

Deriving Classifier Word Order Typology, or Greenberg's Universal 20A, and Universal 20*

One-Soon Her

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

Abstract

The word order typology of numerals (Num), classifier or measure word (C/M), and noun (N) put forth by Greenberg (1990[1972]) can be reduced to a universal principle: N does not come between Num and C/M. Given the affinity between this universal and Greenberg's Universal 20, which concerns the word order typology of D, Num, A, and N, the former is dubbed 'Universal 20A' (Her et al. 2015). This paper first discusses, and ultimately rejects, the two alleged exceptions to Universal 20A, one in Ejagham, the other in some Tai-Kadai and Tibeto-Burman languages. Then, in light of Universal 20A, Cinque's (2005) successful antisymmetric account of Universal 20 and all its exceptions is re-examined and shown to be inadequate for Universal 20A. The analysis I propose adopts Abels and Neeleman's (2012) symmetric derivational account of Universal 20 and, crucially, takes complex numerals into consideration. The final account also integrates a multiplicative theory of C/M (Her 2012a) and is able to explain the base-C/M harmonization, which was first discovered by Greenberg (1990[1978]:292) but has since been overlooked in classifier research, and also offer a functional explanation for Universal 20A.

Key words: classifier, measure word, word order, Universal 20, Universal 20A

1 Introduction

Mandarin Chinese is a textbook example of a language with numeral classifiers, which refer to the unmarked counting or measuring units in a nominal phrase and serve to facilitate the quantification of the noun by a numeral. In general, numeral classifiers come in two types, classifiers (C) and measure words (M), as shown in the examples of (1) and (2), respectively.

- (1) a. *wu ben zazhi*
5 C magazine
'5 magazines'
- b. *shi ke pingguo*
10 C apple
'10 apples'
- c. *babai pi ma*
800 C horse
'800 horses'

* I am sincerely grateful to the editors of this special issue of *Linguistics*, Walter Bisang and Yicheng Wu, and the two anonymous reviewers for their careful and insightful comments, which led to significant improvements of the paper. I also benefited greatly from discussions with Hui-Chin Tsai on the formal issues dealt with in 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 grants by the Ministry of Science and Technology (MOST) of Taiwan: 101-2410-H-004-184-MY3, 103-2633-H-004-001, 104-2410-H-004-164-MY3, and 104-2633-H-004-001.

(2) a. *wu xiang zazhi*
5 M-box magazine
'5 boxes of magazines'

b. *shi gongjin pingguo*
10 M-kilo apple
'10 kilos of apples'

c. *babai qun ma*
800 M-group horse
'800 herds of horses'

Though C/M differ in their semantic properties, as characterized succinctly by Her and Hsieh (2010:544) in the quote below, C/M occupy the same syntactic position in a classifier language like Mandarin and are in complementary distribution.

C refers to an essential property of the noun, which can be restated as the predicate concept in an analytic proposition with the noun as the subject concept; M refers to an accidental property of the noun in terms of quantity, which can be restated as the predicate concept in a synthetic proposition with the noun as the subject concept. Her and Hsieh (2010:544)

Other terms have been used for the C/M distinction, e.g., 'sortal classifier' and 'measural classifier', 'count-classifier' and 'mass-classifier', 'count-noun classifier' and 'mass-noun classifier', 'qualifying classifier' and 'quantifying classifier', 'classifier' and 'massifier', etc. Many also use the term 'classifier' or 'numeral classifier' for both C/M, while others use 'measure word' for both (e.g., Zhang 2007). In this paper, I shall use the term 'classifier' (C) and 'measure word' (M) for the elements between Num and N in (1) and (2), respectively.

Mathematically there can be six orders among Num, C/M, and N, (factorial $3 = 3 \times 2 \times 1 = 6$), shown in (3), but only four are attested, as first discovered by Greenberg (1990[1972]:185) and confirmed by Aikhenvald (2000:104-105).

(3) Six Possible Word Orders of [Num, C/M, N]

- a. ✓ [Num C/M N] (many languages, e.g., Chinese)
- b. ✓ [N Num C/M] (many languages e.g., Thai)
- c. ✓ [C/M Num N] (few languages e.g., Ibibio [Niger-Congo])
- d. ✓ [N C/M Num] (few languages e.g., Jingpho [Tibeto-Burman])
- e. * [C/M N Num] (no languages)
- f. * [Num N C/M] (no languages)

As further noted by Greenberg (1990[1972]:228) and others, there are many more languages with orders (3a) and (3b), where Num > C/M, than those with (3c) and (3d), where C/M > Num. These potential universals regarding C/M's word order are closely related to Greenberg's (1963) well-known Universal 20, which is concerned with the word orders of D, Num, A, and N.

(4) Greenberg's Universal 20

When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite.

A number of genuine exceptions to Greenberg’s Universal 20 have been confirmed in the literature and thoroughly dealt with by Cinque (2005). Clearly, the C/M word order typology in (3) is closely related to Universal 20. To give Greenberg (1990[1972]) the full credit for the discovery of the four attested orders of classifiers in (3), this derived generalization is dubbed ‘Greenberg’s Universal 20A’, stated in (5) (Her et al. 2015; Her *to appear*).

(5) Greenberg’s Universal 20A

Part 1: Of the three elements Num, C/M, and N, any order is possible as long as N does not come between Num and C/M.

Part 2: There are many more languages with Num > C/M orders than languages with C/M > Num orders.

Under Universal 20A, a taxonomy of the four orders can be derived by two binary parameters: C/M-initial/CM-final and N-initial/N-final, as shown in Table 1.

Table 1. Taxonomy of C/M Word Orders

	N-FINAL	N-INITIAL
C/M-FINAL	(A) [Num C/M N]	(B) [N Num C/M]
C/M-INITIAL	(C) [C/M Num N]	(D) [N C/M Num]

Several exceptions have since been claimed concerning Universal 20A. One case is an alleged [C/M N Num] order in Ejagham (Benue-Congo), and the second case is also an alleged [C/M N Num] sequence, where Num must be *one*, in some Tai-Kadai and Tibeto-Burman languages. Inspired by Cinque’s (2005) and Abels and Neeleman’s (2012) successful formal accounts of Universal 20 and its exceptions, this paper aims to likewise account for the two cases of alleged exceptions and derive Universal 20A from a single underlying structure. Ideally, of course, such an account should also derive Universal 20 at the same time. Before offering a formal account, this paper first aims to pursue a functional motivation for Universal 20A.

The paper is organized as follows. Section 2 defends the empirical content of Universal 20A by dismissing the two cases that allegedly contradict Universal 20A. Section 3 then adopts Her’s (2012a) multiplicative theory to justify the unification of C/M as a single syntactic category and to motivate both Part 1 and Part 2 of Universal 20A. Section 4 pursues a formal account of Universal 20A. Complex numerals are introduced, which play a vital role in choosing Abels and Neeleman’s (2012) symmetric solution to Universal 20, over Cinque’s (2005) antisymmetric account, as the basis for extension. Section 5 concludes the paper with a summary and some concluding remarks.

2 Defending Universal 20A

There are two alleged cases that contradict Part 1 of Universal 20A, one from the African language Ejagham, the other from some of the Tai-Kadai and Tibeto-Burman languages. Both cases are alleged instances of the C/M > N > Num order. The other mathematically possible order, Num > N > C/M, has not been reported. Simpson (2005), for example, cites both Ejagham and Nung (Central Tai) to demonstrate that N

can come between Num and C/M. These two cases will be discussed in 2.1 and 2.2 respectively. In 2.3, Part 2 of Universal 20 will be confirmed.

2.1 Ejagham: C/M > N > Num?

The first exception we shall discuss is found in Ejagham, a Benue-Congo Southern Bantoid language, as documented in Watters (1981:310). It was first noted by Aikhenvald (2000:99) that Ejagham's alleged C/M > N > Num order, as shown in (6), is an exception to Greenberg's Universal 20A.

- (6) *à-múg* ' *í-ǰkúid* *á-bá'*
 6-CL:Small.Round GEN6 19-orange.seed 6-two
 'two orange seeds'
 Watters (1981:310)

For Ejagham to be a true exception, one thing must obtain, i.e., the five lexical items that Watters (1981) considers Cs are indeed Cs, not Ns. I contend they are Ns and not Cs. Note that Watters (1981:310) offers no specific argument for the five elements as Cs and in fact concedes that 'some of the classifiers are clearly nominals with a specific meaning, while others are shaped like a nominal with no specific, assignable meaning in isolation.'

Kihm (2005), seemingly unaware of the fact that Ejagham might be a contradiction to Universal 20A, observes specifically that Ejagham Cs behave like nouns, in that they are marked for noun class, e.g., *à-múg* (noun class 6, plural) in (6), like other nouns in the language, e.g., *í-ǰkúid* (noun class 19, singular) in (6). Kihm (2005:498) thus reaches this conclusion: 'It seems we have every reason, therefore, to consider /múg/ a pronoun with the meaning "small round object" morphologically realized as a free form.'

Watters (1981:310) further notes the fact that Num *á-bá'* 'two' (class 6) in (6) concords with the classifier *à-múg* (class 6), not the noun *í-ǰkúid* 'orange seed' (class 19). This fact indicates that this putative classifier is in fact the head noun. In general, Ejagham nouns occur with numbers without any putative C and the order is N > Num, as shown in (7). Thus, should *à-múg* be taken to be an N, (6) is consistent with the general N > Num order in the language.

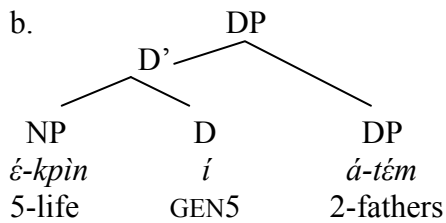
- (7) *bì-yù* *í-bá' é*
 8-yam 8-two
 'two yams'
 (Watters 1981:311)

Note also that, opposite from the word order in English, the head of the Ejagham genitive construction, i.e., the possessed NP, appears before the genitive marker, while the possessor NP appears after it, as shown in (8a), which thus has the structure in (8b), following the standard assumption that the genitive marker is a determiner. The structure of the English counterpart (9a) is thus the same, only with an opposite order, shown in (9b).

