

STRUCTURE OF NUMERALS AND CLASSIFIERS IN CHINESE: HISTORICAL AND TYPOLOGICAL PERSPECTIVES AND CROSS-LINGUISTIC IMPLICATIONS*

One-Soon Her

*Graduate Institute of Linguistics & Research Center for Mind, Brain, and Learning
National Chengchi University*

Abstract

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, 3×10 . The same multiplicative function exists between Num and C/M, e.g., *san-duo hua* '3 C flower' = 3×1 flower, and *san-da hua* '3 dozen flower' = 3×12 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

從歷史與型態學的角度看漢語中數詞與量詞的 語法結構以及其跨語言的意涵

何萬順

語言學研究所暨心智大腦與學習研究中心

國立政治大學

〔摘要〕

分類詞 (C) 與量詞 (M) 究竟是先與數詞 (Num) 結合形成短語 [Num C/M], 還是先與名詞 (N) 形成短語 [C/M N], 一直存有爭議。本文首先簡要回顧現代漢語中的證據, 接著從歷史與型態學的角度提出兩種新的證據, 論證 [Num C/M] 先形成短語。第一種證據來自數詞中的乘法結構, 如「三十」乃 3×10 , 同樣的乘法結構存在於 Num 與 C/M 之間, 如「三打」乃 3×12 , 可見 C/M 與位數詞 (base) 的功能相同, 都是被乘數 (multiplicand)。在漢語三千多年的文字歷史裡, 數詞從來都是位數在後 (base-final), 如「三十」、「三百」, 同樣的 Num 與 C/M 之間也從來都是「C/M 在後」(C/M-final)。我們根據 Greenberg (1990[1978]) 找出了 52 個語言, 發現其 C/M 與位數 (base) 的詞序極為一致。這顯示了「三打」與「三十」一樣形成短語。第二種證據來自 C/M 詞序在型態學上的四種變異: 在 [Num C/M] 形成短語的假設下, 不僅是漢語歷史發展中的 C/M 詞序, 而是所有分類詞語言中的四種詞序, 都可以直接的由中心語參數 (head parameter) 得到解釋。

關鍵字: 分類詞、量詞、短語結構、數詞、乘法、中心語參數

* 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.

STRUCTURE OF NUMERALS AND CLASSIFIERS IN CHINESE: HISTORICAL AND TYPOLOGICAL PERSPECTIVES AND CROSS-LINGUISTIC IMPLICATIONS

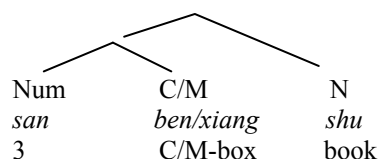
1 Introduction

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.

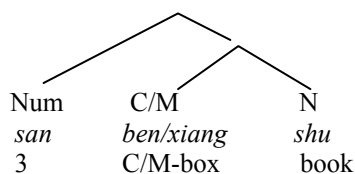
- (1) a. 三 本 書 (Classifier, C)
san ben shu
 3 C book
 ‘3 books’

- 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 will thus avoid being tangled up by 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, 3×12), 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, 3×10). 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 and Wang 1990, Croft 1994, Peyraube 1998, Cheng and 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 in 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* × *multiplicand*], where the multiplicand C is *1*.¹

- (3) a. 三 個 雞蛋
 san ge jidan (3 × 1 egg = 3 eggs; *ge* = 1)
 3 C egg
 '3 eggs'
- b. 三 粒 雞蛋
 san li jidan (3 × 1 egg = 3 eggs; *li* = 1)
 3 C egg
 '3 eggs'
- c. 三 顆 雞蛋
 san ke jidan (3 × 1 egg = 3 eggs; *ke* = 1)
 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 $3 \times 12 \text{ roses} = 36 \text{ roses}$, 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 $3 \times 12 \text{ roses} = 12 \text{ roses} \times 3$, we shall see in section 4 that it does matter linguistically for C/M and numerals.

- (4) a. 三 打 雞蛋
san da jidan (3 × dozen egg ≠ 3 eggs; dozen ≠ 1)
 3 M-dozen egg
 ‘3 dozens of eggs’
- b. 三 磅 雞蛋
san bang jidan (3 × pound egg ≠ 3 eggs; pound ≠ 1)
 3 M-pound egg
 ‘3 pounds of eggs’
- c. 三 箱 雞蛋
san xiang jidan (3 × box egg ≠ 3 eggs; box ≠ 1)
 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 *IM*, *IM* is not *1*. For example, in (4a), [*3 × dozen egg*] (= [*3 × 1dozen 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 the particular semantic feature or features it highlights, be it humanness, animacy, shape, or function, mathematically it is *necessarily 1*.

In short, mathematically C/M converge in being the multiplicand with Num as the multiplier; [Num C/M] thus share 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 and Hsieh (2010) and others.² As shown in (5b), the pre-C adjective scopes over C as well as N.³ In contrast, the pre-M adjective in (5a) does not scope over N.

