a distinct grammatical category on their own. That many non-classifier languages

11. In the literature the two types are often referred to as ‘NC’, short for Num-C, and CN, short for C-Num (e.g. Yang 2005).

have multiplicative numerals indicates that the synchronization between base and C/M is only one way, as shown in (22).

The strict correspondence between base-final and C/M-final is not only observed in all contemporary Sinitic languages, but is also true *diachronically*.<sup>12</sup> As a prelude to the discussions in §5, (23) lists the six word orders among Num, C/M, and N, in the history of Chinese.

(23) Six Word Orders among Num, C/M, and N in History (Peyraube 1998)<sup>13</sup>

- a. Num + N
- b. N + Num
- c. N + Num + M
- d. N + Num + C
- e. Num + M + N
- f. Num + C + N

With no exception, Num precedes C/M. Likewise, the current day decimal base-final numeral system has been consistently employed since the earliest written records more than 3,000 years ago, found in oracle bone inscriptions and bronze inscriptions from the late Shang Dynasty, or around mid-13th c. BC to late 11th c. BC, known as 甲骨文 *Jiaguwen* and 金文 *Jinwen*, respectively. For example, *eight hundred* is formed by topping *hundred* with *eight*, thus 八百 (Hua 2011, chapter 216). The relation between the two is exactly that of [*eight*×*hundred*], or [*n*×*base*]. (24) is another example from Western Zhou bronze inscriptions.

- (24) 俘人                      萬    三千    八十   一人                      (Zhang 2001 (5.2839))  
       *furen*                      *wan*    *sanqian* *bashi* *yi ren*  
       captive-person 10000 3000    80    1 person  
       ‘13,081 captive persons.’

In Chinese there is thus this perfect match between base-final numerals and C/M-final nominals, regardless of the position of N, throughout history. We shall now see evidence in some typologically different, base-initial and C/M-initial, languages within the Tibeto-Burman family.

12. An anonymous reviewer rightly points out that it is paramount to make sure that in the history of the Chinese numeral system, multiplicative complex numerals have always been base-final. This is indeed the case. There is not a shred of evidence found in the literature that would indicate otherwise.

13. Peyraube (1998) in fact listed one more word order,  $N_1 + \text{Num} + N_2$ , e.g. 羌十人 *Qiang shi ren* [Qiang ten person] ‘ten Qiang persons’, and thus seven all together. However, this construction should be seen as an instance of either (23a) [Num + N] or (23d) [N Num C]. There are essentially just six word orders. We shall discuss this in more detail in §4.

## 4.2 Numerals and C/M in Tibeto-Burman<sup>14</sup>

Jiang (2006) claims that in non-Sinitic Tibeto-Burman languages N in general precedes the numeral phrase, which, however, comes in two varieties: [C/M Num] and [Num C/M]. Thus, unlike Sinitic languages, both C/M-initial and C/M-final orders are found. The same study also finds that in C/M-initial languages, e.g. Tibetan, Monba, and Jingpho, the number of C/M is rather low, but in C/M final languages, e.g. Yi, Burmish, Karen, and Qiang languages, C/M are abundant.

In Jiang's (2006), Yang's (2005), and Chan's (2016) surveys, the majority of Tibeto-Burman languages have multiplication-based decimal numeral systems. Also, most allow only the familiar base-final order. For example, Yi, Burmese, Qiang, and Karen languages are exactly like Chinese and allow only base-final numerals and C/M-final nominals, thus [*n base*] and [Num C/M] only. However, some of the C/M-initial languages have both a base-initial order as well as a base-final order.<sup>15</sup> According to Duojie Dongzhi (2005), in Amdo Tibetan, for example, the consistent word order in an NP is C/M-initial and N-initial, thus [N C/M Num] as in (25a), and the use of C is not entirely mandatory, indicating the language's transition from a non-classifier language to a classifier language. According to data found in the website hosted by Eugene Chan (2016), numerals in Amdo lower than a thousand are consistently base-final, e.g. (25b), just like Chinese, but numerals with *thousand* or a larger base are consistently base-initial, e.g. (25c). The C/M-initial order thus still correctly predicts a base-initial order in numerals, though the base-initial order is not exclusive in the language's numeral system.

- (25) a. *ŋə (nda) hsəm*  
           person c     3  
           '3 people'  
       b. *ɣŋi wja*  
           2   hundred  
           '200'  
       c. *rtoŋ     ɣŋi*  
           thousand 2  
           '2000'

14. According to *Ethnologue* (<http://www.ethnologue.com/>), Chinese and Tibeto-Burman are the two major branches under Sino-Tibetan. However, this classical, conventional Shaferian Stammbaum view has been seriously challenged. For example, van Driem (2003, 2007, 2011), among others, has repeatedly argued that Sinitic should be a part of Tibeto-Burman and that Sino-Tibetan should thus be replaced by Tibeto-Burman.

15. This issue will be discussed further in §6.

Yang (2005) speculates that the historical developmental process is this: 1) [N (C)/M Num], then 2) [Num time-unit], Num > 1, then 3) [N (C)/M Num] & [N Num (C)/M], and then finally 4) [N Num C/M]. I shall further speculate that this development from [C/M Num] to [Num C/M] is along the parallel development of the numeral system from 1) [*base n*] only to 2) [*base n*] and [*n base*] co-existing to 3) [*n base*]. I shall, however, leave the verification of these interesting predictions to further research.

In short, while a systematic and comprehensive survey of the entire Tibeto-Burman family is beyond the scope of the paper, data from Jiang's (2006), Yang's (2005), and Chan's (2016) extensive surveys indeed confirm the typological predictions in (22). Furthermore, the correlation between base and C/M in word order also convincingly demonstrates that [Num C/M], like [*n base*] of numerals, form an immediate constituent. However, there is one snag we should deal with before wrapping up the discussions on the correlations between C/M and base.

#### 4.3 Apparent counterexamples: languages with [C/M N Num], Num = 1

The constituency of [Num C/M] in either ordering dictates that Num and C/M be adjacent and thus the two orders [C/M N Num] and [Num N C/M] should not exist. While as far as I am aware the latter has indeed not turned up in any language, a peculiar instance of the former, i.e. [C/M N Num], has presented itself in some Tai-Kadai languages, an independent family in the 16th edition of *Ethnologue*.<sup>16</sup> While in Chinese and Hmong-Mien languages (a.k.a. Miao-Yao) the order is always [Num C/M N], the variation is far greater in Tai-Kadai, or Zhuang-Dong, including [C/M N Num] and [N C/M Num], when Num is *one*, and [Num C/M N] and [N Num C/M], when Num is not *one* (Jiang 2006:16).

I take Maonan as a typical example and examine the relevant facts closely. Maonan, a Kam-Sui language in the Tai-Kadai family, spoken in Guangxi, China, is a language with a mixed C/M typology, i.e. the familiar [Num C/M N], as in (27), and the unexpected [C/M N Num], as in (26) (Zhang 2005), and a mixed base typology, i.e. [*n base*], as in (29), and [*base n*], as in (28) (Chan 2016). Similar

16. Tai-Kadai is known as Zhuang-Dong and considered a branch of Sino-Tibetan by most Chinese linguists, a view generally not shared by scholars outside of China.

facts are found in Bouyei, a Northern Group Tai language in the same family, spoken in Guilin, China.<sup>17</sup>

- (26) a.  $ai^1 zən^1$  ( $deu^{231}$ ) [C/M N Num], C/M-initial  
 c person 1  
 ‘1 person’  
 b.  $*deu^{231} ai^1 zən^1$   
 1 c person  
 ‘(intended) 1 person’
- (27) a.  $ja^{42}/sa:m^{42} ai^1 zən^1$  [Num C/M N], C/M-final  
 2/3 c person  
 ‘2/3 persons’  
 b.  $*ai^1 zən^1 ja^{42}/sa:m^{42}$   
 c person two/three  
 ‘(intended) 2/3 persons’
- (28) a.  $pek^{55}/t^hjen^{42}$  ( $deu^{231}$ ) [*base n*], base-initial  
 hundred/thousand one  
 ‘100/1000’  
 b.  $*deu^{231} pek^{55}/t^hjen^{42}$   
 one hundred/thousand  
 ‘(intended) 100/1000’
- (29) a.  $ja^{42}/sa:m^{42} pek^{55}/t^hjen^{42}$  [*n base*], base-final  
 2/3 hundred/thousand  
 ‘200/300; 2000/3000’  
 b.  $*pek^{55}/t^hjen^{42} ja^{42}/sa:m^{42}$   
 hundred/thousand 2/3  
 ‘(intended) 200/300; 2000/3000’