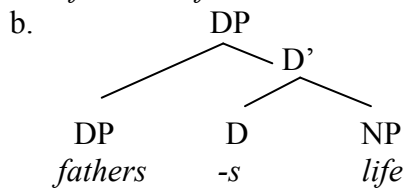
- (8) a. *é-kpìn í* *á-tém*

¹ This example is originally cited in Watters (1981: 310), but here it follows the phonetic spelling, gloss, and notations used in Kihm (2005: 497).

5-life GEN5 2-fathers
 ‘fathers’ life’
 (Watters 1981:354 [61a])

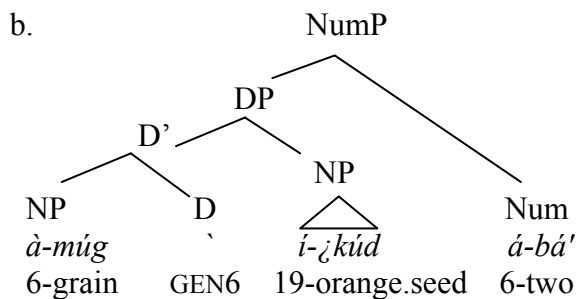


(9) a. *fathers’ life*



The extended structure of (6) is thus headed by Num, again following the standard assumption that Num projects a NumP and takes the DP as its complement, as shown in (10b).

(10) a. à-múg ` í-çkúd á-bá'
 6-grain GEN6 19-orange.seed 6-two
 ‘two grains of orange seeds’



Thus, in terms of noun class, the head of NumP, Num (class 6), naturally concords with the head of its complement DP, D (class 6), as well as the head noun of D’s complement NP (also class 6). I have therefore changed the gloss of the putative C à-múg from Watters’ *CL:Small.Round* to that of a noun *grain*, as the five putative Cs, which may be functionally similar to classifiers, are syntactically Ns. They are very much like English words such as *piece* in *six pieces of paper*, except that Ejagham uses a construct genitive construction, while English uses a free genitive construction. The same conclusion that Ejagham is not a numeral classifier language is independently reached in Doetjes (2012: Section 3.4), where Watters’ (1981) putative Cs are likewise considered nouns. Thus, the alleged [C/M N Num] order in Ejagham turns out to be incorrect and thus irrelevant to Greenberg’s Universal 20A.²

² The fact that if Ejagham were a classifier language, it would violate Greenberg’s Universal 20A, adds to the argument against it being a classifier language.

2.2 Tai-Kadai & Tibeto-Burman: C/M > N > Num?

The second case of apparent exceptions is well-known among linguists that study classifiers in Tai-Kadai and Tibeto-Burman.³ In a study of 194 classifier languages in SMATTI (an acronym for Sinitic, Maio-Yao, Austroasiatic, Tibeto-Burman, Tai-Kadai, and Indo-Aryan), Her et al. (2015) discover that in some languages in Tibeto-Burman and Tai-Kadai, when Num equals *one*, [C/M N Num] obtains.⁴ Specific examples include the Tibeto-Burman language Nuosu Yi (Northern Ngwi) and the Tai-Kadai languages Maonan (Kam-Sui), Mak (Kam-Sui), Bouyei (Northern Tai), Zhuang (Jiang 2007:39), and Nung (Central Tai) (Saul and Wilson 1980). This thus seems to be a serious challenge to Universal 20A.

However, note that in languages with [C/M N Num(=1)], the unmarked order is always [Num C/M N]. In other words, [C/M N Num] and [Num C/M N] must co-exist. The obvious question is, why [C/M N Num] is restricted to *one*? This makes the status of this *one* highly suspicious. Based on the account proposed in Her (to appear), I shall further demonstrate that it is in fact not a numeral and thus not of the category Num; rather, it is an indefinite determiner much like the English *a/an*. I shall further argue that there is a silent *one* in the usual Num position and thus this construction is after all [Num C/M N D], no contradiction to Universal 20A.

In the following discussions, Maonan is used as an example, a Kam-Sui language in the Gunagxi province of China, based on the data and descriptions provided in Zhang (2005), Jiang (2006), Jiang (2007), Chan (2015), and (Her to appear). The unmarked word order of C/M in Maonan is [Num C/M N], as in (11a), and numbers that are not *one* cannot appear after the N, as shown in (11b). The unexpected [C/M N Num] order is only used when the number is *deu*²³¹ ‘one’, as in (12a), and *deu*²³¹ cannot be used before C/M, as shown in (12b) (Zhang 2005).

(11) a. $ja^{42}/sa:m^{42}$ ai^l $zən^l$
 2/3 C person
 ‘2/3 persons’

b. $*ai^l$ $zən^l$ $ja^{42}/sa:m^{42}$
 C person 2/3
 (intended) ‘2/3 persons’

(12) a. ai^l $zən^l$ $(deu^{231}/*tɔ^{231}/*jit^{55})$
 C person 1/1/1
 ‘1 person’

b. $*deu^{231}$ ai^l $zən^l$
 1 C person
 (intended) ‘1 person’

³ Tai-Kadai (a.k.a. Zhuang-Dong) and Hmong-Mien (a.k.a. Miao-Yao), are both independent families in the 16th edition of *Ethnologue* but considered major branches under Sino-Tibetan by most Chinese linguists. The latter classification is followed in this paper.

⁴ In Tai-Kadai, when Num equals *one* both [N C/M Num] and [C/M N Num] are found (Jiang 2006: 25-26). In the Tai group of Tai-Kadai, most of the twenty-plus languages of the Central sub-group and the Northern sub-group have it, according to the Baidu entry of 台语支 *taiyuzhi* ‘Tai Languages’, accessed on November 20, 2011. The URL is <http://baike.baidu.com/view/1004404.htm>.

The first problem to treating deu^{231} as a numeral is the fact that no other numerals can appear in the same position. The second problem is that deu^{231} cannot appear in the position that all other numerals can. The third problem is that deu^{231} in this position is optional, as in (12a), while other numbers cannot be omitted. The fourth problem is that, out of the three different forms referring to the number *one* in Maonan, i.e., $tɔ^{231}$, jit^{55} , and deu^{231} , only deu^{231} can be used in (12a). The contemporary deu^{231} and its cognates in related languages originated from an adjective meaning ‘single’ or ‘only’ (Zhang 2005:303). In all the languages with [C/M N Num(=1)], the only form allowed is always a native form of the numeral *one*, e.g., deu^{231} (Maonan), $diau^1$ (Bouyei) and other cognates in the Northern Tai sub-group and $nunɿ^{35}$ (Nung) and its cognates in the Central Tai sub-group. It is never the form of the numeral *one* borrowed from Chinese, e.g., jit^{55} (Maonan), $ʔit^7$ (Bouyei), and $ɛt^{35}$ (Nung).

Based on these facts, Zhang (2005) and Jiang (2007) insist that deu^{231} and similar elements in [C/M N Num] are not Num. Jiang (2007) further proposes that they behave more like adjectives. Maonan adjectives are indeed post-nominal, as in (13).

(13) a. ai^1 $zən^1$ $da:i^2$
 C person good
 ‘the good person’

 b. $tɔ^2$ $nək^8$ $voŋ^1$
 C bird tall
 ‘the tall bird’

However, deu^{231} as an adjective does not explain why it must always follow other adjectives, as in (14).

(14) a. ai^1 $zən^1$ $voŋ^1$ deu^{231}
 C person tall INDEF.SG
 ‘a tall person’

 b. $*ai^1$ $zən^1$ deu^{231} $voŋ^1$
 C person INDEF.SG tall

As cited in Aikhenvald (2000:100), in some Australian languages the numerals *one* and *two* are not used for counting and thus are not numerals in the strict sense (Dixon 1980:108). Hale (1975) further suggests that these forms are indefinite determiners, comparable to *a* and *some* in English. In Jingpho (Tibeto-Burman) there are also two forms traditionally associated with the numeral *one*, $lǎ^{55}ŋai^{51}$ and mi^{51} . Gu and Wu (2005) argue convincingly that while $lǎ^{55}ŋai^{51}$ is a genuine numeral, mi^{51} is like the determiner *a* in English. Heine (1997:72-76) in fact claims that the use of numeral *one* as an indefinite determiner is common in languages and offers a detailed description of its gradual process of grammaticalization. Mandarin is one notable example, as convincingly demonstrated by Chen (2003, 2004). I submit that the same analysis can be applied to all putative [C/M N Num(=1)] cases. The best proof comes from the fact that deu^{231} and the definite determiner ka^2 are in complementary distribution, as in (15). The two contrastive examples in (16) further confirm that deu^{231} is of the same category as ka^2 ‘that’ and $na:i$ ‘this’.

(15) a. **ai*¹ *zən*¹ *dɛu*²³¹ *ka*²/*na:i*
 C person INDEF.SG that/this

 b. **ai*¹ *zən*¹ *ka*² *dɛu*²³¹
 C person DEF INDEF.SG

(16) a. *ai*¹ *zən*¹ *voj*¹ *ka*²/*na:i*
 C person tall that/this
 ‘that/this tall person’

 b. *ai*¹ *zən*¹ *voj*¹ *dɛu*²³¹
 C person tall INDEF.SG
 ‘a tall person’

Simpson (2005:822), examining similar data in Thai and Nung that also involve a putative Num *one* in the usual position of D, reaches the same conclusion that this *one* in both languages ‘is coming to be an indefinite determiner which contrasts in its indefinite specification with the definiteness encoded by demonstratives’. However, he further proposes that this determiner-like *one* is base-generated in Num and then raises to D. This is an *ad hoc* stipulation, as nothing else base-generated in Num moves to D. For example, in (17a) and (18a), *two* is base-generated in Num and stays there, and (17b) and (18b) show that it cannot move to D is ill-formed. Furthermore, this *one* in fact cannot show up in Num, as shown in (17c) and (18c).