- (5) a. 一 大 箱 蘋果 ≠ 一 箱 大 蘋果 (H&H 13a)
yi da xiang pingguo yi xiang da pingguo
 1 big M-box apple 1 M-box big apple
 ‘1 big box of apples’ ‘1 box of big apples’
- b. 一 大 顆 蘋果 = 一 顆 大 蘋果 (H&H 13b)⁴
yi da ke pingguo yi ke da pingguo
 1 big C apple 1 C 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 and Wang 1990, Cheng and Sybesma 1998, 1999). This is quite simply incorrect, as convincingly demonstrated in many recent works, e.g., Hsieh 2008, Her and Hsieh 2010, Zhang 2011, and Li 2011. Note also that Cheng and 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 (2008:118) 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.’

‘1 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 scope 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/M N]: 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-c). Note that English does not allow numerals to be conjoined with *and*, thus **one and two books*, conjunction of numerals with *or* is fine, thus *one or two books*. This is true for Chinese 和 *han/he* ‘and’, but 又 *you* ‘and’, as in (8d), and 或 *huo* ‘or’ are fine, as in (8b-c).

(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-b), 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 you san xiang e
 he have 3 M-box book I 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-b). 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/*三/十三
*ta you yibai ben shu, wo you sanbai/*san/??shi-san*
 he have 100 C book I have 300/3/13
 你 有 幾 百/*幾?
*ni you ji-bai/*ji?*
 you have 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/*三/十三
*ta you yibai xiang shu, wo you sanbai/*san/??shi-san*
 he have 100 M-box book I have 300/3/13
 你 有 幾 百/*幾?⁵
*ni you ji-bai/*ji?*
 you have 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* [3×1], 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

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.

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).

- (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’
- b. 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 and Wang 1990, Cheng and Sybesma 1999). This is incorrect, as convincingly demonstrated in many recent works, e.g., Hsieh (2008), Her and 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 and 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].⁶

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’
- 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
liang ge ban zhongtou (Hsieh 2008:46 (53a))
 2 C half hour
 ‘2 and a half hours’
- b. 一 個 多 鐘頭
yi ge duo zhongtou (Hsieh 2008:46 (54a))
 1 C more hour
 ‘more than 1 hour’

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 (2013) 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 (2013) 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 complimentizer *de* (e.g., Hsieh 2008, Her 2012b).

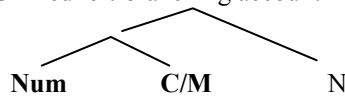
- (14) a. 兩 瓶 半 可樂
liang ping ban kele (Hsieh 2008:46 (53b))
 2 M-bottle half coke
 ‘2 and a half bottles of coke’
- b. 一 瓶 多 可樂
yi ping duo kele (Hsieh 2008:46 (54b))
 1 M-bottle more coke
 ‘more than 1 bottle of coke’

To summarize, C/M converge and diverge at the same time, as observed by many researchers before (e.g., Chao 1968, Wang and Tai 1999, Cheng and 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.

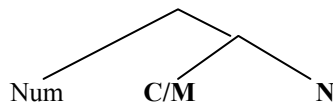
3 Controversy over C/M Constituency

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).

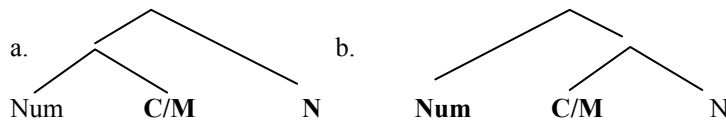
(15) Unified left-branching account



(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 and 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 and 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 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 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, receive a structural solution in Zhang's (2011, 2013) split account. In (18a), pre-M adjectives and numerals c-command M only and thus do not scope over N. In (18b), pre-C adjectives and numerals c-command C as well as N and thus scope over both C and N. 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 the Lexical-Functional Grammar. In short, scope phenomena do not favor either the unified approach or 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., coordination 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.⁷ Under the right-branching account, if Num is assumed to be in the Spec of CIP (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 CIP 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, section 3.4 for more discussions).

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,

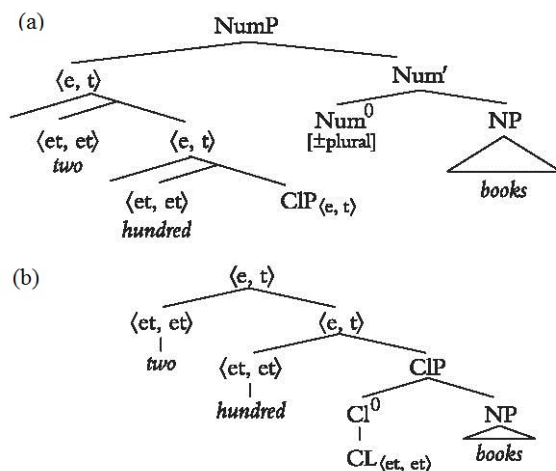
⁷ An anonymous reviewer suggests that the fact that [Num C/M] cannot be dislocated may be evidence against its constituency, as shown in (i). However, while dislocation may be evidence for constituency, its absence is not evidence for non-constituency, because other known modifier phrases also do not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

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.⁸ The point to be made here is that the left-branching (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).

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 × C/M] + *ban/duo*], one that resembles that of complex numerals, e.g., [[2 × 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 and 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. Section 4 offers evidence from the historical data on the internal structure of numerals. Section 5 then does the same based on historical data on word order variations.