However, there are several crucial facts to be noted, according to the data and descriptions in Zhang (2005), Chan (2016), and Jiang (2007). First, the numeral *one* in Maonan has three different forms,  $tɔ^{231}$ ,  $jɪt^{55}$ , and  $deu^{231}$ , and only  $deu^{231}$  can be used in the manner described in (26)–(29), i.e. in [C/M N Num] and [*base n*]. Second, the [C/M N Num] order is highly restricted and in fact is only allowed when the numeral is exactly  $deu^{231}$  ‘1’, and likewise the [*base n*] order is allowed

17. Some of the languages in this family, Thai for example, have [N Num C/M], where Num > 1, and [N C/M Num], instead of the problematic [C/M N Num], where Num = 1 (Jiang 2006:46). Even though the former word order does not violate the universals proposed in this paper, the discussions and conclusion in this section regarding Maonan should apply to languages like Thai as well. Thus, if the [C/M N Num] word order does not exist in Maonan, as we shall argue in this section, the [N C/M Num] order is likewise to be dismissed in languages like Thai.

only when the base is *hundred* or *thousand* or larger and the multiplier *n* is exactly  $\text{deu}^{231}$  ‘1’. Third, in both cases  $\text{deu}^{231}$  is optional and can be omitted. Fourth, adjectives in Maonan are post-nominal, as in (30), just like  $\text{deu}^{231}$ .

- (30) a.  $ai^1 z\grave{a}n^1 \quad \text{voj}^1$   
           c person tall  
           ‘the tall person’  
    b.  $t\check{\jmath}^2 n\grave{o}k^8 \text{da:i}^2$   
           c bird good  
           ‘the good bird’

Based on these facts, I concur with Jiang (2007) that  $\text{deu}^{231}$  in [C/M N  $\text{deu}^{231}$ ] is not Num at all. Jiang (2007) further proposes that  $\text{deu}^{231}$  is an adjective. However, I contend that  $\text{deu}^{231}$  is in fact a determiner marking singularity and indefiniteness, much like the English *a*. Our view seems preferable over Jiang’s (2007) because of the following facts. First, genuine adjectives must always appear closer to N than  $\text{deu}^{231}$ , as in (31). This behavior of  $\text{deu}^{231}$  is consistent with the category D, which takes an NP as complement and thus must either precede or follow the entire NP. Second, a determiner marking definiteness cannot co-occur with  $\text{deu}^{231}$ , as in (32). The fact that  $\text{deu}^{231}$  is mutually exclusive with known determiners in the same position also indicates that it is of the category D.

- (31) a.  $ai^1 z\grave{a}n^1 \quad \text{voj}^1 \text{deu}^{231}$   
           c person tall INDEF. SG.  
           ‘a tall person’  
    b.  $*ai^1 z\grave{a}n^1 \quad \text{deu}^{231} \quad \text{voj}^1$   
           c person INDEF. SG. tall
- (32) a.  $ai^1 z\grave{a}n^1 \quad \text{voj}^1 \text{ka}^2$   
           c person tall DEF. SG.  
           ‘that tall person’  
    b.  $*ai^1 z\grave{a}n^1 \quad \text{deu}^{231} \quad \text{ka}^2$   
           c person INDEF. SG. DEF. SG.  
    c.  $*ai^1 z\grave{a}n^1 \quad \text{ka}^2 \quad \text{deu}^{231}$   
           c person DEF. SG INDEF. SG.

I thus do not consider examples like Maonan  $\text{deu}^{231}$  genuine counterexamples to the claim that [Num C/M] form an immediate constituent. Viewing  $\text{deu}^{231}$  as either adjective or determiner explains why it does not behave like other numerals;

but this does not explain why the other forms of *one* besides *deu*<sup>231</sup>, i.e. *tɔ*<sup>231</sup> and *jit*<sup>55</sup>, also do not behave like other numerals and appear before C/M.<sup>18</sup>

- (33) a. *ja*<sup>42</sup>/*sa:m*<sup>42</sup> *ai*<sup>1</sup> *zən*<sup>1</sup>  
           2/3           c person  
           ‘2/3 persons’  
   b. *\*tɔ*<sup>231</sup>/*\*jit*<sup>55</sup> *ai*<sup>1</sup> *zən*<sup>1</sup>  
           one/one c person  
           ‘1 person’ (intended)
- (34) a. *ai*<sup>1</sup> *zən*<sup>1</sup>  
           c person  
           ‘the person’  
   b. *tɔ*<sup>2</sup> *nɔk*<sup>8</sup>  
           c bird  
           ‘the bird’

I contend that in (34) there is a silent Num *one* before C, as the interpretation of (34) can only be singular. Languages differ typologically regarding the numeral *one* in the context of [Num C/M]. Chinese, for example, also allows the numeral *one* to be silent in certain contexts.

- (35) a. 這 (一) 本 書  
           *zhe (yi) ben shu*  
           the 1 c book  
           ‘This (1) book’  
   b. 這 (一) 箱 玫瑰  
           *zhe (yi) xiang meigui*  
           the 1 M-box rose  
           ‘This (1) box of roses’

In (35), Chinese and English alike, the interpretation is the same with or without the overt numeral *one*, which is entirely redundant as a multiplier mathematically and is thus also redundant semantically. Typologically, languages like Chinese and English *allow* this numeral *one* to be omitted, while languages like Maonan *require* it to be silent.

18. Part of the oddity in structural variety that these Tai-Kadai languages allow in the C/M construction may be explained by the heavy borrowing of Sinitic numerals and C/M system into the underlying native structures. The three words for *one* are thus of different origins: *jit*<sup>55</sup> is a borrowing from Middle Chinese *\*jit* or Old Chinese *\*ɣit* ‘one’; *tɔ*<sup>231</sup> is, on the other hand, native in Kam-Sui; and *deu*<sup>231</sup>, though likewise native, is indeed not a numeral and should be glossed as *single* or *only one*. This lexical item is reconstructed as Proto-Tai *\*diəuw* A (> Siamese *diəw* A1, Longzhou *dei* B1, etc.) by Pittayaporn (2009:358).



To summarize, I have demonstrated in this section that in Chinese the [Num C/M] constituency and word order are consistent with the internal structure of numerals, i.e. [*n base*] and this parallel internal structure and word order between C/M and numerals have existed in the last three thousand years. Furthermore, in some of the typologically different Tibeto-Burman languages, where both C/M-final and base-final and C/M-initial and base-initial word orders are found, C/M and numerals likewise share the same multiplication-based internal structure and word orders. The only apparent exception is the existence of the [C/M N Num] word order, where Num must be one specific form of the numeral *one*. Yet, upon closer examination, here *one* is more likely to be a determiner and not a numeral. The insights gained from the typology of C/M ordering and base ordering in numerals support the unified left-branching approach where [Num C/M] always form an immediate constituent.

## 5. Evidence from word order variations in history

In this section I shall find support for the [Num C/M] constituency in the word order variations in the history of Chinese. Even though C/M already appeared in the earliest oracle bone inscriptions, it took more than a millennium for C/M to develop fully and become an independent grammatical category. The history of Chinese can be divided into four periods: *Shanggu Hanyu*, or Old Chinese, a.k.a. Archaic Chinese, (500 BC–200 AD, Shang to Han), *Zhonggu Hanyu*, or Middle Chinese, (201–1000 AD, end of Han to end of Tang), *Jindai Hanyu*, or Pre-Modern Chinese, (1001–1900, end of Tang to mid-Qing), and *Xiandai Hanyu*, or Modern Chinese, (mid-Qing to the present) (e.g. Sun 1996:3). I shall now sketch the development of C/M in terms of these four stages. Note that the focus of this sketch is C/M's word order variations only; thus, there will be no in-depth discussions about the motivation or origin of any of the particular constructions or the pragmatic or grammatical functions of any of the word orders.<sup>19</sup>

In Archaic Chinese in oracle bone and bronze inscriptions, the predominant structure for counting objects is [Num N], as in (36a), followed by [N Num], as in (36b). While genuine C's are only in their embryonic stage at best, with at most half a dozen rudimentary C's in the [N Num C] construction, as in (37), there

---

19. The reader is referred to Behr (2009) for an excellent overview of the origin of C/M's and the various nominal constructions in Archaic Chinese and Yang-Drocourt (2004) for a meaningful discussion on these two important issues: how and when C/M became obligatory in Chinese and why and how the earlier predominant [N Num C/M] order was taken over by the contemporary [Num C/M N] order.

are a number of M's, including container and standard measures, as in (38a), and collective measures (38b) appearing in the [N Num M] construction (e.g. Huang 1964; Behr 2009).