(17) a. *ja*⁴² *ai*¹ *zən*¹ (Maonan)
 2 C person
 ‘2 persons’

 b. **ai*¹ *zən*¹ *ja*⁴²
 C person 2
 (intended) ‘2 persons’

 c. **dɛu*²³¹ *ai*¹ *zən*¹
 1 C person
 (intended) ‘1 person’

(18) a. *slong* *ohng* *dehk* (Nung)
 2 CL child
 ‘2 children’
 (Saul and Wilson 1980)

 b. **ohng* *dehk* *slong* (Nung)
 CL child 2
 (intended) ‘2 children’

 c. **nuhng* *ohng* *dehk* (Nung)
 1 CL child
 (intended) ‘1 child’

A straightforward solution is to base-generate this *one* in D. Consequently, the putative [C/M N Num(=1)] sequence is more appropriately [C/M N D], thus not a contradiction to Universal 20A.

Finally, an explanation is needed as to why *deu*²³¹ and any of the forms in Maonan associated with the number *one*, i.e., *tɔ*²³¹ and *jit*⁵⁵, do not appear before in [Num C/M N] like other numbers. A more comprehensive look at other classifier languages reveals a typology of the omission of Num *one* in the C/M construction, shown in (19).

- (19) Typology of the omission of Num *one* in classifier phrases
- a. Prohibited: e.g., Min
 - b. Optional: e.g., Cantonese, Mandarin, Wu
 - c. Obligatory: e.g., Maonan

In Cheng and Sybesma's (2005) study of four Chinese languages, listed in (19a-b), Min, unlike the other three, is the only one that does not allow the omission of Num *one*. (20) and (21) are examples from Cantonese and Mandarin.

- (20) *Zek gau soeng gwo maalou.*
 CL dog want cross road
 'The dog wants to cross the road.'
 (Cheng and Sybesma 2005:9 [24c])

- (21) a. *Ta mai-le (yi) ben shu.*
 s/he buy-ASP 1 C book
 'S/he bought a book.'
- b. *Zhe (yi) xiang meigui hen zhong.*
 the 1 M-box rose very heavy
 'This (one) box of roses is very heavy.'

Maonan and other similar languages represent another type, one that *requires* the omission of Num *one*; alternatively, one may follow Her, Chen, and Tsai's (2015) account, where a silent counterpart of *one* is available in syntax. This phenomenon is in fact not unique to languages with the alleged [C/M N Num]; it is also found elsewhere, e.g., Amis (Formosan) (Tang 2004:389), Tat (Southwestern Iranian), and Khasi (Austroasiatic) (Greenberg 1990[1972]:168). A functional explanation is provided for this silent *one* in the next section.

In conclusion, the putative [C/M N Num(=1)] sequence in Maonan and other languages turns out to be [C/M N D], a construction where the real Num *one* is deleted or a silent element. The alleged [C/M N Num(=1)] is therefore merely an instance of [Num C/M N D], no contradiction to Universal 20A.

2.3 C/M-final more common than C/M-initial

Part 2 of Universal 20A, i.e., languages with the C/M-final, or [Num C/M], orders are far more common than those with C/M-initial, or [C/M Num], orders, has never been refuted. However, no studies have ever offered any concrete statistics on this observation, until Her et al. (2015), where among the 194 classifier languages in SMATTI, 165, or 85%, of them have C/M-final orders; see Table 2.

Table 2. Distribution of C/M-final & C/M-initial Languages in SMATTI

C/M-final	C/M-initial	Total Languages
165 (85%)	29 (15%)	194 (100%)

In sum, both Part 1 and Part 2 of Universal 20A are now confirmed: N does not come between Num and C/M, and languages with C/M-final orders far outnumber languages with C/M-initial orders.

3 Unifying C/M and motivating Universal 20A

Universal 20A implicitly assumes that in classifier languages C/M belong to the same syntactic category, even though there is no consensus in formal syntax. It is, however, well-established that C and M can be distinguished semantically. As concisely summarized by Zhang (2011) and Her (2012b), most syntactic accounts of Chinese assign C/M the same category in a unified structure, which should be left-branching to some but right-branching to others. Zhang (2011, 2013), however, proposes a split analysis where C involves a right-branching structure, and M, left-branching, while Li's (2011) account also requires both structures, the left-branching structure for the measuring reading and the right-branching one for the counting reading. In this section, 3.1 first establishes C/M as a single syntactic category, and then in 3.2, Her's (2012a) multiplicative theory is employed for their distinction, as well as their unification. The significant implications for motivating Universal 20A are revealed in 3.3.

3.1 C/M as a single syntactic category

The best evidence for unifying C/M under a single syntactic category comes from two universal facts: C/M are mutually exclusive and C/M share the same word orders. Compare the Mandarin examples in (22) and (23).

(22) *wu ben/ce/xiang/he zazhi*
 5 C/C/M-box/M-box magazine
 '5 magazines/boxes of magazines'

(23)a. **wu xiang ben zazhi*
 5 M-box C magazine
 b. **wu ben xiang zazhi*
 5 C M-box magazine
 c. **wu ce ben zazhi*
 5 C C magazine
 d. **wu xiang he zazhi*
 5 M-box M-box magazine

There are languages that have only Ms and no Cs, e.g., Archaic Chinese (Peyraube 1998) and Canglo Monpa (Tibeto-Burman) (Jiang 2006:50). Archaic Chinese (500BC-200AD, Shang to Han), for example, already has a number of Ms, including container measure, standard measures, and collective measures, appearing in the [N Num M] construction; a classifier system is not fully developed until Middle Chinese (201-1000, end of Han to end of Tang) (e.g., Huang 1964). C/M, being in the same syntactic category, naturally share the same word orders in history. (24) shows the four orders observed by Peyraube (1998), which can be reduced to (25).

(24) C/M's Four Word Orders in Chinese History (Peyaube 1998)

- a. N > Num > M
- b. N > Num > C
- c. Num > M > N
- d. Num > C > N

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

- a. N > Num > C/M
- b. Num > C/M > N

Chinese thus has undergone a word order change from N-initial to N-final, but has remained C/M-final throughout its 3,000 years of recorded history. Greenberg's (1975:30) observation that C/M do not deviate in word order has never been challenged (*Her to appear*), and indeed no exceptions are found in a database of 439 classifier languages (Her et al. 2015).

3.2 A multiplicative theory of C/M

As for the semantic distinction between C/M, the multiplicative theory found in Her (2012a) offers an insightful explanation, which is succinctly stated in (26).

(26) C/M Distinction in Mathematical Terms (Her 2012a:1679)

[Num X N] = [[Num × X] N], where X = C *iff* X = 1, otherwise X = M.

The relation between Num and C/M is seen as [*multiplier* × *multiplicand*] mathematically (Au Yeung 2005, Au Yeung 2007; Her 2012a). C/M differ in their respective value: C equals 1, M does not. This view thus unifies C/M under the concept of multiplicand and distinguishes them in terms of their difference in value. Therefore, (27a) returns (3 × 12) roses; (27b) returns (3 × 1) roses. Yet, crucially, the M *da* 'dozen' and the C *duo* are mutually exclusive, as in (28), even though the stacking of multiplicands is fine mathematically, as in (29). This means that the mutual exclusivity of C/M can only be a formal constraint not attributable to semantic factors.

(27) a. *san da meigui* (3 × 12 rose)

3 M-dozen rose
'3 dozens of roses'

b. *san duo meigui* (3 × 1 rose)

3 C rose
'3 roses'

(28) a. **san da duo meigui* (3 × 12 × 1 rose)

3 M-dozen C rose

b. **san duo da meigui* (3 × 1 × 12 rose)

3 C M-dozen rose

(29) a. *san-shi-wan duo meigui* (3 × 10¹ × 10⁴ × 1 rose)

300,000 C rose
'300,000 roses'

b. *san-shi-wan da meigui* (3 × 10¹ × 10⁴ × 12 rose)

300,000 C rose

'300,000 dozens of roses'

Semantically, *duo* ‘C’, *da* ‘dozen’, *shi* ‘10¹’, *wan* ‘10⁴’ can all function as multiplicands. Mathematically, there is no reason why they cannot all be stacked, and stacking of numeral bases (which function exactly as multiplicands) is in general allowed, though not without constraints, e.g., *3 hundred thousand* ($3 \times 10^2 \times 10^3$) and *san-shi-wan* ($3 \times 10^1 \times 10^4$). In contrast, though *duo* ‘C’ and *da* ‘dozen’ also function as multiplicands, they belong to the formal category C/M, not Num. The well-formed two examples in (29) thus indicate that the ill-formedness of C/M stacking in (28) has nothing to do with semantics and can only be seen as a formal constraint.

In short, C/M belong to the same syntactic category and function the same as a multiplicand, but at the same time they are two subcategories: C has the value *1*, and M, anything but *1*.

3.3 A functional motivation for Universal 20A

Having defended the implicit assumption in Universal 20A, that C and M belong to a single formal category, and its empirical content, one naturally asks the next logical question ‘why’. First, why should Num and C/M reject the intervention by N? A natural explanation is immediately obtainable within the multiplicative theory, i.e., intervention by N would interrupt the multiplicative unit, [*multiplier* × *multiplicand*], formed by Num and C/M. Second, why should there be more languages with the C/M-final orders than those with the C/M-initial orders? The motivation here is less obvious but likewise resides in the multiplicative function between Num and C/M.

Note that multiplication is transparent in a complex numeral, such as *three hundred* in English and *san-bai* ‘300’ in Chinese. In both examples, *three* and *hundred* form a multiplicative unit, [*n* × *base*].⁵ The order between *n* and *base*, just like the order between Num and C/M, is reversible, as either order between a multiplier and a multiplicand gives the same product. Both C/M-final and C/M-initial orders are found in languages; likewise, in the numeral systems in the world’s languages, some are base-final, while others, base-initial.

An important generalization is made by Greenberg (1990[1978]:292) that the order between Num and C/M and the order between *n* and *base* should ‘harmonize’. This ‘harmonization’ is restated in more explicit terms in (31), which assumes the two parameters in (30) (Her *to appear*, Her et al. 2015).⁶

(30) Base-parameter & C/M-parameter

- a. Base-parameter: base-final [*n base*] or base-initial [*base n*]
- b. C/M-parameter: C/M-final [Num C/M] or C/M-initial [C/M Num]

(31) Harmonization between base-parameter & C/M-parameter

- a. C/M-final order ⇒ base-final numerals
- b. C/M-initial order ⇒ base-initial numerals

Chinese is a good example, as it has always had a base-final decimal numeral system and a consistently C/M-final system. *Ibibio*, a Niger-Congo language, is also a good example, which has a C/M-initial [C/M Num N] order (Greenberg

⁵ I follow the terminology used by Comrie (2006), where the most prevailing pattern in the world’s numeral systems is depicted as (*n* × *base*) + *m*, where *m* < *base*.