⁸ As convincingly demonstrated in Her (2012a) with ample corpus data, the plural marker *-men* do co-occur with numerals, though optionally. The example below is taken from a web site of National Chiayi University in Taiwan. The URL is: http://www.nccyu.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.'

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 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.⁹ In this section, we shall see that, consistently throughout the 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.¹⁰ 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 ... $xn + y$, 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's 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 $[\text{base } n]$

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's call this the **C/M-parameter**.¹¹

(21) C/M-parameter:

- a. C/M-final, thus $[\text{Num } \underline{\text{C/M}}]$
- b. C/M-initial, thus $[\underline{\text{C/M}} \text{ Num}]$

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

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.

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

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 $[n \text{ base}]$ and $[\text{Num C/M}]$ both function as $[\text{multiplier} \times \text{multiplicand}]$, C/M and base, as multiplicands, should of course follow the same order.

The crucial fact is therefore that, though 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 $[\text{multiplier} > \text{multiplicand}]$. 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 form a constituent. In a language that encodes the multiplicative $[n \text{ base}]$, e.g., *san-shi* (three-ten) $[3 \times 10]$, as a syntactic constituent, it is only logical that it encodes the same multiplicative $[\text{Num C/M}]$, e.g., *san-ge* (three-C) $[3 \times 1]$ and *san-da* (three-M_{-dozen}) $[3 \times 12]$, 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 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 also true *diachronically*.¹² As a prelude to the discussions in section 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)¹³

- 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. BCE to late 11th c. BCE, 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 $[\text{eight} \times \text{hundred}]$, or $[n \times \text{base}]$. (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*
 '13081 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 will 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] and is thus redundant. There are essentially just six word orders. We shall discuss this in more detail in section 4.

4.2 Numerals and C/M in Tibeto-Burman¹⁴

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 (2013) 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.¹⁵ According to Duojie (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 (2013), 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. $\eta\partial$ (nda) $hs\partial m$
 person C 3
 '3 people'
- b. ηi $w\text{.}jja$
 2 hundred
 '200'
- c. $rto\eta$ ηi
 thousand 2
 '2000'

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 will 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 (2013) 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*¹⁶. While in Chinese and Hmong-Mien (a.k.a. Miao-Yao) languages the order is always [Num C/M N], the variation is far greater in Tai-Kadai, or Zhuang-Dong, languages, 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).

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 section 6.

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.

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 2013). Similar facts are found in Bouyei, a Northern Group Tai language in the same family, spoken in Guilin, China.¹⁷

(26) a. $ai^1 \ zən^1 \ (dɛu^{231})$ [C/M N Num], C/M-initial
 C person 1
 ‘1 person’

b. $*dɛu^{231} \ ai^1 \ zən^1$
 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. $pɛk^{55}/tʰjɛn^{42} \ (dɛu^{231})$ [*base n*], base-initial
 hundred/thousand one
 ‘100/1000’

b. $*dɛu^{231} \ pɛk^{55}/tʰjɛn^{42}$
 one hundred/thousand
 ‘(intended) 100/1000’

(29) a. $ja^{42}/sa : m^{42} \ pɛk^{55}/tʰjɛn^{42}$ [*n base*], base-final
 2/3 hundred/thousand
 ‘200/300; 2000/3000’

b. $*pɛk^{55}/tʰjɛn^{42} \ ja^{42}/sa : m^{42}$
 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 (2013), and Jiang (2007). First, the numeral *one* in Maonan has three different forms, $tɔ^{231}$, jit^{55} , and $dɛu^{231}$, and only $dɛu^{231}$ 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 $dɛu^{231}$ ‘1’, and likewise the [*base n*] order is allowed only when the base is *hundred* or *thousand* or larger and the multiplier *n* is exactly $dɛu^{231}$ ‘1’. Third, in both cases $dɛu^{231}$ is optional and can be omitted. Fourth, adjectives in Maonan are post-nominal, as in (30), just like $dɛu^{231}$.

(30) a. $ai^1 \ zən^1 \ voj^1$
 C person tall
 ‘the tall person’

b. $tɔ^2 \ nɔk^8 \ da : i^2$
 C bird good

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.

‘the good bird’

Based on these facts, I concur with Jiang (2007) that $d\epsilon u^{231}$ in [C/M N $d\epsilon u^{231}$] is not Num at all. Jiang (2007) further proposes that $d\epsilon u^{231}$ is an adjective. However, I contend that $d\epsilon u^{231}$ 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 $d\epsilon u^{231}$, as in (31). This behavior of $d\epsilon u^{231}$ 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 $d\epsilon u^{231}$, as in (32). The fact that $d\epsilon u^{231}$ is mutually exclusive with known determiners in the same position also indicates that it is of the category D.

(31) a. ai^l $z\partial n^l$ voj^l $d\epsilon u^{231}$
 C person tall Indef. Sg.
 ‘a tall person’

b. $*ai^l$ $z\partial n^l$ $d\epsilon u^{231}$ voj^l
 C person Indef. Sg. tall

(32) a. ai^l $z\partial n^l$ voj^l ka^2
 C person tall Def. Sg.
 ‘that tall person’

b. $*ai^l$ $z\partial n^l$ $d\epsilon u^{231}$ ka^2
 C person Indef. Sg. Def. Sg.

c. $*ai^l$ $z\partial n^l$ ka^2 $d\epsilon u^{231}$
 C person Def. Sg. Indef. Sg..