- (36) a. 五 人 一 牛 (Hu 1983 (01060))  
*wu ren yi niu*  
 5 person 1 ox  
 '5 persons and 1 ox'
- b. 獲 鳥 二百 十二, 兔 一 (Hu 1983 (41802))  
*huo niao erbai shier tu yi*  
 capture bird 200 12 hare 1  
 'captured 212 birds and 1 hare'
- (37) a. 馬 四 匹 (Zhang 2001 (3.754))  
*ma si pi*  
 horse 4 c  
 '4 horses'
- b. 俘 車 三十 輛 (Zhang 2001 (5.2839))  
*fu che sanshi liang*  
 capture chariot 30 c  
 'captured 30 chariots'
- (38) a. 其 新 蒸 鬯 二 升 一 (Zhang 2001 (30973))  
*qi xin zheng chang er sheng yi*  
 his new steam millet-brand 2 m-sheng-pint 1  
 'his 2.1 sheng-pint of newly steamed millet-brand'
- b. 俘 車馬 五 乘 (Zhang 2001 (5.2779))  
*fu chema wu cheng*  
 capture carriage-horse 5 pair  
 'captured 5 pairs of carriage-horses'

Also, a well-recognized typological feature at this time is the extensive use of the so-called 'echo classifiers' (e.g. Jiang 2006:106), also referred to as 'identical classifiers', 'repeaters', 'self-classifiers' and 'auto-classifiers' (Aikhenvald 2000:103), as shown in (39).

- (39) 俘 牛 三百 五十五 牛, 羊 二十八 羊 (Zhang 2001 (5.2839))  
*fu niu sanbai wushiwu niu yang ershiba yang*  
 capture ox 300 55 ox goat 28 goat  
 'captured 355 oxen and 28 goats'

Within the Sino-Tibetan languages, it has been noted often that this [N<sub>1</sub> Num N<sub>1</sub>] echo construction is a prelude to the [N Num C/M] construction and is thus

found in the history of some of the classifier languages, e.g. Chinese (Y. Li 2000). A similar and related construction is  $[N_1 \text{ Num } N_2]$ , where the second  $N$  is less constrained and need not be identical to  $N_1$ ; rather,  $N_2$  names an instance of  $N_1$  and thus in this sense functions also like a  $C$ . (40) is an example from oracle bones.

- (40) 羌 三 人 (Hu 1983 (336))  
*qiang san ren*  
 Qiang 3 person  
 ‘3 Qiang persons’

Even though both the echo classifier and  $N_2$  in the  $[N_1 \text{ Num } N_2]$  construction do have the same functions as  $C$ ’s, it is controversial whether syntactically they can be seen as genuine  $C$ ’s (Jiang 2006:109). For our purpose here, however, we can either treat them as simply  $N$ ’s, and thus  $[N_1 \text{ Num } N_1]$  and  $[N_1 \text{ Num } N_2]$  are just cases of  $[\text{Num } N]$ . Or we can view them as  $C$ ’s, and thus they are cases of  $[N \text{ Num } C]$ .

Both constructions,  $[N_1 \text{ Num } N_1]$  and  $[N_1 \text{ Num } N_2]$ , however, all but disappeared during the Han Dynasties (206 BC–220 AD), giving way to a sharp rise of genuine  $C$ ’s. Furthermore, the now more than fifty  $C$ ’s are also much more fine-tuned semantically, including the first general classifier in Chinese, i.e. 枚 *mei* (Huang 1961). With the preceding pre-Qin period as the embryonic stage, the Han period is thus often considered as the transitional stage, through which Chinese is to become a full-fledged classifier language in the following era of Wei Jin Nanbeichao (220–589 AD) (Jiang 2006).<sup>20</sup>

In his influential study of  $C/M$ ’s during Wei Jin Nanbeichao, Liu (1965) finds that  $[\text{Num } N]$  and  $[N \text{ Num}]$  are by now scarce and thus claims that the use of  $C/M$  is mandatory between  $\text{Num}$  and  $N$ . Along with a drastic increase of shape-specific and function-specific  $C$ ’s, the dominant word order is now  $[\text{Num } C/M \text{ } N]$ , replacing the previously dominant  $[N \text{ Num } C/M]$ . It is uncontroversial that in the eight hundred years from the end of Han to the end of Tang, a classifier system is fully developed and established in Middle Chinese, though some believe it is as early as Wei Jin Nanbeichao and others think it is as late as Tang. In Yang-Drocourt’s (2004) study, it is found that a mature system close to the modern system in Chinese is in place near the end of Song (13th c. AD).

The  $[\text{Num } C/M \text{ } N]$  order continues to stabilize and dominate, eventually at the expense of the demise of  $[N \text{ Num } C/M]$  in Pre-modern Chinese during Yuan and Ming. In Modern Chinese, the unmarked order remains  $[\text{Num } C/M \text{ } N]$ , while  $[\text{Num } N]$  and  $[N \text{ Num } C/M]$  are highly marked and used only if stylistically

20. Wei Jin Nanbeichao, the period of China’s disunity after Han and before Sui, is also known as *Liuchao*, or Six Dynasties, which includes *Sanguo*, or Three Kingdoms, Jin Dynasty, and Nanbeichao, or Southern and Northern Dynasties.

required and only in very specific contexts. Huang (1964) summarizes the historical development of the various word orders as follows.

(41) Developmental Stages of C/M Word Orders (Huang 1964)

Developmental Stage	I	II	III	IV	V
Word Order	[N Num]	[Num N]	[N Num N]	[N Num C/M]	[Num C/M N]
Examples	馬五 <i>ma wu</i> horse 5	五馬 <i>wu ma</i> 5 horse	馬五馬 <i>ma wu ma</i> horse 5 horse	馬五匹 <i>ma wu pi</i> horse 5 C	五匹馬 <i>wu pi ma</i> 5 C horse

A concise summary of the word variation among Num, C/M, and N is provided in the seminal work of Peyraube (1998), validated by many subsequent works, e.g. Yang-Drocourt (2004), Pan (2005), Behr (2009), Wu et al. (2006), and Jiang (2006).

(42) C/M's Seven Word Orders in Chinese History (Peyraube 1998)<sup>21</sup>

- a. Num + N (一牛 *yi niu* 1 ox)
- b. N + Num (虎一 *hu yi* tiger 1)
- c.  $N_1$  + Num +  $N_2$  (羌十人 *qiang shi ren* Qiang 10 person)
- d. N + Num + M (貝廿朋 *bei nian peng* cowrie 20 M-double-strands)
- e. N + Num + C (馬三匹 *ma san pi* horse 3 C)
- f. Num + M + N (一杯羹 *yi bei geng* 1 M-cup soup)
- g. Num + C + N (一株松 *yi zhu song* 1C pine-tree)

Note that I have already pointed out that (42c), [ $N_1$  Num  $N_2$ ], like the echo construction [ $N_1$  Num  $N_1$ ], can be viewed as either a case of [Num N] or [N Num C], depending on whether the second N is in fact N or C. With that, (42) can be reduced to six word orders, which can be further reduced with C and M united. In essence, Num and C/M must always be adjacent, like other languages (Greenberg 1990[1972]:185, Aikhenvald 2000:104–105).

(43) C/M's Word Orders in Chinese History (Generalized)

- a. Num + N
- b. N + Num
- c. N + Num + C/M
- d. Num + C/M + N

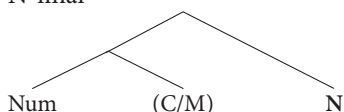
Note that this four-way variation in history can be fully, and indeed quite elegantly, accounted for by the head parameter under the simple assumption that Num

21. The reader is referred to Yang-Drocourt (2004), an excellent study largely based on corpus data that documents the developmental stages the Chinese language underwent from a language without C to its current state with an intricate system of largely obligatory C's in numeral constructions with the predominant [Num C N] order.

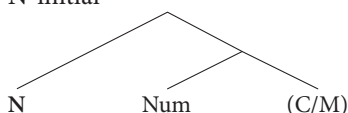
and C/M form a constituent, which then merges with N, in either a head-initial or head-final fashion, as indicated in (44). This is the simplest, and in fact the only, way to ensure that Num and C/M are always adjacent, without resorting to movement.

- (44) Word Order Variation under [Num C/M] Constituency and Head-parameter

a. N-final



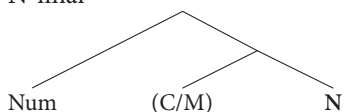
b. N-initial



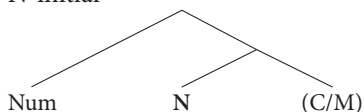
I have already discussed in §3 the parallel constituency and fixed word order between [Num C/M] in NPs and [*n base*] in numerals. The historical development of C/M word order variation is thus simply due to the rise of C/M coupled with a shift from N being head-initial to being head-final. Since adjacency is a necessary, though not sufficient, condition for constituency, the structure in (44) at least fulfills this necessary condition. However, this necessary condition cannot be met under a right-branching analysis, where the same application of the head parameter, regardless of N or C/M as the head, yields drastically different, and incorrect, word order variations, as shown in (45).