⁶ In Her (to appear) and Her et al. (2015), this is called ‘synchronization’. In this paper, I use the term ‘harmonization’, following Greenberg (1990[1978]).

1990[1972]:185), and its numeral system is indeed base-initial (Chan 2015). A sample of its numerals is given in (31), taken from the *Ibibio Dictionary Online*.⁷

(32) Ibibio numbers

- a. ikie 100
- b. ikie iba 200 (100×2)
- c. ikie ita 300 (100×3)
- d. tosin 1000
- e. tosin iba 2000 (1000×2)
- f. tosin ita 3000 (1000×3)

However, again, no studies have ever offered any concrete statistics on this observation by Greenberg, until Her et al. (2015), where among the 194 classifier languages in SMATTI, only 5 exceptions are found. Base-C/M harmonization obtains in 189 languages, or 97.5%. See Table 3; the shaded areas are exceptions.

Table 3. Base-C/M Synchronization in SMATTI

	C/M-final	C/M-initial	Total Languages
Base-final	164 (84.5%)	4 (2%)	168 (86.5%)
Base-initial	1 (0.5)	25 (13%)	26 (13.5)
Total Languages	165 (85%)	29 (15%)	194 (100%)

Greenberg suggests that such harmonization exists within the domain of a language; however, restricting its domain of application to a nominal phrase is more insightful. A perfect example is Rabha, a Tibeto-Burman language which has all four word orders in Universal 20A; relevant facts confirming base-C/M harmonization are summarized in (33) (Joseph 2007).

(33) Base-C/M harmonization in Rabha (Joseph 2007)

- a. [Num C/M N] (C/M-final loan C/Ms & base-final loan numerals)
- b. [N Num C/M] (C/M-final loan C/Ms & base-final loan numerals)
- c. [C/M Num N] (C/M-initial native C/Ms & native numerals 1-3)
- d. [N C/M Num] (C/M-initial native C/Ms & native numerals 1-3)

Heavily influenced by Assamese, a dominant Indo-Aryan language with 30 million speakers in the region, Rabha has two numeral systems, one native, the other borrowed from Assamese, and two C/M systems, again one native, the other borrowed from Assamese, which is strictly base-final and C/M-final. The theory of harmonization predicts that the native C/M-initial orders co-occur only with numerals in a base-initial system. However, native numerals have died out except the numerals for *one*, *two*, and *three*. Fortunately, Joseph (2007:393, 844-845) documents an earlier native system that is base-initial, e.g., *gota-campa* [hundred-five] ‘500’, indicating that the three surviving numerals are in fact part of an extinct base-initial system. A similar scenario is found in Meche, also a Tibeto-Burman language influenced also by an Indo-Aryan language, Nepali in this case (Kiryu 2009).

So, back to the second question, why should there be more classifier languages with the C/M-final orders than those with the C/M-initial orders? The answer is now clear: because there are more languages with base-final numerals than those with

⁷ *Ibibio Dictionary Online*, accessed on 23 March, 2015, is available at: www.myibibio.com.

base-initial numerals. The data in Table 3 reveals that among 194 classifier languages in SMATTI, 168 (86.5%) are base-final, only 26 (13.5) are base-initial.

But then, why are there more languages with base-final numerals? One possible reason has to do with sociolinguistic factors. Comrie (2006) claims that among all grammatical features, the numeral system of a language is most vulnerable in language contact. Historically, language groups with base-final numerals, e.g., Sinitic, Tai, and Indo-Aryan, happened to be more powerful and have caused their neighboring languages with base-initial numerals to change to base-final.

Another possible answer has to do with syntax, i.e., base-final is the default setting in UG for the base-parameter, while base-initial is marked. This in turn means that the C/M-final order must likewise be the default setting. This possibility will be discussed in the next section, where a formal account of Universal 20A is pursued.

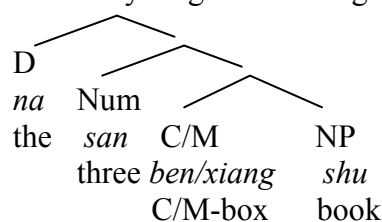
4 Pursuing a formal account of Universal 20A & Universal 20

Our focus is now on a proper account to derive the four attested orders in Universal 20A, as well as Cinque's Universal 20, which refers to the attested orders of D, Num, A, and N covered by Greenberg's Universal 20 and its attested exceptions dealt with in Cinque (2005). First, the three approaches to the structure of the classifier phrase are summarized in 4.1, followed by a discussion of the internal structure of complex numerals in 4.2. Cinque's (2005) account of Universal 20 is then extended, unsuccessfully, to Universal 20A in 4.3. The account I propose in 4.4 for Universal 20A is based on Abels and Neeleman's (2012) symmetric account of Cinque's Universal 20.

4.1 The three approaches: Left, right, and split

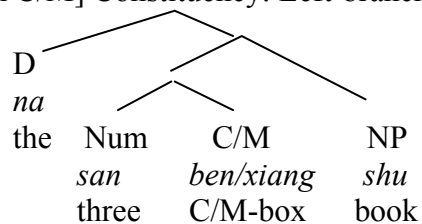
There are in general two different approaches to the syntactic status of C/M (e.g., Simpson 2005; Zhang 2011; Her 2012b). In one approach, Num and C/M are two distinct heads in a right-branching fashion, as shown schematically in (34), e.g., Tang (1990b:413), Cheng and Sybesma (1998, 1999), Borer (2005), Watanabe (2006), Zhang (2009), Huang et al. (2009), among others. In this more recent approach, popular among formalists, C/M and N form a constituent before merging with Num.

(34) [C/M N] Constituency: Right-branching Approach



The opposite takes place in the other approach, i.e., Num and C/M form a constituent before merging with N. In this traditional approach, supported by Greenberg (1975), some treat this [Num C/M] constituent as a complex functional head (e.g., Muromatsu 1998; Bhattacharya 1999, Bhattacharya 2001), while others see it as a phrasal constituent (e.g., Li and Thompson 1981:105; Paris 1981:105-117; Huang 1982; Croft 1994:151; Lin 1997:419; Fukui and Takano 2000; Hsieh 2008), as shown schematically in (35).

(35) [Num C/M] Constituency: Left-branching Approach⁸



As Simpson (2005) and Zhang (2011) both point out, there are surprisingly few explicit arguments justifying either approach, which is often assumed in the various works on C/M without much further discussion. Yet, Simpson and Zhang come to very different conclusions. Simpson (2005) argues against the [Num C/M] constituency and adopts the right-branching [C/M N] approach. Zhang (2011, 2013), however, claims that both structures are needed for Chinese. In the leftist camp, Hsieh (2008) and Her (2012b) stand out, and both offer specific arguments for [Num C/M] constituency. Li (2014) is the most prominent among the rightist and defends the [C/M N] approach. The three approaches are summarized in Table 4.

Table 4. Three Approaches to Universal 20A (Her 2012b)

Unified	Left-branching
	Right-branching
Split	Both structures necessary

4.2 The missing link: Structure of complex cardinals

In the literature on the structure of noun phrases in general and the [Num C/M N] constituency in particular, Num is always assumed to be simple numerals. The internal structure of a complex numeral as Num has thus not played any role in this left-or-right debate.⁹ However, as demonstrated by the base-C/M harmonization, the internal structure of Num may have significant consequences and thus must be taken into consideration.

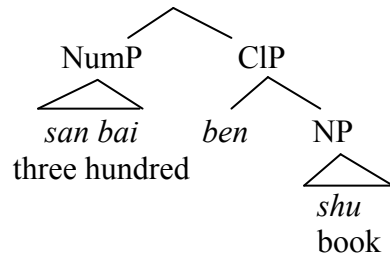
Since Hurford's (1975, 1987) pioneering work on numerals, the conventional assumption is that complex numerals like *three hundred* are phrasal constituents (e.g., Corver and Zwarts 2006; among others).¹⁰ An expression in Chinese such as *san bai ben shu* (three hundred C book) 'three hundred books' thus has two possible constituency analyses, depending on whether the overall nominal structure is right-branching or left-branching, as in (36a) and (36b), respectively.

(36) a. Complex numerals as constituents in a right-branching classifier phrase

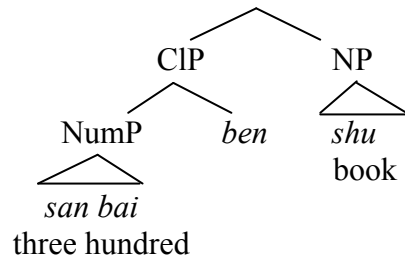
⁸ Note that in (35), the constituency of [[Num C/M] NP] is left-branching, though the overall tree is also partly right-branching.

⁹ The only two possible exceptions I am aware of are Her (2012b) and Li (2014), with the latter cites Ionin & Matushansky's (2006) right-branching solution to argue against the former's position that the left-branching approach better accounts for the multiplicative relation between Num and C/M. However, neither has shown how the internal structure of complex numerals favors one approach over the other.

¹⁰ Complex cardinals involving both multiplication and addition, e.g., *two hundred (and) ten*, will not be discussed in this paper, as addition involves some form of coordination and is not directly relevant to the issues under investigation here. See Ionin & Matushansky (2006: Section 4).