I thus do not consider examples like Maonan $d\epsilon u^{231}$ genuine counterexamples to the claim that [Num C/M] form an immediate constituent. Viewing $d\epsilon u^{231}$ 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 $d\epsilon u^{231}$, i.e., $t\sigma^{231}$ and jit^{55} , also do not behave like other numerals and appear before C/M.¹⁸

(33) a. ja^{42}/sa : m^{42} ai^l $z\partial n^l$
 2/3 C person
 ‘2/3 persons’

b. $*t\sigma^{231}/*jit^{55}$ ai^l $z\partial n^l$
 one/one C person
 ‘1 person’

(34) a. ai^l $z\partial n^l$
 C person
 ‘the person’

b. $t\sigma^2$ $n\partial k^8$
 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.

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^{55} is a borrowing from Middle Chinese $*\eta j i t$ or Old Chinese $*\eta j i t$ ‘one’; $t\sigma^{231}$ is, on the other hand, native in Kam-Sui; and $d\epsilon u^{231}$, 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 $*d\epsilon u w$ A (> Siamese $d\epsilon u w$ A1, Longzhou $d\epsilon i w$ B1, etc.) by Pittayaporn (2009: 358).

- (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.

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 will 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, (500BCE-200CE, Shang to Han), *Zhonggu Hanyu*, or Middle Chinese, (201–1000, 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 will 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.¹⁹

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 dozen rudimentary C’s in the [N Num C] construction, as in (37), there 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’

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.

- (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-brandy 2 M-sheng-pint 1
‘his 2.1 *sheng*-pint of newly steamed millet-brandy’
- 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 [N₁ Num N₁] 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 [N₁ Num N₂], where the second N is less constrained and need not be identical to N₁; rather, N₂ names an instance of N₁ 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 N₂ in the [N₁ Num N₂] 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 [N₁ Num N₁] and [N₁ Num N₂] are just cases of [Num N]. Or we can view them as C’s, and thus they are cases of [N Num C].

Both constructions, [N₁ Num N₁] and [N₁ Num N₂], however, all but disappeared during the Han Dynasties (206BCE-220CE), 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 (220-589CE) (Jiang 2006).²⁰

In his influential study of C/M’s during Wei Jin Nanbeichao, Liu (1965) finds that [Num N] and [N Num] are by now scarce and thus claims that the use of C/M is mandatory between Num and N. Along with a drastic increase of shape-specific and function-specific C’s, the dominant word order is now [Num C/M N], replacing the previously dominant [N Num C/M]. It is uncontroversial that in the eight

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.

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 other 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. CE).

The [Num C/M N] order continues to stabilize and dominate, eventually at the expense of the demise of [N Num C/M] in Pre-modern Chinese during Yuan and Ming. In Modern Chinese, the unmarked order remains [Num C/M N], while [Num N] and [N Num C/M] are highly marked and used only if stylistically 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)²¹

- a. Num + N (一牛 *yi niu* 1 ox)
- b. N + Num (虎一 *hu yi* tiger 1)
- c. N₁ + Num + N₂ (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), [N₁ Num N₂], like the echo construction [N₁ Num N₁], 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).

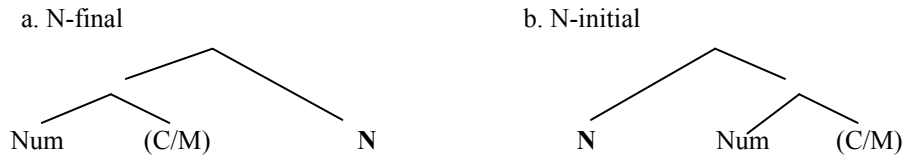
(43) C/M's Word Orders in Chinese History (Generalized)

- 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 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.

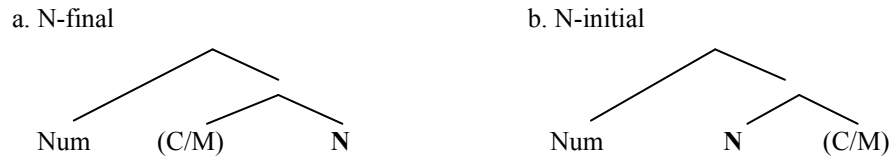
²¹ The reader is referred to Yang-Drocourt (2004), an excellent study largely based on corpus data that documents the developmental stages the Chinese language underwent from a language without C to its current state with an intricate system of largely obligatory C's in numeral constructions with the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

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



I have already discussed in section 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



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 these findings above, I will 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 (2011), a study that covers 1154 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 section 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., Type 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 1000 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 (2013) 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 Type 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, taken from the *Ibibio Dictionary Online*.²²

(47) Ibibio numerals

ked	1
iba	2
ita	3
inaang	4
ition	5
itioked	6 (5+1)
itiaba	7 (5+2)
itia-ita	8 (5+3)
usuk-ked	9 (??1) ²³
duop	10
duopo-ked	11 (10+1)
duop-eba	12 (10+2)
duop-eta	13 (10+3)
duop-enaang	14 (10+4)
efid	15
efid-aked	16 (15+1)
efid-eba	17 (15+2)
efid-eta	18 (15+3)
efid-enaang	19 (15+4)
edip	20
edip mme duop	30 (20+10)

²² *Ibibio Dictionary Online*, accessed on 23 September, 2011, is available at: www.myibibio.com.