- (45) Word Order Variation under [C/M N] Constituency and Head-parameter

a. N-final



b. N-initial



The structure in (45b) allows Num and C/M to be separated by N and therefore fails to capture the parallelism between [Num C/M] and [*n base*] in numerals. Furthermore, the two possibilities of (45b), [Num N C] and [Num N M], never occurred in the history of Chinese. Yet, two of the orders that did occur, [N Num C] and [N Num M], are not accounted for by (45a). The only way for (45a) to work is

to view Num as the Spec of C/M and to resort to the raising of N. Movement, however, is seen as costly under the Minimalist Program. Therefore, historical facts of C/M word orders also favor the [Num C/M] constituency, or the left-branching structure.

## 6. Cross-linguistic implications

Based on the findings above, I shall now venture to propose that the [Num C/M] constituency, where the ordering may vary depending on whether it is C/M-initial or C/M-final, is not only true for Chinese but also true for all classifier languages. In other words, N never intervenes between Num and C/M in a constituent composed of Num, C/M and N, and C/M always forms a constituent with Num first. The argumentation is based on cross-linguistic data related to C/M word order and the internal structure of numerals.

### 6.1 Correlation between the C/M-parameter and the base-parameter

In Dryer (2013), a study that covers 1,154 languages, more than half of the languages, 608 to be exact, have the noun following numerals, e.g. English *eight fish*. With N as the head in [Num N], such languages can be said to be **N-final**. 479 languages have the noun preceding numerals, e.g. in Pumi, a Tibeto-Burman language in Northwestern Yunnan, China, the same English expression would be *fish eight*. Such languages can be said to be **N-initial**. No dominant order is found in 65 languages, and in 2, numerals can only modify verbs. The data indicate that languages can be largely divided as head-initial or head-final, concerning a head noun and a numeral quantifier.

Given that [Num C/M], e.g. [*san da*] ‘three dozen’, regardless of the internal word order, form a constituent and function exactly like a numeral, it is no surprise that the N-final order [[Num C/M] N] and the N-initial [N [Num C/M]] order are both found in classifier languages. In turn, as discussed in §4, internally numerals can be either **base-final** or **base-initial**, corresponding to the **C/M-final** or **C/M-initial** order. Thus, accordingly, [[Num C/M] N] and [N [Num C/M]] can be expanded to four variations. And these are precisely the four word orders that Greenberg’s (1990[1972]:185) has found in his study of the world’s classifier languages, which has been confirmed in Aikhenvald’s (2000:104–105) influential book.

(46) Four Types of Word Order under [Num C/M] Constituency

	<i>N-final</i>	<i>N-initial</i>
<i>C/M-final</i>	(A) [[Num C/M] N] e.g. Chinese, Vietnamese, Hmong, Miao, Uzbek	(B) [N [Num C/M]] e.g. Thai, Burmese, Japanese, Khmer, Mal (Austro-Asiatic)
<i>C/M-initial</i>	(C) [[C/M Num] N] e.g. Ibibio (Niger-Congo), Kiriwina (a.k.a. Kilivila, Oceanic)	(D) [N [C/M Num]] e.g. Louisiade Archipelago (Oceanic), Jingpho, Bodo (Tibeto-Burman)

I have already demonstrated that the C/M-final orders, i.e. Types A and B, consistently correspond to a base-final order in numerals in the same respective language. Type A languages include Sinitic and Miao-Yao (a.k.a. Hmong-Mien) languages, and Type B languages include Old Chinese, Japanese, Thai, as well as some other Tai-Kadai, or Zhuang-Dong, languages. Given the fact that the majority of human languages employ a base-final numeral system, the fact that C/M-final orders are much more than C/M-initial orders is also consistent with the nexus base directionality and C/M directionality this paper tries to establish.

Thus, logically, the C/M-initial orders of Type C and D should likewise correspond to a base-initial order in numerals. I have demonstrated that it is true in some of the Tibeto-Burman languages with a C/M-initial Type D order, [N C/M Num]. Amdo Tibetan and Jingpho, for example, have the numerals 1,000 and above constructed in a base-initial fashion, i.e. [*thousand*×*n*], even though lower numerals have a base-final internal order, e.g. [*n* × *ten*]. In Chan’s (2016) survey, many of the Atlantic languages of the Niger-Congo group have largely base-initial numeral systems. Interesting enough, Greenberg (1990[1972]) has identified Ibibio, a Niger-Congo language of southern Nigeria, as one with the Type C order. Since Types C and D are both rather rare, Ibibio has become a standard citation in the literature for its [C/M Num N] order. Yet, even more interestingly, the language has an exclusively base-initial order in numerals that involve multiplication. Here I give a sample of its numerals.<sup>22</sup>

(47) Ibibio numerals

ked	1
iba	2
ita	3
inaang	4

22. See *Numbering In Ibibio* at: <https://pencillors.wordpress.com/2016/06/25/numbering-in-ibibio/> (accessed 8 August, 2015).

<b>ition</b>	<b>5</b>
itioked	6 (5+1)
itiaba	7 (5+2)
itia-ita	8 (5+3)
usuk-ked	9 (??1) <sup>23</sup>
<b>duop</b>	<b>10</b>
duopo-ked	11 (10+1)
duop-eba	12 (10+2)
duop-eta	13 (10+3)
duop-enaang	14 (10+4)
<b>efid</b>	<b>15</b>
efid-aked	16 (15+1)
efid-eba	17 (15+2)
efid-eta	18 (15+3)
efid-enaang	19 (15+4)
<b>edip</b>	<b>20</b>
edip mme duop	30 (20+10)
<b>aba</b>	<b>40</b>
aba mme duop	50 (40+10)
<b>ata</b>	<b>60</b>
ata mme duop	70 (60+10)
<b>anaang</b>	<b>80</b>
anaang mme duop	90 (80+10)
<b>ikie</b>	<b>100</b>
ikie iba	200 (100×2)
ikie ita	300 (100×3)
<b>tosin</b>	<b>1000</b>
tosin iba	2000 (1000×2)
tosin ita	3000 (1000×3)

While Chinese is a model example of the perfect match between C/M-final and base-final orders, Ibibio has very pleasantly turned out to be an example of the perfect match between its C/M-initial and base-initial orders. Another language with Type C order is Kilivila, the Austronesian language of the Trobriand Islanders. The examples in (48) are quoted from Senft (2000:18–21) and illustrate the C/M-initial

23. The composition of this number is difficult to make out, as the source does not contain any information on what *usuk* refers to.



order, [C/M Num N]. The numerals in Kilivila are thus base-initial accordingly (Senft 1986:77–80), as shown in (49).

- (48) a. *na-tala yena*  
 CL.animal-one fish  
 ‘one fish’  
 b. *kevala-lima yena*  
 CL.batch.drying-five fish  
 ‘five batches of smoked fish’  
 c. *oyla-lima yena*  
 CL.string-five fish  
 ‘five strings with stringed-on fish’  
 d. *pwasa-lima pwasa-tala yena*  
 CL.rotten-five CL.rotten-one fish  
 ‘six rotten fish’

(49) Kilivila numerals

-tala	1
-yu	2
-tolu	3
-vasi	4
<b>-lima</b>	<b>5</b>
-lima-tala	6 (5+1)
-lima-yu	7 (5+2)
-lima-tolu	8 (5+3)
-lima-vasi	9 (5+4)
<b>-luwatala</b>	<b>10 (10×1)</b>
-luwayu	20 (10×2)
-luwatolu	30 (10×3)
<b>-lakatutala</b>	<b>100 (100×1)</b>
-lakatuyu	200 (100×2)
-lakatutolu	200 (100×3)
-lakatuvasi	200 (100×4)

The next example I shall give is Lhiimaqalhqama, a Totonacan language family of Mesoamerica. Kung (2006) has found that the language has a consistent [C/M Num N] order. Even though some thirty classifiers have been elicited, only eight appeared in the data analyzed and in fact the majority of speakers use only the general classifier *laqa-* and the human classifier *puma-*. The classifiers are prefixed to the numerals and quantifiers, as shown in (50).