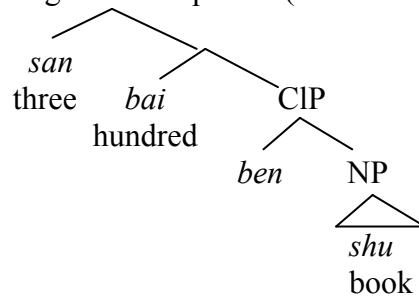


b. Complex numerals as constituents in a left-branching classifier phrase

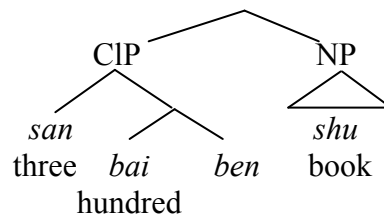


However, this constituency approach to complex numerals has recently been challenged by Ionin and Matushansky's (2006) non-constituency account, where an expression such as *three hundred books* is analyzed as [*three* [*hundred* [*books*]]], thus having the syntactic structure of head and complement. For classifier languages like Chinese, Ionin and Matushansky (2006:328 [22a] & [22b]) remain uncommitted and leave open two possible accounts, as in (37a) and (37b), one right-branching like that of English, and the other left-branching, unlike that of English.¹¹

(37) a. Right-branching classifier phrase (Ionin and Matushansky 2006:328 [22a])¹²



b. Left-branching classifier phrase (Ionin and Matushansky 2006:328 [22b])



¹¹ The two trees are adopted from Ionin & Matushansky (2006:328 (22a) and (22b)), respectively, with some simplification.

¹² In Ionin & Matushansky (2006), the right-branching structure of (37b) is labeled NumP with [\pm plural] as the head Num⁰. Not to confuse their label Num, which refers to number distinction of singular and plural, and our use of Num, which refers to numerical quantifiers, I have left out some of the labels in (37b).

In this paper, the constituency versus non-constituency is not our concern; rather, the focus is on the left-or-right debate over the constituency of the nominal structure. As it will become clear momentarily, the left-branching options jibe much better with Universal 20A.

4.3 Rejecting the split approach

The most prominent advocates of the split approaches are Zhang (2011, 2013) and Li (2011), though their accounts are drastically different, with Zhang arguing that individual, individuating, and kind C/Ms have a right-branching structure, and others, left-branching, and Li insisting that the left-branching structure is needed for all C/M's measuring reading and right-branching for all C/M's counting reading. 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) also proposes that different classifier languages may have different structures. Ionin and Matushansky (2006:328, fn.14), though not taking sides on this left-or-right debate, wade in and suggest that 'more than one structure may be available in a given language and cross-linguistically' because of the word order variation.

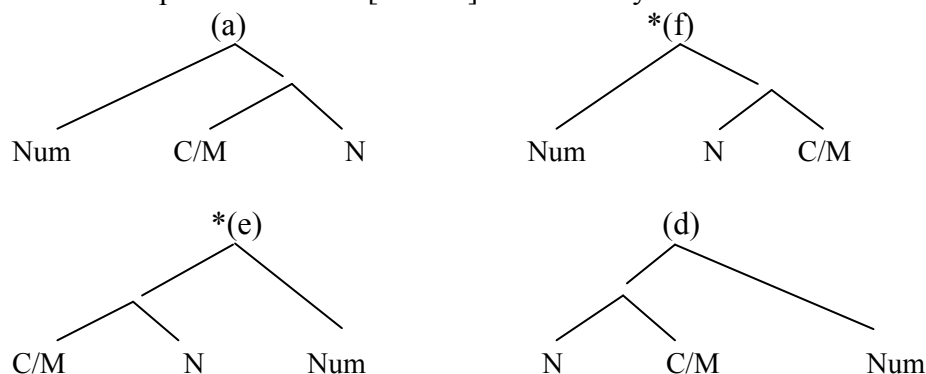
In this study I will not review the specific arguments such split accounts pose, or the specific arguments offered previously by either the leftist camp or the rightist camp, and will refer the readers to the references cited. I will point out, however, that between a unified account and a split account, everything else being equal, Occam's Razor eliminates the latter, as the latter is less restricted. More importantly, everything is not equal. Let's look at the four orders in Universal 20A again, repeated in (38).

(38) Four word orders in Universal 20A

- a. [Num C/M N]
- b. [N Num C/M]
- c. [C/M Num N]
- d. [N C/M Num]
- e. *[C/M N Num]
- f. *[Num N C/M]

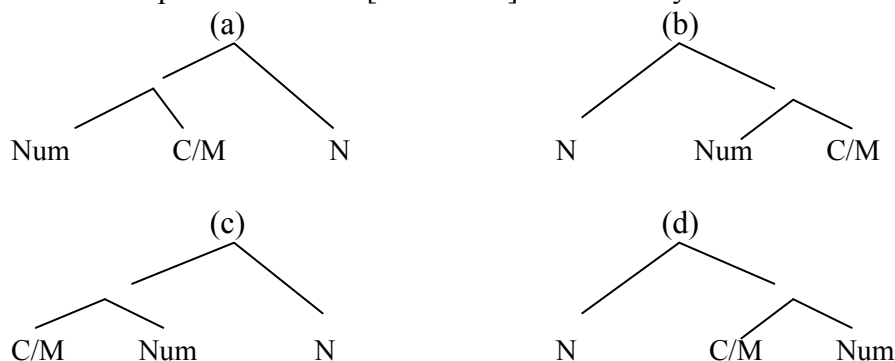
As shown in (39), assuming [C/M N] constituency and the head parameter, two of the orders predicted, (38e) and (38f), are unattested, while two of the attested orders, (38b) and (38c), are missing. One thus must adopt a derivational framework, preferably without the head parameter, and allow movement to work out only the desired results.

(39) Four orders predicted under [C/M N] constituency



As shown in (40), assuming [Num C/M] constituency and the head parameter, the orders predicted are right on, no over-generation, nor under-generation. One can of course still abandon the head parameter and insist on generating the right results from a single underlying order by movement, which is however a more costly operation than base-generation.

(40) Four orders predicted under [Num C/M] constituency



Detrimental to the split approach is that it must ‘over kill’, as inevitably all the well-formed orders are structurally ambiguous. The only way to avoid such redundancy is to adopt either (39) or (40), i.e., a unified approach.

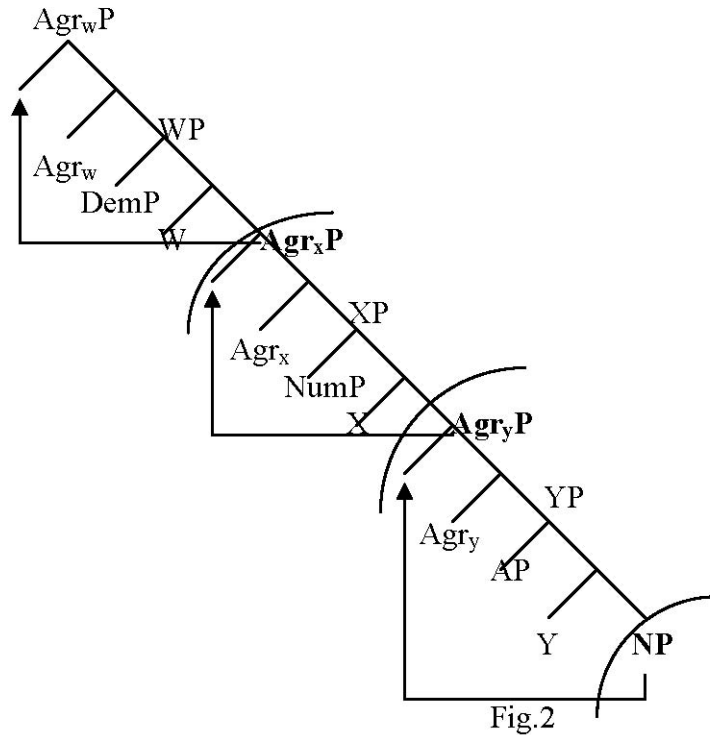
As mentioned earlier, in languages that have both Cs and Ms and also have more than one word order, C/M are always in unison. Under a split account that assigns two different structures to Cs and Ms, such unison is rendered accidental, while in a unified account, such unison is a natural consequence.

With the split approach rejected, our focus is now back on the left-or-right verdict. If the head parameter is allowed, then (40) is the clear winner. If movement is allowed, then the battle is not over just yet. The strategy of finding the ultimate solution is to up the ante and bring in Universal 20 and piggyback on a successful account and see which of (39) and (40) fares better. Fortunately, two such formal accounts exist, Cinque (2005) and Abels and Neeleman (2012), and precisely along the theoretical bend of (39) and (40), respectively, no less. Cinque’s account of Universal 20 assumes an antisymmetrical framework, where specifier, head, and complement can only be base-generated in that strict order and variation can only be accounted for by movement. Abels and Neeleman (2012), unsatisfied with Cinque’s (2005) theoretical assumptions, propose a more restrictive framework that allows the head parameter, or symmetry.¹³

4.4 Extending Cinque (2005) to Universal 20A

The empirical content of Cinque’s Universal 20 is 14 attested orders among D, Num, A, and N, out of 24 possibilities, as shown in Table 5, where the shaded area indicates orders unattested in languages and the columns are arranged according the relative position of N.

¹³ Dryer (2009) offers an informal account as an alternative to Cinque (2005) and points out a few more exceptions to Greenberg’s Universal 20 not covered by Cinque (2005), i.e., Num-N-Dem-A (III_f in Table 5), Dem-A-Num-N (I_b), and N-Num-Dem-A (IV_f). We will leave these exceptions for future studies and also will not discuss Dryer’s informal account, which is stated in terms of semantic categories independent of their syntactic realization. Instead, we shall focus on Cinque’s (2005) and Abels and Neeleman’s (2012) formal accounts and the same range of data covered in both.



However, besides the above three orders, there are 11 more reported. Cinque (2005) is able to derive all of the 14 attested orders, without also deriving any of the 10 unattested ones. This impressive account assumes Kayne's (1994) theory that symmetric base-generation of modifiers is impossible and also the following:

- (41) a. Order of merge: [...[WPDem ...[XPNum ...[YP A [NP N]]]]]
- b. Parameters of movement:
- i) No movement (unmarked), or
 - ii) NP movement plus Pied-piping of the *whose picture*-type (unmarked), or
 - iii) NP movement without Pied-piping (marked), or
 - iv) NP movement plus Pied-piping of the *picture of who*-type (more marked)
 - v) *total* (unmarked) vs. *partial* (marked) movement of the NP with or without Pied-piping (in other words, the NP raises all the way up or just partially around its modifiers).
 - vi) Neither head movement nor movement of a phrase not containing the NP are possible (except perhaps for focus-related movements to a DP initial position).
 - vii) Severity of markedness: partial movement < movement without Pied-piping < movement with Pied-piping of *picture of who*-type

Also directly relevant to our discussion is Cinque's (2005) proposal to include the category of classifier in a universal order of merge within the overall nominal constituent, shown in (42). Thus, limiting the elements to what concerns us immediately, i.e., Num, C/M, and N, the universal order of merge is in (43a), a right-branching structure compatible with the right-branching solution (37a) suggested by Ionin and Matushansky (2006), as shown in the example of (43b).