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

aba	40
aba mme duop	50 (40+10)
ata	60
ata mme duop	70 (60+10)
anaang	80
anaang mme duop	90 (80+10)
ikie	100
ikie iba	200 (100 ×2)
ikie ita	300 (100 ×3)
tosin	1000
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 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 |
|
 | |
| -lima | 5 |
| -lima-tala | 6 (5+1) |
| -lima-yu | 7 (5+2) |
| -lima-tolu | 8 (5+3) |
| -lima-vasi | 9 (5+4) |
|
 | |
| -luwatala | 10 (10×1) |
| -luwayu | 20 (10 ×2) |
| -luwatolu | 30 (10 ×3) |
|
 | |
| -lakatutala | 100 (100×1) |
| -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'

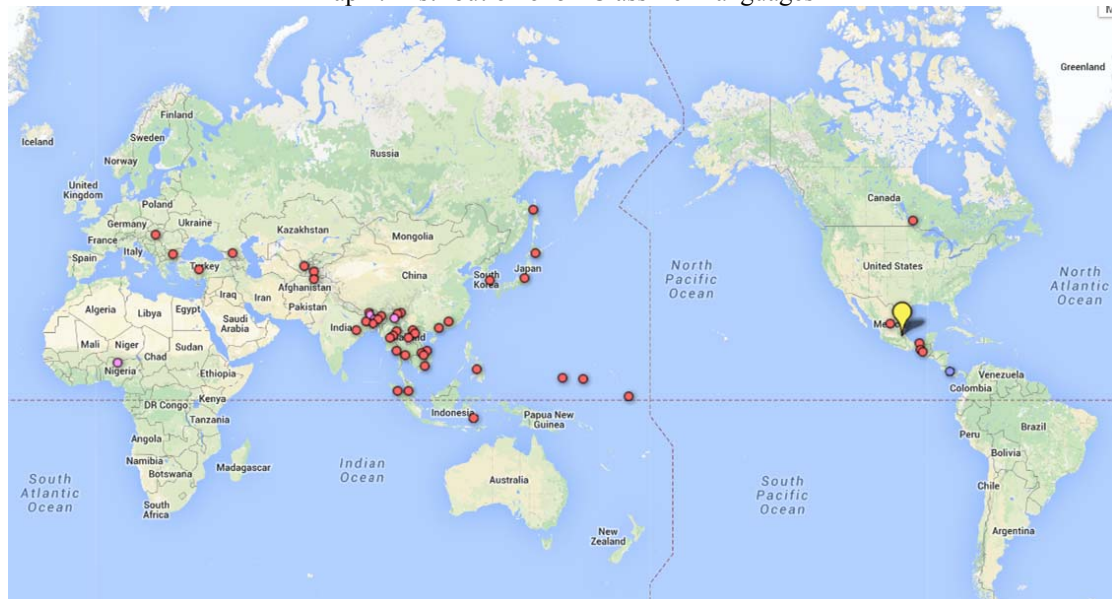
In terms of numerals, one through five are native to the language but those higher than five are all borrowed from Spanish, whose numerals are base-final. However, I do not think we should jump to the conclusion that Lhiimaqalhqama' is therefore a counterexample to the claim that a C/M-initial order 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), Jiang (2006), and Chan (2013) 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.²⁴ 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.²⁵ A summary of the base and C/M correlation is given in Table 1, where the language families these 52 languages belong to are also indicated.

Map 1. Distribution of 52 Classifier Languages



²⁴ The information on the base-parameter in the numeral systems is mostly obtained from observing the data available from Chan (2013), 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.

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 it 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, 1000 and above being base-initial, i.e., [*thousand* × *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 northeastern 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 of 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 *vegesimal* 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.²⁶ This universal tendency finds a convincing underlying motivation in

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.

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., *three C horse*)
- (B) [N Num C/M] (e.g., *horse three C*)
- (C) [C/M Num N] (e.g., *C three horse*)
- (D) [N C/M Num] (e.g., *horse C three*)
- (E) *[C/M N Num] (e.g., *C horse three*)
- (F) *[Num N C/M] (e.g., *three horse C*)

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 section 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.

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 languages²⁷; 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] form 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

Type 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;²⁸ rather, the [Num C/M]/[C/M Num] constituency is universal, and the internal word order

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 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 kpa-a-bee nee (Ikoro 1994: 19-21, cited in Aikhenvald (2000:112))
one CL:GENERIC bald-head person
'one bald person'

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.

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 × 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 3000 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.