- (50) *maqa-tam*      *luu*  
 CL:flexible-one snake  
 ‘one snake’

In terms of numerals, one through five are native to the language but those higher than five are all borrowed from Spanish, whose numerals are base-final. However, I do not think we should jump to the conclusion that Lhiimaqalhqama’ is therefore a counterexample to the claim that a C/M-initial order must correlate with a base-initial order in numerals or that it is a language where C/M developed prior to a multiplication-based numeral system. I suspect that the language had a native base-initial numeral system but later lost it due to the dominant Spanish influence. In §6.2, I shall provide some circumstantial evidence for this scenario in several other languages.

Clearly, not all languages with a multiplication-based numeral system have C/M’s. However, the parallel between C/M and base makes two predictions. First, a language with a C/M system should have a multiplication-based numeral system. This prediction is beyond the scope of the paper and remains to be confirmed with further research. The second prediction is that in a classifier language the C/M-parameter should correlate with its base-parameter. The evidence provided so far has been anecdotal and the generalization is based on surveys presented in other works, such as Comrie (2011), Jiang (2006), Yang (2005), and Chan (2016), remain impressionistic. A more systematic survey with much more concrete statistics will now be given.

## 6.2 A survey based on Greenberg’s (1990[1972]) 103 classifier languages

Greenberg (1990[1972]:167, fn.5) lists 103 classifier languages, not including Mandarin, but no specific information on C/M-ordering or base-ordering is given. Out of these 103 languages, there are 52 for which I was able to find sources with reliable data on their numeral systems as well as the C/M word orders.<sup>24</sup> A list of these 52 languages can be found in the Appendix. Map 1 illustrates the geographical distribution of the 52 languages in a GIS map.<sup>25</sup> A summary of the base and C/M correlation is given in Table 1, where the language families that these 52 languages belong to are also indicated.

24. The information on the base-parameter in the numeral systems is mostly obtained from observing the data available from Chan (2016), while that of the C/M-parameter from various papers or books on the respective languages.

25. The information on the precise coordinates of these languages is obtained from SIL.



Map data © 2016 Google, SK telecom, ZENRIN  
Language data from WLMS 16 [www.gmi.org/wlms](http://www.gmi.org/wlms)

Map 1. Distribution of 52 Classifier Languages

Table 1. Summary of Base & C/M Correlation in 52 Classifier Languages

Lang Group	Number of Classifier Languages	Base- & C/M-final	Base- & C/M-initial	Mixed Orders
Afro-Asiatic	1			1
Algic	1	1		
Altaic	2	2		
Austro-Asiatic	8	8		
Austronesian	7	7		
Chibchan	1		1	
Dravidian	1	1		
Indo-European	6	6		
Japonic	1	1		
Language Isolates	3	3		
Mayan	3	3		
Sino-Tibetan	9	6	1	2
Tai-Kadai	6	6		
Totonac	1			1
Uralic	1	1		
Uto-Aztecan	1	1		
Total	52	46	2	4

Out of the 52 languages, 46 have base-final and C/M-final orders, 2 are base-initial and C/M-initial. Among the 4 that have mixed orders, Jingpho and Bodo are Tibeto-Burman, Hausa is Afro-Asiatic, and finally there is a Totonacan language in Mexico. Note, crucially, that these 4 languages with mixed orders are exceptional in exactly the same way: C/M-initial mixed with base-final. This fact is significant, as will become evident momentarily.

**Jingpho**, a.k.a. Kachin, is a C/M-initial language. Note that though its numerals are base-final in general, 1,000 and above being base-initial, i.e. [*thousand*  $\times$  *n*]. Given Jingpho's close historical contact with, and thus strong influence by Chinese, an unyielding C/M-final and base-final language, this 'split' in its numeral system indicates a conflict between its **indigenous base-initial** system and the borrowed Chinese system. The use of Cs is optional and is often absent with numerals over ten (Dai & Xu 1992:128). Garo, another Tibeto-Burman language surveyed, for example, is consistently base-initial and C/M-initial, thus thoroughly maintaining the indigenous systems.

The pattern found in **Bodo**, also a C/M-initial language, spoken in the north-eastern Indian state of Assam, is similar to Jingpho in that Bodo numerals are generally base-final; however, the numeral 100 can be either base-final or base-initial. Thus, compared to Jingpho, the **indigenous base-initial** system of Bodo has deteriorated further, most likely due to the influence of Assamese, the dominant language of the region, which, much like Chinese, is consistently base-final and C/M-final.

**Hausa**, according to Newman (2000) and Zimmermann (2008), has only one numeral classifier, which comes before the numeral and is thus C/M-initial. Its numerals again have two orders. Numerals above 100 are **indigenous and base-initial**, which correlates with the classifier; yet, numerals such as 20, 30, 40, etc. are borrowed from Arabic and base-final. Most Hausa-speakers regardless of ethnic-affiliation are Muslims. Thus, again, the split in base-ordering is due to the concurrent decline of the native base-initial system and the rise of the base-final system borrowed from a dominant language.

**Totonac** is likewise mixed with base-final and C/M-initial orders. Its numerals are consistently base-final, whether they are indigenous or borrowed from Spanish. Crucially, however, as noted by McFarland (2009:105), numeral classifiers are obligatorily prefixed on numbers up to twenty but not on any higher numbers. Now, the fact that its indigenous system is vigesimal means that classifiers, in the form of C-Num, do not co-occur with numerals with a conflicting order, i.e. [*n base*]. In other words, though this base-final C/M-initial conflict exists in the language, it never manifests itself in a noun phrase.

In short, even though the number of base-initial, C/M-initial languages is small, this survey of 52 classifier languages suggests that there is a strong universal

tendency for the word order of C/M and that of the base to be correlated.<sup>26</sup> This universal tendency finds a convincing underlying motivation in the common multiplicand function between C/M and base. It also supports the [Num C/M] constituency in all classifier languages.

### 6.3 Established C/M word orders and [Num C/M] constituency

Returning to word order variations, there are however six, not four, mathematical possibilities among Num, C/M, and N, the other two besides the four in (46) being [C/M N Num] and [Num N C/M], which I shall call **Type E** and **Type F** respectively, as shown in (51).

(51) Six Types of Word Order among Num, C/M, and N

- |                 |                              |
|-----------------|------------------------------|
| (A) [Num C/M N] | (e.g. <i>three C horse</i> ) |
| (B) [N Num C/M] | (e.g. <i>horse three C</i> ) |
| (C) [C/M Num N] | (e.g. <i>C three horse</i> ) |
| (D) [N C/M Num] | (e.g. <i>horse C three</i> ) |
| (E)*[C/M N Num] | (e.g. <i>C horse three</i> ) |
| (F)*[Num N C/M] | (e.g. <i>three horse C</i> ) |

Greenberg (1990[1975]) claims that the last two types do not exist. Indeed, as mentioned earlier, in our research Type F indeed has not turned up in the literature. Though Type E, [C/M N Num], has been identified in some of the Tai-Kadai languages, it is never the dominant order for C/M and without exception it is only allowed when Num is a specific form of the numeral *one*. I have thus demonstrated in §4 that the *one* here behaves more like an indefinite article or an adjective and thus should not be seen as Num. I have argued further that there is in fact a covert *one* and thus the order should be viewed as [Num C/M N]. Typologically, classifier languages either allow (e.g. Chinese) or require (e.g. Maonan) Num to be omitted when it is exactly *one*. The reason for this is again mathematical. The underlying operation in [Num C/M] is multiplication, much like that in a numeral [*n base*]; both can be seen as [multiplier × multiplicand] (Au Yeung 2005, 2007; Her 2012a, 2012b). When Num, the multiplier, is *one*, it is redundant mathematically and can thus be omitted and the result will be the same.

26. A much more extensive survey of numeral classifier languages is under way to confirm this universal tendency. Meanwhile, it should be expected that the number of base-initial, C/M-initial languages is relatively much smaller than the base-final, C/M-final languages, for the simple reason that in the hot spot of numeral classifier languages, i.e. East Asia and South Asia, cultures that employ the base-final numeral systems are by far more dominant than those that use the base-initial ones.

Thus, the [Num C/M] constituency makes a very simple story, i.e. as long as [Num C/M] remains a constituent uninterrupted by N, all word orders are attested in languages;<sup>27</sup> the two unattested word order types, i.e. (51E) and (51F) both violate the [Num C/M] constituency. The story would be drastically different under an analysis where [C/M N] forms a constituent first and then merge with Num. This analysis predicts that C/M and N are never separated and thus gives rise to the following typology.

(52) Four Types of Word Order under [C/M N] Constituency

	<i>Num-final</i>	<i>Num-initial</i>
<i>C/M-final</i>	(D) [[N C/M] Num] e.g. Louisiade Archipelago (Oceanic), Bodo (Tibeto-Burman)	(F) *[Num [N C/M]] (does not exist)
<i>C/M-initial</i>	(E) *[[C/M N] Num] (does not exist)	(A) [Num [C/M N]] e.g. Chinese, Vietnamese, Bengali

Types E and F, which do not exist, are predicted to be viable options in languages, and yet two of the orders actually found in languages, i.e. Type B, [N Num C/M], and Type C, [C/M Num N], are rendered impossible.