- (42) Universal order of Merge in the nominal structure (Cinque 2005:328)¹⁴
 [Q_{univ.} [Dem. [Num_{ord.} [RC. [Num_{card.} [Clf. [A. NP]]]]]]]]
- (43) Universal order of Merge among Num, C/M, N
 a. [Num_{card.} [C/M. N]]
 b. [*three* [*hundred.* [C/M. N]]]

Now we can apply Cinque's framework to Universal 20A and see if the same success can be achieved, namely if all and only the attested orders can be derived.

- (44) C/M Word Orders Derived within Cinque's (2005) Account
- (a) [Num C/M N] (*attested, very many languages, e.g., Mandarin*)
 N does not raise; unmarked: very many languages.
Prediction correct.
- (b) [N Num C/M] (*attested, many languages, e.g., Thai*)
 N raises around C/M and Num (without Pied-piping, thus marked); one marked option: fewer languages.
Prediction correct.
- (c) [C/M Num N] (*attested, few languages, e.g., Ibibio*)
 N does not move, but the two elements to its left are in the wrong order of merge; cannot be derived.
Prediction incorrect; under-generates.
- (d) [N C/M Num] (*attested, few languages, e.g., Jingpho*)
 N moves around C/M and Num with Pied-piping of the *whose picture*-type (partially, in order not to separate C/M and Num, thus marked); one marked option: fewer languages.
Prediction correct.
- (e) [C/M N Num] (*unattested, no languages*)
 N raises one notch, but the two elements to its left are in the wrong order of merge; cannot be derived. Or, N raises one notch with Pied-piping of *picture of who*-type (thus marked), one marked option: fewer languages.
Prediction incorrect; over-generates.
- (f) [Num N C/M] (*unattested, no languages*)
 N moves around C/M (partially, thus marked), with vacuous Pied-piping of the *whose picture*-type (unmarked), one marked option (degree 1): fewer languages.
Prediction incorrect; over-generates.

With three misses out of six targets, the account's performance is no better than chance. Cinque's (2005) account thus fails to account for Universal 20A, even when only simple numerals are considered in [Num C/M N].

However, as an anonymous reviewer points out, the left-branching option of classifiers in view of Universal 20A is still available under Cinque's (2005) analysis of Universal 20. Yet, given the strictly antisymmetric right-branching [spec [head complement]] structure assumed, the left-branching structure of [[Num C/M] N] is not

¹⁴ See Borer (2005), Cinque (2006), and Kayne (2006), among others, for the claim that the category of classifier should also exist in languages that have no classifiers.

straightforwardly compatible. A possible antisymmetric solution is to view [Num C/M] as a complex functional head (e.g., Tang 1990a; Muromatsu 1998; Bhattacharya 1999, Bhattacharya 2001), but then copious movements are needed to account for classifier word order typology, or Universal 20A (see Liu 2006 for an example of such movements).

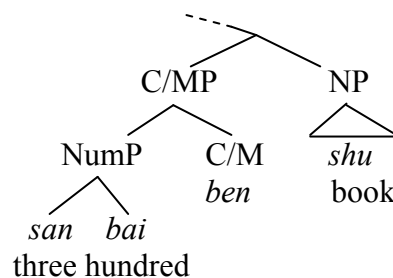
Furthermore, with complex numerals taken into consideration, the derivation of Universal 20A becomes more complicated. Incorporating a complex numeral, which has the form [*n base*] or [*base n*], the six mathematically possible orders of Num, C/M, N are thus expanded to twelve, as in (45), where only four are well-formed, which observe the base-C/M harmonization.

(45) Twelve Possible Word Orders of [*n, base, C/M, N*]

- a. √ [*n base C/M N*] base-C/M harmonization
- a' * [*base n C/M N*]
- b. √ [*N n base C/M*] base-C/M harmonization
- b' * [*N base n C/M*]
- c. √ [*C/M base n N*] base-C/M harmonization
- c' * [*C/M n base N*]
- d. √ [*N C/M base n*] base-C/M harmonization
- d' * [*N C/M n base*]
- e. * [*C/M N base n*]
- e' * [*C/M N n base*]
- f. * [*n base N C/M*]
- f' * [*base n N C/M*]

We shall first explore the constituency approach to complex numerals, under Cinque's (2006) antisymmetric framework, allowing [Num C/M] as a constituent. Given an expression such as *san bai ben shu* (three hundred C book) 'three hundred books', assuming the order of merge in (42) and the order of [*n base*], (45a) is the only base-generated order, shown schematically in (46). To derive by movements the other three attested orders in (45), however, is complicated by the internal structure of the C/MP as well as that of the complex numeral, NumP, as discussed in (47).

(46)



(47) C/M Word Orders Derived within Cinque's (2005) Account

- (a) [[[*n base*] C/M] N] (*attested, very many languages, e.g., Mandarin*)

N does not raise; unmarked: very many languages.

Prediction correct.

- (b) [N [[*n base*] C/M]] (*attested, many languages, e.g., Thai*)

N raises around the complex head C/MP (without Pied-piping, thus marked); one marked option: fewer languages.

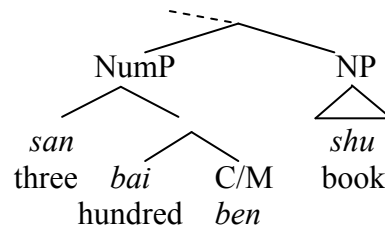
Prediction correct.

- (c) [[C/M **base** n] N] (*attested, few languages, e.g., Ibibio*)
 N does not move, but the two elements to its left, C/M and [base n], are in the wrong order of merge; base and n are also in wrong order of merge; cannot be derived.
Prediction incorrect; under-generates.
- (d) [N [C/M **base** n]] (*attested, few languages, e.g., Jingpho*)
 N raises around C/MP (without Pied-piping, thus marked), but the two elements C/M and [base n] are in the wrong order of merge; base and n are also in the wrong order of merge; cannot be derived.
Prediction incorrect; under-generates.

The good news is that the account does not over-generate, as all the unattested orders are ruled out due to a wrong order of merge involving the two elements C/M and [base n] and/or the two elements base and n. The bad news is, as seen in (47), the theory is too restrictive and thus under-generates.

We now test the non-constituency approach to complex numerals under Cinque's (2006) antisymmetric framework, likewise assuming [Num C/M] as a constituent. Again, the base-generated (45a) is straightforward, as shown schematically in (48).

(48)



- (49) C/M Word Orders Derived within Cinque's (2005) Account
- (e) [[[n **base**] C/M] N] (*attested, very many languages, e.g., Mandarin*)
 N does not raise; unmarked: very many languages.
Prediction correct.
- (f) [N [n [**base** C/M]]] (*attested, many languages, e.g., Thai*)
 N raises around the complex head NumP (without Pied-piping, thus marked); one marked option: fewer languages.
Prediction correct.
- (g) [[[C/M **base**] n] N] (*attested, few languages, e.g., Ibibio*)
 N does not move, but the three elements to its left, C/M, base and n are in the wrong order of merge; cannot be derived.
Prediction incorrect; under-generates.
- (h) [N [[C/M **base**] n]] (*attested, few languages, e.g., Jingpho*)
 N raises around NumP (without Pied-piping, thus marked), but the three elements to its left, C/M, base and n are in the wrong order of merge; cannot be derived.
Prediction incorrect; under-generates.

This account is likewise unsatisfactory and under-generates. In short, Cinque's (2005) antisymmetric account of Universal 20 is too restrictive to accommodate a left-branching [[Num C/M] N], whether Num is a simple numeral or complex numeral,

in view of Universal 20A. We shall now pursue a solution in an alternative symmetric framework.

4.5 Extending Abels and Neeleman (2012) to Universal 20A

Abels and Neeleman (2012) is a reaction to Cinque (2005). The empirical content dealt with here is the same, as shown in Table 6, where the orders are arranged differently but the shaded area again indicates unattested orders. The table is divided into two parts, each with a pair of columns, and the orders in each pair of columns are mirror images of each other. A pair of mirror images belongs to the symmetry part if both orders are attested or both are unattested, thus symmetrical. A pair of mirror images is asymmetrical if one is attested, and the other, unattested.

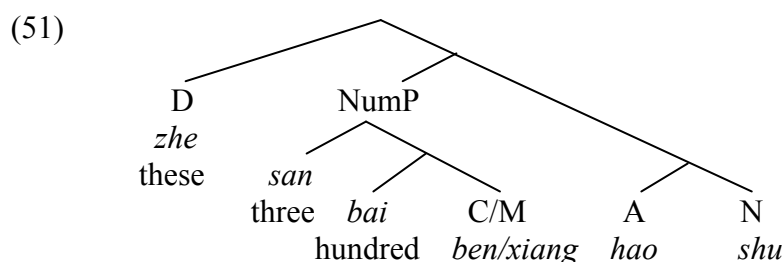
Table 6. Universal 20 Revised
Attested orders of D, A, Num, and N (Abels and Neeleman 2012)

	Symmetry		Asymmetry	
	I	II	III	IV
a.	Dem ₁ Num ₂ A ₃ N ₄	N ₄ A ₃ Num ₂ Dem ₁	Dem ₁ N ₄ Num ₂ A ₃	A ₃ Num ₂ N ₄ Dem ₁
b.	Dem ₁ Num ₂ N ₄ A ₃	A ₃ N ₄ Num ₂ Dem ₁	N ₄ Dem ₁ Num ₂ A ₃	A ₃ Num ₂ Dem ₁ N ₄
c.	Dem ₁ A ₃ N ₄ Num ₂	Num ₂ N ₄ A ₃ Dem ₁	A ₃ N ₄ Dem ₁ Num ₂	Num ₂ Dem ₁ N ₄ A ₃
d.	Dem ₁ N ₄ A ₃ Num ₂	Num ₂ A ₃ N ₄ Dem ₁	N ₄ Num ₂ A ₃ Dem ₁	Dem ₁ A ₃ Num ₂ N ₄
e.	A ₃ Dem ₁ Num ₂ N ₄	N ₄ Num ₂ Dem ₁ A ₃	N ₄ Dem ₁ A ₃ Num ₂	Num ₂ A ₃ Dem ₁ N ₄
f.	A ₃ Dem ₁ N ₄ Num ₂	Num ₂ N ₄ Dem ₁ A ₃	N ₄ A ₃ Dem ₁ Num ₂	Num ₂ Dem ₁ A ₃ N ₄

- (50) a. The underlying hierarchy Dem > Num > A > N is assumed (> = c-command);
 b. all (relevant) movements move a subtree containing N;
 c. all movements target a c-commanding position;
 d. all movements are to the left.