References

- Adams, Karen Lee. 1989. *Systems of Numeral Classification in the Mon-Khmer, Nicobarese and Aslian Subfamilies of Austroasiatic*. ANU: Canberra (= Pacific Linguistics Series B - No. 101).
- Aikhenvald, Alexandra Y. 2000. *Classifiers: a typology of noun categorization devices*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press
- Au Yeung, Wai Hoo Ben. 2005. *An interface program for parameterization of classifiers in Chinese*. Ph.D. Dissertation, Hong Kong University.
- Au Yeung, Wai Hoo Ben. 2007. Multiplication basis of emergence of classifier. *Language and Linguistics* 8 (4): 835-861.
- Behr, Wolfgang. 2009. Classifiers, determinatives and lexical tone: a look at their emergence, diachrony and interaction in Old Chinese and beyond. Paper presented at the Diachrony of Classifiers Workshop, March 11, 2009, Netherlands Institute of Advanced Studies, Wassenaar.
- Borer, Hagit. 2005. *Structuring sense*, Vol. 1: *In Name Only*. Oxford: Oxford University Press.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chan, Eugene S. L. 2013. *Numeral systems of the world's languages*. Available online at <http://lingweb.eva.mpg.de/numeral/>. Accessed on 2011/08/11.
- Cheng, Lisa L.-S., and 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., and Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30 (4): 509-542.
- Chu, Fuzhang. 1994. Keji hanyu zhong de shuming jiegou (The Num-N construction in Chinese for technology). *Hanyu Xuexi* (Learning Chinese) 1994(3), 40-41.
- Comrie, Bernard. 2006. Numbers, language, and culture. Paper presented at the Jyväskylä Summer School, Jyväskylä, Finland.
- Comrie, Bernard. 2011. Numeral Bases. In *The world atlas of language structures online*, eds. Dryer, Matthew S. and Haspelmath, Martin. Munich: Max Planck Digital Library, chapter 131. Available online at <http://wals.info/feature/131>. Accessed on 2011/08/03.
- Croft, William. 1994. Semantic universals in classifier systems. *Word* 45: 145-171.
- Dai, Qing-Xia and Xu Xi-Jian. (1992). *The Grammar of Kachin*. Beijing: ZhongYang Minzu Xueyuan Chubanshe.
- Dempsey, James M. 1995. *A reconsideration of some phonological issues involved in reconstructing Sino-Tibetan numerals*. Ph.D. dissertation, University of Washington.
- Duojie, Dongzhi (多杰东智). 2005. 藏语安多方言中的量词 Zangyu Anduo fangyan zhong de liangci 'The classifiers in Amdo Tibetan'. In 漢藏語系量詞研究 *Hanzangyu Liangci Yanjiu* 'Studies of Classifiers in Sino-Tibetan', eds. Li Jingfang (李錦方) and Hu Suhua (胡素華), 47-54. Beijing: The Central University of Nationalities Press.
- Dryer, Matthew S. 2011. Order of Numeral and Noun. In *The world atlas of language structures online*, eds. Dryer, Matthew S. and Haspelmath, Martin. Munich: Max Planck Digital Library, chapter 89. Available online at <http://wals.info/feature/89>. Accessed on 2011/08/03.
- Duanmu, San. 2000. *The phonology of Standard Chinese*. Oxford: Oxford University Press.
- Duanmu, San. 2005. The tone-syntax interface in Chinese: some recent controversies. In *Proceedings of the Symposium "Cross-Linguistic Studies of Tonal Phenomena, Historical Development, Tone-Syntax Interface, and Descriptive Studies"*, December 16-17, 2004, ed. Shigeki Kaji. Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies, pp. 231-264.
- Gil, David. 1994. Summary: Numeral classifiers. *Linguist List* Vol 5: 466
- Greenberg, Joseph H. 1990[1972]. Numerical classifiers and substantival number: problems in the genesis of a linguistic type. In *On language. Selected writings of Joseph H. Greenberg*, eds. K. Denning and Kemmer, 166-193. Stanford, CA: Stanford University Press. [First published 1972 in *Working Papers on Language Universals* 9: 1-39. Stanford, CA: Department of Linguistics, Stanford University.]
- Greenberg, Joseph H. 1990[1974]. Studies in numerical system: Double numeral system. In *On language. Selected writings of Joseph H. Greenberg*, eds. K. Denning and Kemmer, 194-206. Stanford, CA: Stanford University Press. [First published 1974 in *Working Papers on Language Universals* 14: 75-89. Stanford, CA: Department of Linguistics, Stanford University.]
- Greenberg, Joseph H. 1990[1975]. Dynamic aspects of word order in numeral classifier. In *On language. Selected writings of Joseph H. Greenberg*, eds. K. Denning and Kemmer, 227-240. Stanford, CA: Stanford University Press. [First published 1974 in *Word order and word order change*, ed. C. Li, 27-46. Austin, University of Texas Press.]
- Greenberg, Joseph H. 1990 [1978]. Generalizations about numeral systems. In *On Language: selected writings of Joseph H. Greenberg*, ed. by Keith Denning and Suzanne Kemmer, 271—309. [Originally published in *Universals of Human Language*, ed by Joseph H. Greenberg, Charles A. Ferguson, & Edith A. Moravcsik, Vol 3, 249—295. Stanford; Stanford University Press.]
- Gu, Yang and Wu, Da. 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). In *Hanzangyu liangci Yanjiu* 'Studies of Classifiers in Sino-Tibetan', eds. Li Jingfang and Hu Suhua, 55-89. Beijing: The Central University of Nationalities Press.
- He, Chuansheng. 2015. Complex numerals in Mandarin Chinese are constituents. *Lingua* 164:189-214.
- Her, One-Soon. 2012a. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122.14: 1668-1691. (SSCI, AHCI)
- Her, One-Soon. 2012b. Structure of classifiers and measure words: A lexical functional Account. *Language and Linguistics* 13.6:1211-1511.
- Her, One-Soon and Hsieh, Chen-Tien. 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. No. 2. Taipei: Crane Publishing.
- Hu, Houxuan (胡厚宣) (ed.). 1983. 甲骨文合集 *Jiaguwen heji* 'The great collection of the oracle inscriptions', Beijing: China Social Sciences Publishing House.
- Hua, Qiang (華強). 2011. 甲骨文比較研究 *Jiaguwen bijiao yanjiu* 'A comparative study of oracle bones'. Xian: Sanqin Shushe (Sanqin Publishing House).
- Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Ph.D. Dissertation. MIT.
- Huang, C.-T. James, Audrey Li and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge University Press.