Finally, a split analysis, e.g. Zhang (2011) and Li (2011), where both [Num /M] and [C/M N] are employed, predicts that all six possible variations are viable options in languages. The lack of Type E and F in human languages indicates that a split analysis is quite too powerful and simply incorrect. I therefore conclude with confidence that C/M and N do not form a constituent first and then merge with Num;<sup>28</sup> rather, the [Num C/M]/[C/M Num] constituency is universal, and the internal word order between Num and C/M in a language is consistent with the internal word order of [*n base*]/[*base n*] in numerals in that language.

27. The claim that Adams (1989:12) makes is thus correct, that the classifier and number are never separated from each other. Her further conjecture is thus also correct, that in the [C/M N *one*] construction in some Tai languages, 'one' is not a number.

28. Aikhenvald (2000:111–112) claims that Kana, a Cross River language in southeastern Nigeria, is the only exception, citing Ikoro's (1994:19–21) tone sandhi evidence showing C/M forming a prosodic unit with the head noun, not the numeral. However, this may not be the correct *syntactic* analysis upon close examination, as C can in fact be separated from N by an adjectival phrase, as in (i) below.

(i) *zii ka kpaa-bee nee*  
one CL:GENERIC bald-head person  
'one bald person'

(Ikoro 1994:19–21, cited in Aikhenvald (2000:112))

## 7. Conclusion

This paper is concerned with the fundamental issue whether a classifier (C) or measure word (M) first forms a constituent with the numeral (Num), i.e. [Num C/M], or the noun (N), thus [C/M N], in a phrase composed of Num, C/M, and N. The language examined most closely is Mandarin Chinese from both synchronic and diachronic perspectives; however, the issue is also carefully considered from cross-linguistic typological perspectives. The conclusion that C/M first merges with Num, based on such argumentation, thus applies not only to Mandarin Chinese but also has significant implication on all classifier languages.

I first summarized the advantages of the unified left-branching [[Num C/M] N] analysis for Modern Mandarin Chinese and then offered two kinds of new evidence, from historical and typological data. I demonstrated that in Chinese the [Num C/M] constituency and its internal word order are entirely consistent with those of the round figures, i.e. [*n base*], with multiplication as the common underlying operation. Thus, [Num C/M] and [*n base*] can be unified as [multiplier  $\times$  multiplicand]. Moreover, this parallel internal structure between C/M and numerals and the consistent multiplicand-final order, manifested as the corresponding C/M-final and base-final word orders, has existed in the language throughout its 3,000 years of history.

I also examined some of the typologically different Tibeto-Burman languages within the Sino-Tibetan family, where both varieties, i.e. [C/M-final and base-final] and [C/M-initial and base-initial], are found. Instances of [C/M N Num], where N intervenes between C/M and N when Num is of the value 1, are rejected as genuine counterexamples to the generalizations between C/M and numerals, as here the so-called numeral *one* is in fact not a numeral and is instead an indefinite article, like the English *a/an*, or an adjective meaning *single*. Furthermore, a systematic survey of 52 genetically diverse classifier languages also confirms the implicational universal that base-parameter and C/M-parameter are correlated. The above insights gained from the typology of C/M ordering and base ordering in numerals are better accounted for within the unified left-branching approach where [Num C/M] form a constituent that quantifies N.

Historically, the order [Num + C/M] has always remained the same in Chinese. The unified left-branching analysis for Modern Chinese can quite straightforwardly and comprehensively account for all historical word order variations by the simple head-parameter. However, a right branching analysis, i.e. [Num [C/M N]], together with the head-parameter, predicts incorrectly that Num and C/M can be separated by N in Chinese and also fails to capture the consistent parallelism between C/M and numeral bases throughout history.

Among the six word orders logically available among Num, C/M, and N, only four are attested. Essentially, all orders are allowed as long as N does not come in between Num and C/M. This typology is again easily accounted under the left-branching [Num C/M] constituency and the head-parameter. Yet, under the right-branching [C/M N] constituency and likewise the head-parameter, both under-generation and over-generation obtain. This again indicates that the left-branching approach is on the right track.

## Acknowledgements

I am sincerely grateful to the editors of *Language and Linguistics* and its two anonymous reviewers for their careful and insightful comments, which led to significant improvements of the paper. However, all remaining errors are my own. I also gratefully acknowledge that the research reported in the paper was largely funded by the following MOST (previously NSC) grants: 101-2410-H-004-184-MY3, 103-2633-H-004-001, 104-2410-H-004-164-MY3, and 104-2633-H-004-001. I also offer my heartfelt thanks to the many graduate students who have worked on these projects.

## References

- Adams, Karen Lee. 1989. *Systems of Numeral Classification in the Mon-Khmer, Nicobarese and Aslian Subfamilies of Austroasiatic (Pacific Linguistics B101)*. Canberra: ANU Press.
- Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices (Oxford Studies in Typology and Linguistic Theory)*. Oxford: Oxford University Press
- Au Yeung, Ben Wai Hoo. 2005. *An Interface Program for Parameterization of Classifiers in Chinese*. Kowloon: Hong Kong University of Science and Technology dissertation. doi: 10.14711/thesis-b922442
- Au Yeung, Ben Wai Hoo. 2007. Multiplication basis of emergence of classifiers. *Language and Linguistics* 8.4:835–861.
- Behr, Wolfgang. 2009. Classifiers, lexical tone, determinatives: a look at their emergence and diachrony in Old Chinese and beyond. Paper presented at the Diachrony of CLF Workshop, March 11, 2009. Wassenaar: Netherlands Institute of Advanced Studies.
- Borer, Hagit. 2005. *Structuring Sense*, Volume 1: In Name Only. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199263929.001.0001
- Chan, Eugene S. L. 2016 (ongoing). *Numeral Systems of the World's Languages*. Available online at <http://lingweb.eva.mpg.de/moved.php> (accessed 11 August 2011).
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Cheng, Lisa L.-S., & Rint Sybesma. 1998. Yi-wan tang, yi-ge tang: classified and massifiers. *Tsing Hua Journal of Chinese Studies*, New Series 28.3:385–412.
- Cheng, Lisa L.-S., & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30.4:509–542. doi: 10.1162/002438999554192



- Chu, Fuzhang. 1994. Keji Hanyu zhong de shuming jieyou [The Num-N construction in Chinese for technology]. *Hanyu Xuexi [Chinese Language Learning]* 3:40–41.
- Comrie, Bernard. 2006. Numbers, language, and culture. Paper presented at the 16<sup>th</sup> Jyväskylä Summer School, July 24–August 11, 2006. Jyväskylä: University of Jyväskylä.
- Comrie, Bernard. 2011. Numeral bases. *The World Atlas of Language Structures Online*, ed. by Martin Haspelmath, Matthew S. Dryer, David Gill & Bernard Comrie. Munich: Max Planck Digital Library. Available online at <http://wals.info/feature/131> (accessed 3 August 2011).
- Croft, William. 1994. Semantic universals in classifier systems. *Word* 45.2:145–171. doi: 10.1080/00437956.1994.11435922
- Dai, Qing-Xia, & Xi-Jian Xu. 1992. *The Grammar of Kachin*. Beijing: ZhongYang Minzu Xueyuan Chubanshe.
- Dempsey, James Martin. 1995. *A Reconsideration of Some Phonological Issues Involved in Reconstructing Sino-Tibetan Numerals*. Seattle: University of Washington dissertation.
- Dryer, Matthew S. 2013. Order of numeral and noun. *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer & Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/feature/89> (accessed 3 October 2011).
- Duanmu, San. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.
- Duanmu, San. 2005. The tone-syntax interface in Chinese: some recent controversies. *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena: Historical Development, Tone-Syntax Interface, and Descriptive Studies*, December 16–17, 2004, ed. by Shigeki Kaji, 231–264. Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.
- Duojie Dongzhi. 2005. Zangyu Anduo fangyan zhong de liangci [The classifiers in Amdo Tibetan]. *Hanzangyuxi Liangci Yanjiu [Studies of Classifiers in Sino-Tibetan]*, ed. by Jingfang Li & Suhua Hu, 47–54. Beijing: Central University for Nationalities Press.
- Greenberg, Joseph H. 1990[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 & Suzanne Kemmer, 166–193. Stanford: Stanford University Press. First published 1972 in *Working Papers on Language Universals* 9:2–39.
- Greenberg, Joseph H. 1990[1974]. Studies in numerical system: double numeral system. *On Language. Selected Writings of Joseph H. Greenberg*, ed. by Keith Denning & Suzanne Kemmer, 194–206. Stanford: Stanford University Press. First published 1974 in *Working Papers on Language Universals* 14:75–89.
- Greenberg, Joseph H. 1990[1975]. Dynamic aspects of word order in numeral classifier. *On Language. Selected Writings of Joseph H. Greenberg*, ed. by Keith Denning & Suzanne Kemmer, 227–240. Stanford: Stanford University Press. First published 1975 in *Word Order and Word Order Change*, ed. by Charles N. Li, 27–46. Austin: University of Texas Press.
- Greenberg, Joseph H. 1990[1978]. Generalizations about numeral systems. *On Language. Selected Writings of Joseph H. Greenberg*, ed. by Keith Denning & Suzanne Kemmer, 271–309. Originally published in *Universals of Human Language: Word Structure*, ed. by Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik, Vol 3, 249–295. Stanford: Stanford University Press.
- Gu, Yang, & Da Wu. 2005. Cong Jingpoyu han Yiyu de liangci duanyu kan mingci duanyu de zhishe tezheng [Implications of Jingpho and Yi classifier phrase on noun phrase referentiality]. *Hanzangyuxi Liangci Yanjiu [Studies of Classifiers in Sino-Tibetan]*, ed. by Jingfang Li & Suhua Hu, 55–89. Beijing: Central University for Nationalities Press.