The eight orders under symmetry, i.e., (Ia-d) and (IIa-d), are all base-generated by simply applying the head parameter to (50a), and the unattested pairs (Ie-f) and (IIe-f) cannot be base-generated. The remaining six in (III) are all derived by movement in accordance with (50). This account satisfactorily generates the same range of data as Cinque (2005), but is able to do so in a general framework that is more restrictive and more elegant. Can the same success be achieved for Universal 20A?

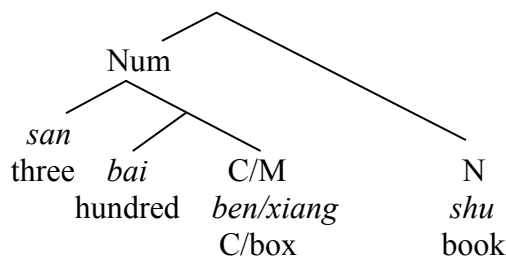
Let's bring in the complex numerals and test the left-branching [[Num C/M] N] approach. In (51), a complex numeral is shown not be a constituent, as suggested by Ionin and Matushansky (2006) and integrated in the default structure assumed by Abels and Neeleman. Note, however, whether or not a complex numeral forms a constituent on its own does not affect the outcome of the argumentation pursued here. The crucial factor is that [Num C/M] is a constituent, [C/M N] is not.



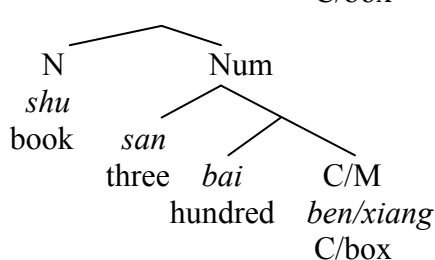
C/box good book

NumP, as a single constituent in this overall structure Dem > Num > A > N, has no effect on the generation of Cinque's Universal 20. Thus, to generate Universal 20A, we only need to be concerned with the relevant three elements, as in (52a), and its N-initial counterpart (52b).

(52) a.



b.



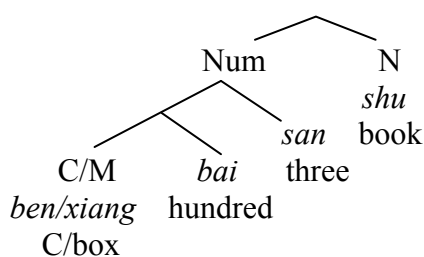
Recall that a taxonomy of the four orders under Universal 20A can be derived by two binary parameters: C/M-initial/CM-final and N-initial/N-final, as shown in Table 1, repeated in Table 7.

Table 7. Taxonomy of C/M Word Orders

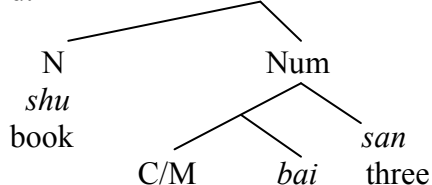
	N-FINAL	N-INITIAL
C/M-FINAL	(A) [Num C/M N]	(B) [N Num C/M]
C/M-INITIAL	(C) [C/M Num N]	(D) [N C/M Num]

The orders (A) and (B) are base-generated with the N-final and N-initial options, in (52a) and (52b), respectively, under the C/M-final option. For the other two orders, (C) and (D), we look inside of Num and this time base-generate the C/M-initial order, coupled with the N-final and N-initial options, as in (52c) and (52d), respectively.

(52) c.

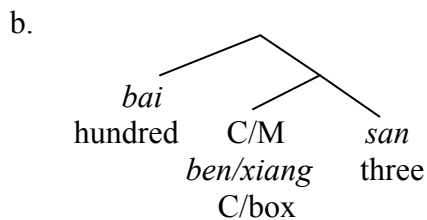
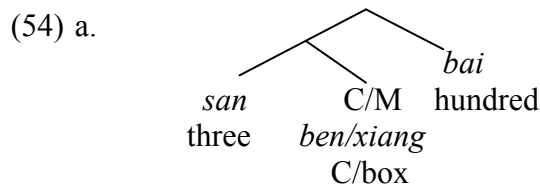
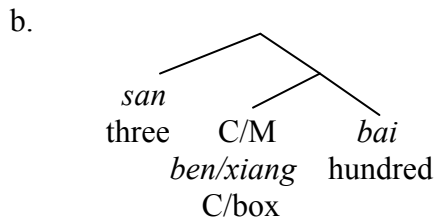
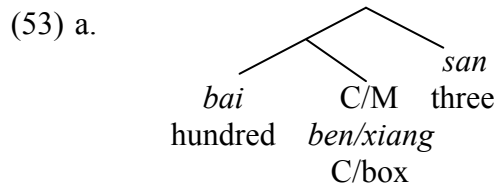


d.



ben/xiang hundred
C/box

Note that a complex numeral is used here deliberately to demonstrate the base-C/M harmonization. Formally, *n* subcategorizes for *base* as complement, which in turn subcategorizes for C/M as complement. Functionally, the two parts of *san-bai* ‘300’, i.e., *n* and *base*, must stay adjacent for multiplication to hold, and base and C/M as multiplicands must also stay adjacent. The base-C/M harmonization obtains as a consequence. The universal ill-formedness of the structures in (53a-b) and (54a-b) likewise obtains.



The final issue is how to account for Part 2 of Universal 20A, i.e., C/M-final languages are more common than C/M-initial ones. The functional explanation put forth earlier is that, given the base-C/M harmonization, it is merely a consequence of the fact that base-final classifier languages are more common than base-initial classifier languages, a fact attributed to sociolinguistic factors. Perhaps a formal explanation can be achieved too. The D-Num-A-N order is assumed to be the default by Abels and Neeleman (2012), called the ‘straight’ order. It is also the only base-generated order in Cinque (2005), all other orders derived via movement, a more costly operation than merger. Following the same intuition, the *n-base(-C/M)* order can be seen as the default order in UG and that explains the higher frequency of base-final numeral systems.

5 Concluding remarks

The four attested word orders of the three elements, number (Num), classifier/measure word (C/M), and noun (N): [Num C/M N], [N Num C/M], [N C/M

Num], and [C/M Num N], discovered first by Greenberg (1990[1972]), can be reduced to a simple generalization of Part 1 of (55), which, together with his observation regarding the distribution of these word orders in languages, as in Part 2 of (55), is dubbed Greenberg's Universal 20A.

(55) Greenberg's Universal 20A

Part 1: Of the three elements Num, C/M, and N, any order is possible as long as Num and C/M are adjacent.

Part 2: There are many more languages with Num > C/M orders than languages with C/M > Num orders.

This paper first defends Part 1 by dismissing the two alleged exceptions, one in Ejagham, the other in some Tai-Kadai and Tibeto-Burman languages, and confirms Part 2 with concrete statistics from SMATTI (short for 'Sinitic, Maio-Yao, Austroasiatic, Tibeto-Burman, Tai-Kadai, and Indo-Aryan'). A multiplicative theory is then employed to justify C/M as a single syntactic category, assumed in Universal 20A, and also to provide a functional motivation.

Complex numerals have long been neglected in previous studies on numeral classifiers; thus, this study adopts Ionin and Matushansky's (2006) account of complex numerals in its pursuit of a formal account of Universal 20A. There exist two formal accounts, Cinque (2005) and Abels and Neeleman (2012), that account for all the attested word orders of D, Num, A, and N, thus Universal 20 and all its genuine exceptions. Our strategy is to extend these accounts to Universal 20A and thus kill two birds with one stone. However, Cinque's antisymmetrical account is shown to be inadequate for Universal 20A, as it over-generates and under-generates. Abels and Neeleman, unsatisfied with Cinque's (2005) theoretical assumptions, propose a more restrictive framework that allows both symmetry and asymmetry, and is able to come up with a more elegant account for Cinque's Universal 20. Implementing this framework with the left-branching solution of complex numerals and C/M that Ionin and Matushansky's (2006) suggest proves satisfactory. The account not only generates all and only the attested four orders, it is also fully compatible with the multiplicative theory that provides the functional motivation for Universal 20A.

Typology is fittingly a rich ground for testing theories and frameworks that aspire to be compatible with UG. This study further demonstrates that one need not, and in fact should not, reinvent the wheel when attempting to account for a typological phenomenon formally. The accumulative nature of scientific advances suggests that the winning strategy is to extend a theory or framework that has been proven successful empirically and sufficiently restrictive. In that sense, this current study on Universal 20A also serves as an endorsement of Abels and Neeleman's (2012) account of Universal 20 as well as the overall framework. Finally, this study demonstrates that formal accounts and functional explanations not only can co-exist, they need each other, as an insightful account in one tradition can in fact facilitate the exploration in the other.

References

Abels, Klaus & Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15(1). 25-74.