- Huang, C.-T. James and Masao Ochi. 2011. Classifiers and nominal structure: a parametric approach and its consequences. In Gao, Mingle (ed.) *Universals and Variation: Proceedings of Glow-in-Asia VIII*, Beijing Languages and Culture University Press, pp. 191-197.
- Huang, Chu-Ren. 1989. *Mandarin Chinese NP de -A Comparative Study of Current Grammatical Theories*. Taipei: Academia Sinica.
- Huang, Shengzhang (黃盛璋). 1961. 兩漢時代的量詞 Lianghan shidai de liangci 'Classifiers in the two-Han period'. *Zhongguo Yuwen* 1961.8: 21-28.
- Huang, Zaijun (黃載君). 1964. 從甲文金文量詞的應用考察漢語量詞的起源與發展 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* 1964.6:432-441.
- Ikoro, Suanu M. 1994. Numeral classifiers in Kana. *Journal of African Languages and Linguistics* 15, 7-28.
- Jenks, Peter. 2010. Evidence for the syntactic diversity of numeral classifiers. Poster in the 84th Annual Meeting of the Linguistics Society of America. Baltimore, MD.
- Jenks, Peter. 2011. *The Hidden Structure of Thai Noun Phrases*. PhD Dissertation, Harvard University.
- Jiang, Julie L. 2012. *Nominal arguments and language variation*. PhD Dissertation, Harvard University.
- Jiang, Renping (蔣仁萍). 2007. 基數詞和序數詞的類型學研究 *Jishuci he xushuci de leixingxue yanjiu* 'A typological study of cardinals and ordinals'. MA thesis, Nanchang University. Available online at: <http://wenku.baidu.com/view/816cd0d4b9f3f90f76c61b68.html>; accessed on June 13, 2012.
- Jiang, Ying (蔣穎). 2006. 漢藏語系名量詞研究 *Hanzang yuxi mingliangci yanjiu* 'A study of nominal classifiers in Sino-Tibetan'. Ph.D. Dissertation, The Central University of Nationalities.
- Joseph, Thattil Umbavu Varghese. 2006. *Rabha*. Leiden: Brill Academic Publishers.
- Kubo, Miori. 1996. Some considerations on noun classes and numeral classifiers: A study of (pseudo-)partitives in Japanese and English. *Studies in Theoretical Linguistics I*, eds. by Y. Nishiyama and Y. Otsu, 89-124. Keio, Japan: Institute of Cultural and Linguistic Studies, Keio University.
- Kung, Susan Smythe. 2006. Numeral classifiers in Lhiimaqalhqama'. In *Topics in the Morphosyntax of Underrepresented Languages: Papers from the 9th Texas Linguistics Society Conference, Austin, Texas, November 4-6, 2005*, volume eds. Douglas S. Bigham, Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu, Jessica White, and series ed. Stephen Wechsler, Stanford, CA: CSLI Publications. Available online at: <http://csli-publications.stanford.edu/TLS/>. Accessed on 2011/09/018.
- LaPolla, Randy, J. 2003. Overview of Sino-Tibetan morphosyntax. *The Sino-Tibetan Languages*, eds. by Graham Thurgood and Randy J. LaPolla. London and New York: Routledge.
- Li, Yen-Hui Audrey. 1998. Argument determiner phrases and number phrases. *Linguistic Inquiry* 29 (4), 693-702.
- Li, Yen-Hui Audrey. 2014. Structure of noun phrases: Left or right? *Taiwan Journal of Linguistics* 12.2:1-32.
- Li, Charles and Sandra 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.
- Li, Xu-Ping. 2008. Semantics of pre-classifier adjectives. *Proceedings of ConSOLE XVI*, 117-134.
- Li, Xu-Ping. 2011. *On the semantics of classifiers in Chinese*. Ph.D. Dissertation, Bar-Ilan University.
- Li, Yuming (李宇明). 2000. 拷貝型名量詞在漢藏語系量詞發展中的地位. Kaobeixing lingci zai hanzang yuxi lianci fazhan zhong de diwei 'The status of echo classifiers in the development of classifiers in Sino-Tibetan'. *Zhongguo Yuwen* 2000.1: 27-34.
- Lin, Jo-wang. 1997. Noun phrase structure in Mandarin Chinese: DP or NP? *Chinese Languages and Linguistics* 3: 401-434.
- Liu, Hongyong. 2006. The Structure of Complex Nominals: Classifiers, Possessives and Relatives. PhD thesis, the Chinese University of Hong Kong.
- Liu, Shiru (劉世儒). 1965. 魏晉南北朝量詞研究 *Wejin Nanbeichao liangci yanjiu* 'A Study of Classifiers in Wejin Nanbeichao'. Beijing: Zhonghua Book Company.
- Löbel, Elisabeth. 2001. Classifiers and the notion of 'Semi-Lexical Head'. *Semi-Lexical Categories*, eds. by van Riemsdijk, H. v. and N. Corver. Berlin/New York: Mouton de Gruyter.
- McFarland, Teresa Ann. 2009. *The Phonology and Morphology of Filomeno Mata Totonac*. PhD Dissertation. University of California, Berkeley. P.107
- 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) 2011.3.
- 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.
- Newman, Paul. 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven: Yale University Press.
- Pan Yukun (潘玉坤). 2005. *Xi-Zhou jinwen yuxu yanjiu* 西周金文语序研究 'Studies in the syntax of Western Zhou bronze inscriptions', Shanghai: Huadong 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. In *Studia Linguistica Serica*, ed. 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*. Ph.D. Dissertation, Cornell University.
- Robertson, Teresa. 2008. Essential vs. accidental properties. In *The Stanford Encyclopedia of Philosophy*, ed. Edward Zalta. Accessed online, December 1, 2009. Available online at: <http://plato.stanford.edu/entries/essential-accidental>
- Senft, Gunter. 1986. *Kilivila: The language of the Trobriand Islanders*. Berlin: Mouton de Gruyter.
- Senft, Gunter. 2000. What do we really know about nominal classification systems? In *Systems of nominal classification*, ed. Gunter Senft, 11-49. Cambridge: Cambridge University Press.
- Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. *The Oxford Handbook of Comparative Syntax*, eds. by Guglielmo Cinque and Richard Kayne, 806-838. Oxford University Press.
- Sun, Chaofen. 1996. *Word-Order change and grammaticalization in the history of Chinese*. California: Stanford University Press.
- Tai, H.-Y. James. and L. 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: 337-354.