- He, Chuansheng. 2015. Complex numerals in Mandarin Chinese are constituents. *Lingua* 164:189–214. doi: 10.1016/j.lingua.2015.06.014
- Her, One-Soon. 2012a. Distinguishing classifiers and measure words: a mathematical perspective and implications. *Lingua* 122.14:1668–1691. doi: 10.1016/j.lingua.2012.08.012
- Her, One-Soon. 2012b. Structure of classifiers and measure words: a lexical functional account. *Language and Linguistics* 13.6:1211–1511.
- Her, One-Soon, & Chen-Tien Hsieh. 2010. On the semantic distinction between classifiers and measure words in Chinese. *Language and Linguistics* 11.3:527–551.
- Hsieh, Miao-Ling. 2008. *The Internal Structure of Noun Phrases in Chinese (Taiwan Journal of Linguistics: Book Series in Chinese Linguistics)*. Taipei: Crane.
- Hu, Houxuan (ed.) 1983. *Jiaguwen Heji [The Great Collection of the Oracle Inscriptions]*. Beijing: China Social Sciences Press.
- Hua, Qiang. 2011. *Jiaguwen Bijiao Yanjiu [A Comparative Study of Oracle Bones]*. Xian: Sanqin Shushe.
- Huang, C.-T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Cambridge: MIT dissertation.
- Huang, C.-T. James, Y.-H. Audrey Li, & Yafei Li. 2009. *The Syntax of Chinese*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166935
- Huang, C.-T. James, & Masao Ochi. 2011. Classifiers and nominal structure: a parametric approach and its consequences. *Universals and Variation: Proceedings of Glow-in-Asia VIII 2010*, ed. by Ming-le Gao, 191–197. Beijing: Beijing Language and Culture University Press.
- Huang, Chu-Ren. 1989. *Mandarin Chinese NP de: A Comparative Study of Current Grammatical Theories*. Taipei: Institute of History and Philology, Academia Sinica.
- Huang, Shengzhang. 1961. Lianghan shidai de liangci [Classifiers in the two-Han period]. *Zhongguo Yuwen [Studies of the Chinese Language]* 1961.8:21–28.
- Huang, Zaijun. 1964. Cong jiawen jinwen liangci de yingyong kaocha Hanyu liangci de qiyuan yu fazhan [A study of the origin and development of classifiers in Chinese from the perspectives of the use of classifiers in oracle bone and bronze inscriptions]. *Zhongguo Yuwen [Studies of the Chinese Language]* 6:432–441.
- Ikoro, Suanu M. 1994. Numeral classifiers in Kana. *Journal of African Languages and Linguistics* 15:7–28. doi: 10.1515/jall.1994.15.1.7
- Ionin, Tania, & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23:315–360. doi: 10.1093/jos/ffl006
- Jenks, Peter. 2010. Evidence for the syntactic diversity of numeral classifiers. LSA Annual Meeting Extended Abstracts 2010. Washington, DC: Linguistic Society of America.
- Jenks, Peter. 2011. *The Hidden Structure of Thai Noun Phrases*. Cambridge: Harvard University dissertation.
- Jiang, Julie Li. 2012. *Nominal Arguments and Language Variation*. Cambridge: Harvard University dissertation.
- Jiang, Renping. 2007. *Jishuci he Xushuci de Leixingxue Yanjiu [A Typological Study of Cardinals and Ordinals]*. Nanchang: Nanchang University MA thesis.
- Jiang, Ying. 2006. *Hanzangyuxi Mingliangci Yanjiu [A Study of Nominal Classifiers in Sino-Tibetan]*. Beijing: Central University for Nationalities dissertation.
- Kubo, Miori. 1996. Some considerations on noun classes and numeral classifiers: a study of (pseudo-) partitives in Japanese and English. *Keio Studies in Theoretical Linguistics I*, ed. by Yuji Nishiyama & Yukio Otsu, 89–124. Tokyo: Institute of Cultural and Linguistic Studies, Keio University.

- Kung, Susan Smythe. 2006. Numeral classifiers in Lhiimaqalhqama'. *Texas Linguistic Society IX: The Morphosyntax of Underrepresented Languages*, ed. by Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu, Jessica White & Stephen Wechsler, 185–202. Stanford: CSLI Publications. Available online at <http://cslipublications.stanford.edu/TLS/> (accessed 18 September 2011).
- LaPolla, Randy, J. 2003. Overview of Sino-Tibetan morphosyntax. *The Sino-Tibetan Languages*, ed. by Graham Thurgood & Randy J. LaPolla. London & New York: Routledge.
- Li, Charles N., & Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Li, Wendan. 2000. The pragmatic function of numeral-classifiers in Mandarin Chinese. *Journal of Pragmatics* 32.8:1113–1133. doi: 10.1016/S0378-2166(99)00086-7
- Li, XuPing. 2009. Semantics of pre-classifier adjectives. *Proceedings of ConSOLE XVI*, ed. by Sylvia Blaho, Camelia Constantinescu & Bert Le Bruyn, 117–134. Leiden: Universiteit Leiden. Available online at <http://media.leidenuniv.nl/legacy/console16-li.pdf> (accessed 18 September 2011).
- Li, XuPing. 2011. *On the Semantics of Classifiers in Chinese*. Ramat Gan: Bar-Ilan University dissertation.
- Li, Yen-Hui Audrey. 1998. Argument determiner phrases and number phrases. *Linguistic Inquiry* 29.4:693–702. doi: 10.1162/ling.1998.29.4.693
- Li, Yen-Hui Audrey. 2014. Structure of noun phrases: left or right? *Taiwan Journal of Linguistics* 12.2:1–32.
- Li, Yuming. 2000. Kaobeixing minglingci zai Hanzangyuxi lianci fazhan zhong de diwei [The status of echo classifiers in the development of classifiers in Sino-Tibetan]. *Zhongguo Yuwen [Studies of the Chinese Language]* 2000.1:27–34.
- Lin, Jo-wang. 1997. Noun phrase structure Mandarin Chinese: DP or NP? *Chinese Languages and Linguistics* 3:401–434.
- Liu, Hongyong. 2006. *The Structure of Complex Nominals: Classifiers, Possessives and Relatives*. Shatin: The Chinese University of Hong Kong dissertation.
- Liu, Shiru. 1965. *We-jin Nanbeichao Liangci Yanjiu [A Study of Classifiers in Wejin Nanbeichao]*. Beijing: Zhonghua Book Company.
- Löbel, Elisabeth. 2001. Classifiers and semi-lexicity: functional and semantic selection. *Semi-Lexical Categories: the Function of Content Words and the Content of Function Words*, ed. by Norbert Corver & Henk van Riemsdijk, 223–272. Berlin & New York: Mouton de Gruyter. doi: 10.1515/9783110874006.223
- Ma, Aimin. 2011. Hanyu jiaoxue zhong ying zhuyi mingliangci de fenlei wenti [Nominal classifiers should be properly classified for pedagogical purposes]. *Xiandai Yuwen [Modern Chinese]* 3:37–39.
- Matisoff, James A. 1997. *Sino-Tibetan Numerals and the Play of Prefixes (Pacific Linguistics B-114)*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- McFarland, Teresa Ann. 2009. *The Phonology and Morphology of Filomeno Mata Totonac*. Berkeley: University of California at Berkeley dissertation.
- Newman, Paul. 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven: Yale University Press.
- Pan, Yukun. 2005. *Xi-Zhou Jinwen Yuxu Yanjiu [Studies in the Syntax of Western Zhou Bronze Inscriptions]*. Shanghai: East China Normal University Press.
- Paris, Marie-Claude. 1981. *Problèmes de Syntaxe et de Sémantique en Linguistique Chinoise*. Paris: Collège de France, Institut des Hautes Etudes Chinoises.