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Au Yeung, Ben Wai Hoo. 2005. *An interface program for parameterization of classifiers in Chinese*. Hong Kong University PhD. dissertation.
- Au Yeung, Ben Wai Hoo. 2007. Multiplication basis of emergence of classifiers. *Language and Linguistics* 8(4). 835-861.
- Bhattacharya, Tanmoy. 1999. *The structure of the Bangla DP*. University College London PhD. dissertation.
- Bhattacharya, Tanmoy. 2001. Numeral/quantifier-classifier as a complex head in semi-lexical heads. In Henk van Riemsdijk & Norbert Corver (eds.), *Semi-lexical heads*, 191-222. Mouton de Gruyter.
- Borer, Hagit. 2005. *Structuring sense*, Vol. 1: *In name only*. Oxford: Oxford University Press.
- Chan, Eugene S. L. 2015. *Numeral system of the world's languages*. Available online at <http://lingweb.eva.mpg.de/numeral/>.
- Chen, Ping. 2003. Indefinite determiner introducing definite referent: A special use of *yi* 'one' + classifier in Chinese. *Lingua* 113(12). 1169-1184.
- Chen, Ping. 2004. Identifiability and definiteness in Chinese. *Linguistics* 42(6). 1129-1184.
- Cheng, Lisa L.-S. & Rint Sybesma. 1998. *Yi-wan Tang, Yi-ge Tang*: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies*, New Series 28(3). 385-412.
- Cheng, Lisa L.-S. & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4). 509-542.
- Cheng, Lisa L.-S. & Rint Sybesma. 2005. Classifiers in four varieties of Chinese. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 259-292. Oxford: Oxford University Press.
- Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36(3). 315-332.
- Cinque, Guglielmo. 2006. Are all languages 'numeral classifier languages'? *Rivista di grammatica generativa* 31. 119-122.
- Comrie, Bernard. 2006. Numbers, language, and culture. Paper presented at the Jyvaskyla Summer School, Jyvaskyla, Finland. 24 July-11 August.
- Corver, Norbert & Joost Zwarts. 2006. Prepositional numerals. *Lingua* 116(6). 811-35.
- Croft, William. 1994. Semantic universals in classifier systems. *Word* 45. 145-171.
- Dixon, R. M. W. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.
- Doetjes, Jenny. 2012. Count/mass distinctions across languages. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, Vol. 3, 2559-2580. Berlin: De Gruyter.
- Dryer, Mathew S. 2009. The order of demonstrative, numeral, adjective, and noun: An alternative to Cinque. Paper presented at the Conference on Theoretical Approaches to Disharmonic Word Orders, University of Newcastle, 30 May-1 June, 2009.
- Fukui, Naoki & Yuji Takano. 2000. Nominal structure: An extension of the symmetry principle. In P. Svenonius (ed.), *The derivation of VO and OV*, 219-254. Amsterdam: John Benjamins.

- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (ed.), *Universals of language*, 73-113. Cambridge, Mass.: MIT Press.
- Greenberg, Joseph. 1990[1972]. Numerical classifiers and substantival number: Problems in the genesis of a linguistic type. In K. Denning & Kemmer (eds.), *On language. Selected writings of Joseph H. Greenberg*, 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. 1975. Dynamic aspects of word order in numeral classifier. In Charles Li (ed.), *Word order and word order change*, 27-46. Austin, TX: University of Texas Press.
- Greenberg, Joseph. 1990[1978]. Generalizations about numeral systems. In K. Denning & Kemmer (eds.), *On language. Selected writings of Joseph H. Greenberg*, 271-309. Stanford, CA: Stanford University Press. [First published 1978 in *Universals of the Human Language*, Vol. 3. 249-295. Stanford, CA: Stanford University Press.]
- Gu, Yang & 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 Jingfang Li & Suhua Hu (eds.), *Hanzangyu liangci yanjiu* [Studies of classifiers in Sino-Tibetan], 55-89. Beijing: The Central University of Nationalities Press.
- Hale, Kenneth L. 1975. Gaps in grammar and culture. In M. Dale Kinkade, Kenneth L. Hale & Oswald Werner (eds.), *Linguistics and anthropology. In honor of C.F. Voegelin*. Lisse: The Peter de Ridder Press.
- Heine, Bernd. 1997. *Cognitive foundations of grammar*. New York: Oxford University Press.
- Her, One-Soon. 2012a. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122(14). 1668-1691.
- Her, One-Soon. 2012b. Structure of classifiers and measure words: A lexical functional account. *Language and Linguistics* 13(6). 1211-1251.
- Her, One-Soon. To appear. Structure of numerals and numeral classifiers in Chinese: Historical and typological perspectives and cross-linguistic implications. To appear in *Language and Linguistics*.
- Her, One-Soon, Ching Perng Chen & Hui-Chin Tsai. 2015. Justifying silent elements in syntax: The case of a silent numeral, a silent classifier, and two silent nouns in Mandarin Chinese. *International Journal of Chinese Linguistics* 2(2). 193-226.
- Her, One-Soon & Chen-Tien Hsieh. 2010. On the semantic distinction between classifiers and measure words. *Language and Linguistics* 11(3). 527-550.
- Her, One-Soon, Hui-Chin Tsai, Kun-Han Lin, Marc Tang, & Meng-Chang Lee. 2015. Unification of numeral bases and numeral classifiers: Evidence from SMATTI. Paper presented at the Workshop on the syntax and semantics of numerals, the 48th Annual Meeting of the Societas Linguistica Europaea. Leiden University Centre for Linguistics (LUCL), Leiden, 2-5 September.
- 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.
- Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. MIT PhD. dissertation.

- Huang, C.-T. James, Audrey Li & Yafei Li. 2009. *The syntax of Chinese*. Cambridge University Press.
- Huang, C.-T. James & Masao Ochi. 2011. Classifiers and nominal structure: A parametric approach and its consequences. In Mingle Gao (ed.), *Universals and variation: Proceedings of Glow-in-Asia VIII*, 191-197. Beijing Languages and Culture University Press.
- Huang, Zaijun. 1964. Cong jiawen jinwen liangci de yingyong kaocha hanyu liangci de qiyuan yu fazhan [A study of the origin and development of classifiers in Chinese from the perspectives of the use of classifiers in oracle bones and bronze inscriptions]. *Zhongguo Yuwen* 1964(6). 432-441.
- Hurford, James. 1975. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
- Hurford, James. 1987. *Language and number*. Oxford: Blackwell.
- Ionin, Tania & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23. 315-360.
- 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.
- Jiang, Ying. 2006. *Hanzang yuxi mingliangci yanjiu* [A study of nominal classifiers in Sino-Tibetan]. The Central University of Nationalities PhD. dissertation.
- Jiang, Renping. 2007. *Jishuci he xushuci de leixingxue yanjiu* [A typological study of cardinals and ordinals]. Nanchang University MA thesis.
- Joseph, Thattil Umbavu Varghese. 2007. *Rabha*. Leiden: Brill Academic Publishers.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, Mass., MIT Press
- Kayne, Richard S. 2006. On parameters and on principles of pronunciation. In H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz & J. Koster (eds.), *Organizing grammar. Linguistic studies in honor of Henk van Riemsdijk*, 289-299. Berlin: Mouton de Gruyter.
- Kihm, Alain. 2005. Noun class, gender, and the lexicon-syntax-morphology interfaces: A comparative study of Niger-Congo and Romance languages. In Guglielmo Cinque & Richard Kayne (eds.), *Handbook of comparative syntax*, 459-512. Oxford University Press.
- Kiryu, Kazuyuki. 2009. On the rise of the classifier system in Newar. In Yasuhiko Nagano (ed.), *Issues in Tibeto-Burman historical linguistics*, 51-69. *Senri Ethnological Studies* (75). Osaka, Japan: National Museum of Ethnology.
- Li, Charles & Sandra Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, Xu-Ping. 2011. *On the semantics of classifiers in Chinese*. Bar-Ilan University PhD. dissertation.
- Li, Yen-Hui Audrey. 2014. Structure of noun phrases: Left or right? *Taiwan Journal of Linguistics* 12(2). 1-32.
- Lin, Jo-wang. 1997. Noun phrase structure in Mandarin Chinese: DP or NP?. *Chinese Language and Linguistics* 3. 401-434.
- Liu, Hongyong. 2006. The Structure of complex nominals: classifiers, possessives and relatives. The Chinese University of Hong Kong PhD. dissertation.
- Muromatsu, Keiko. 1998. *On the syntax of classifiers*. University of Maryland PhD. dissertation.
- 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 Benjamin K. T'sou (ed.) *Studia Linguistica Serica*, 131-145. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.
- Saul, Janice E. & Nancy F. Wilson. 1980. *Nung Grammar*. Summer Institute of Linguistics, University of Texas at Arlington.
- Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. In Guglielmo Cinque & Richard Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, 806-838. Oxford University Press.
- Tang, Chih-Chen Jane. 1990a. *Chinese phrase structure and the extended X'-theory*. Cornell University PhD. dissertation.
- Tang, Chih-Chen Jane. 1990b. A note on the DP analysis of Chinese noun phrases. *Linguistics* 28. 337-354.
- Tang, Chih-Chen Jane. 2004. Two types of classifier languages: A typological study of classification markers in Paiwan noun phrases. *Language and Linguistics* 5(2). 377-407.
- Watanabe, Akira. 2006. Functional projections of nominals in Japanese: Syntax of classifiers. *Natural Languages and Linguistic Theory* 24. 241-306.
- Watters, John R. 1981. *A phonology and morphology of Ejagham –With notes on dialect variation*. Los Angeles, University of California PhD. dissertation.
- Zhang, Hong. 2007. Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16(1). 43-59.
- Zhang, Jingni. 2005. Maonanyu de liangci duanyu [The classifier phrase in Maonan]. In Jingfang Li & Suhua Hu (eds.), *Hanzangyu linagci Yanjiu [Studies of classifiers in Sino-Tibetan]*, 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://www.ccunix.ccu.edu.tw/~lgnz/index.files/May%202009.pdf>
- Zhang, Niina Ning. 2011. The constituency of classifier constructions in Mandarin Chinese. *Taiwan Journal of Linguistics* 9(1). 1-50.
- Zhang, Niina Ning. 2013. *Classifier structures in Mandarin Chinese*. Berlin: Mouton de Gruyter.