- Tang, Chih-Chen Jane. 1990b. *Chinese phrase structure and extended X'-theory*. Ph.D dissertation, Cornell University.
- 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. In Werner Winter, Brigitte L. M. Bauer and Georges-Jean Pinault (eds.) *Language in time and space: a Festschrift for Werner Winter on the occasion of his 80th birthday*, Walter de Gruyter, pp. 101-119.
- van Driem, George. 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *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: 1-48.
- Vos, Riet. 1999. *A Grammar of Partitive Constructions*. Ph.D dissertation, Tilburg University.
- 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 2004(2): 113-116.
- Watanabe, Akira. 2006. Functional Projections of Nominals in Japanese: Syntax of Classifiers. *Natural Language and Linguistic Theory* 24: 241-306.
- Wu, Fuxiang (吳福祥), Feng, Shengli (馮勝利), & Huang, Cheng-Teh James (黃正德). 2006. 漢語“數+量+名”格式的來源 Hanyu shu+liang+ming geshi de lai yuan (On the origin of the [Num Cl N] construction). *Zhongguo Yuwen* 2006.05.
- Wu, Yicheng and Adams Bodomo. 2009. Classifiers≠determiners. *Linguistic Inquiry*. 40 (3), 487-503.
- Yang, Jiangling (楊將領). 2005. 藏緬語數量短語從CN到NC的演變機制 Zhangmianyu shuliang duanyu cong CN dao NC de yanbian jizhi (The mechanism of change in classifier phrases from CN to NC within Tibeto-Burman). In 漢藏語系量詞研究 *Hanzangyu Liangci Yanjiu* 'Studies of Classifiers in Sino-Tibetan', eds. Li Jingfang (李錦方) and Hu Suhua (胡素華), 29-46. Beijing: The Central University of Nationalities Press.
- Yang-Drocourt, Zhitang. 2004. *Evolution syntaxique du classificateur en chinois, du XIIIe siècle av. J.-C. au XVIIe siècle* (Hanyu liangci de chansheng ji qi zaiju fashang de yanbian 漢語量詞的產生及其在句法上的演變), Paris: École des hautes études en sciences sociales.
- Yi, Byeong Uk. 2009. Chinese classifiers and count nouns. *Journal of Cognitive Science* 10: 209-225.
- Yi, Byeong Uk. 2011. What is a numeral classifier? *Philosophical Analysis* 23: 195-258.
- Zhang, Jingni (張景霓). 2005. 毛南語的量詞短語 Maonanyu de liangci duanyu 'The classifier phrase in Maonan'. In 漢藏語系量詞研究 *Hanzangyu liangci yanjiu* 'Studies of Classifiers in Sino-Tibetan', eds. Li Jingfang (李錦方) and Hu Suhua (胡素華), 296-308. Beijing: The Central University of 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, National Chung Cheng University. Accessed online, December 1, 2009. http://ling.auf.net/lingbuzz/000854/current.pdf?_s=iNqV2Jyn9dGKP_DS
- 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: Mouton de Gruyter.
- Zhang, Yachu (張亞出) (ed.). 2001. 殷周金文集成引得 *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. In Lisa Matthewson (ed.), *Quantification: a cross-linguistic perspective*. Bingley: Emerald, 415-475.

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
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	Totonac	final-to-initial