- Peyraube, Alain. 1998. On the history of classifiers in Archaic and Medieval Chinese. *Studia Linguistica Serica*, ed. by Benjamin K. T'sou, 131–145. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.
- Pittayaporn, Pittayawat. 2009. *The Phonology of Proto-Tai*. Ithaca: Cornell University dissertation.
- Senft, Gunter. 1986. *Kilivila: The Language of the Trobriand Islanders*. Berlin: de Gruyter Mouton. doi: 10.1515/9783110861846
- Senft, Gunter. 2000. What do we really know about nominal classification systems? *Systems of Nominal Classification*, ed. by Gunter Senft, 11–49. Cambridge: Cambridge University Press.
- Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. *The Oxford Handbook of Comparative Syntax*, ed. by Guglielmo Cinque & Richard Kayne, 806–838. Oxford: Oxford University Press.
- Sun, Chaofen. 1996. *Word-Order Change and Grammaticalization in the History of Chinese*. Stanford: Stanford University Press.
- Tai, James H-Y., & Lianqing Wang. 1990. A semantic study of the classifier Tiao. *Journal of the Chinese Language Teachers Association* 25.1:35–56.
- Tang, Chih-Chen Jane. 1990a. A note on the DP analysis of Chinese noun phrases. *Linguistics* 28.2: 337–354. doi: 10.1515/ling.1990.28.2.337
- Tang, Chih-Chen Jane. 1990b. *Chinese Phrase Structure and the Extended X'-theory*. Ithaca: Cornell University dissertation.
- Tang, Chih-Chen Jane. 2005. Nouns or classifiers: a non-movement analysis of classifiers in Chinese. *Language and Linguistics* 6.3:431–472.
- Tsai, Hui-Chin. 2011. On Chinese ordinal constructions. *Taiwan Journal of Linguistics* 9.1: 89–122.
- van Driem, George. 2003. Tibeto-Burman vs. Sino-Tibetan. *Language in Time and Space: A Festschrift for Werner Winter on the Occasion of His 80th Birthday*, ed. by Brigitte L. M. Bauer & Georges-Jean Pinault, 101–120. Berlin & New York: Mouton de Gruyter. doi: 10.1515/9783110897722.101
- van Driem, George. 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Zhongguo Yuyanxue Jikan [Bulletin of Chinese Linguistics]* 1.2:211–270.
- van Driem, George. 2011. Tibeto-Burman subgroups and historical grammar. *Himalayan Linguistics* 10.1:31–39.
- van Riemsdijk, Henk. 1998. Categorical feature magnetism: the endocentricity and distribution of projections. *Journal of Comparative Germanic Linguistics* 2.1:1–48. doi: 10.1023/A:1009763305416
- Vos, Riet. 1999. *A Grammar of Partitive Constructions*. Tilburg: Tilburg University dissertation.
- Wang, Hanwei. 2004. Liangci de fenlei yu duiwai Hanyu liangci jiaoxue [Classification of classifiers and teaching classifiers in Chinese as a second language]. *Jinan Xuebao [Journal of Jinan University (Humanities and Social Sciences Edition)]* 2:113–116.
- Watanabe, Akira. 2006. Functional projections of nominals in Japanese: syntax of classifiers. *Natural Language and Linguistic Theory* 24.1:241–306. doi: 10.1007/s11049-005-3042-4
- Wu, Fuxiang, Shengli Feng, & C.-T. James Huang. 2006. Hanyu shu+lianhg+ming geshi de laiyan [On the origin of the construction of 'numeral+classifier+noun' in Chinese]. *Zhongguo Yuwen [Studies of the Chinese Language]* 2006.5:387–400.
- Yang, Jiangling. 2005. Zhangmianyu shuliang duanyu cong CN dao NC xing de yanbian jizhi [The mechanism of change in classifier phrases from CN to NC within Tibeto-Burman].

Hanzangyuxi Liangci Yanjiu [Studies of Classifiers in Sino-Tibetan], ed. by Jingfang Li & Suhua Hu, 29–46. Beijing: Central University for Nationalities Press.

Yang-Drocourt, Zhitang. 2004. *Evolution Syntaxique du Classificateur en Chinois, du XIIIe Siècle av. J.-C. au XVIIe Siècle*. Paris: École des Hautes Études en Sciences Sociales.

Yi, Byeong-Uk. 2009. Chinese classifiers and count nouns. *Journal of Cognitive Science* 10:209–225. doi: 10.17791/jcs.2009.10.2.209

Yi, Byeong-Uk. 2011. What is a numeral classifier? *Philosophical Analysis* 23:195–258.

Zhang, Jingni. 2005. Maonanyu de liangci duanyu [The classifier phrase in Maonan]. *Hanzangyuxi Liangci Yanjiu [Studies of Classifiers in Sino-Tibetan]*, ed. by Jingfang Li & Suhua Hu, 296–308. Beijing: Central University for Nationalities Press.

Zhang, Niina Ning. 2009. Syntactic properties of numeral classifiers in Mandarin Chinese. Talk given on April 10, 2009, at the Graduate Institute of Linguistics. Taipei: National Chung Cheng University.

Zhang, Niina Ning. 2011. The constituency of classifier constructions in Mandarin Chinese. *Taiwan Journal of Linguistics* 9.1:1–50.

Zhang, Niina Ning. 2012. De and the functional expansion of classifiers. *Language and Linguistics* 13.3:569–582.

Zhang, Niina Ning. 2013. *Classifier Structures in Mandarin Chinese*. Berlin: de Gruyter Mouton. doi: 10.1515/9783110304992

Zhang, Yachu (ed.) 2001. *Yin-Zhou Jinwen Jicheng Yinde [Index to Collection of Inscriptions of the Yin-Zhou Period]*. Beijing: Zhonghua Book Company.

Zimmermann, Malte. 2008. Quantification in Hausa. *Quantification: A Cross-Linguistic Perspective*, ed. by Lisa Matthewson, 415–475. Bingley: Emerald.

Appendix. Base & C/M Synchronization in 52 languages

Language	Base & C/M Correlation	Language	Base & C/M Correlation
Ahom	final-to-final	Malay	final-to-final
Ainu	final-to-final	Mon	final-to-final
Archaic Chinese	final-to-final	Northern Uzbek	final-to-final
Assamese	final-to-final	Northwestern Ojibwa	final-to-final
Batak Karo	final-to-final	Osetin	final-to-final
Bengali	final-to-final	Pohnpeian	final-to-final
Bulgarian	final-to-final	Rawang	final-to-final
Burmese	final-to-final	Ruching Palaung	final-to-final
Cebuano	final-to-final	Sauria Paharia	final-to-final
Chuukese	final-to-final	Shan	final-to-final
Classical Nahuatl	final-to-final	Tabasco Chontal	final-to-final
Eastern Farsi	final-to-final	Tai Dam	final-to-final

Language	Base & C/M Correlation	Language	Base & C/M Correlation
Eastern Jakalteko	final-to-final	Tai Don	final-to-final
Eastern Katu	final-to-final	Tajiki	final-to-final
Gilyak	final-to-final	Thai	final-to-final
Hakka Chinese	final-to-final	Turkish	final-to-final
Hungarian	final-to-final	Venustiano Carranza Tzotzil	final-to-final
Japanese	final-to-final	Vietnamese	final-to-final
Kharia	final-to-final	Wolio	final-to-final
Khasi	final-to-final	Yue Chinese	final-to-final
Khmu	final-to-final	<i>Bribri</i>	<i>initial-to-initial</i>
Kiribati	final-to-final	<i>Garó</i>	<i>initial-to-initial</i>
Korean	final-to-final	Bodo	mixed
Lao	final-to-final	Hausa	mixed
Lave	final-to-final	Jingpho	mixed
Lisu	final-to-final	<b>Totonac</b>	<b>final-to-initial</b>

### *Author's address*

One-Soon Her  
Graduate Institute of Linguistics and Research Center for Mind, Brain, and Learning  
National Chengchi University  
64, Sec. 2, ZhiNan Road  
Taipei 116  
Taiwan  
hero@nccu.edu.tw

### **Publication history**

Date received: 14 August 2015  
Date accepted: 1 February 2016