

# Unification of Numeral Classifiers and Plural Markers: Empirical Facts and Implications

**One-Soon Her**

Graduate Institute of Linguistics & Research  
Center for Mind, Brain, and Learning  
National Chengchi University  
Taipei, Taiwan  
hero@nccu.edu.tw

**Yun-Ru Chen**

Graduate Institute of Linguistics  
National Chengchi University  
Taipei, Taiwan  
99555015@nccu.edu.tw

## Abstract

The use of an obligatory numeral classifier (C) on N in general does not co-occur with mandatory plural marking (PM) (Greenberg 1990[1972], Sanches and Slobin 1973). Borer (2005) and Her (2012a) take this generalization further and see Cs and PMs as the same category. This unification implies that C/PM are mutually exclusive on N. In this paper, we first provide a mathematical foundation for this unification, i.e., C/PM both function as a multiplicand with the precise value of 1 (Her 2012a), and then explore empirically to what extent C/PM's complimentary distribution is borne out. We obtain from the WALS database a total of 22 languages with both Cs and PMs, including Mandarin, Japanese, and Vietnamese. Our survey finds C/PM co-occurring on N in 11 languages. We then set out to formally account for the unification of C/PM and explain its exceptions, taking Mandarin as an example, with a double-headed classifier construction. This study thus adds merit to the unification of C/PM and concludes with its implication on a universal lexical count/mass distinction.

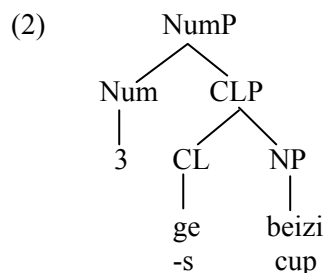
## 1 Introduction

Greenberg (1972) and Sanches and Slobin (1973) made the initial observation that languages with obligatory numeral classifiers (Cs) on nouns do not have compulsory morphological marking of nominal plurality, and vice versa. This generalization has been supported by a number of researchers, e.g., Tsuo (1976), Borer (2005), Her (2012a), Doetjies (2012), among others. To

explain this generalization, Greenberg (1972) links the emergence of Cs in a language to its loss of plural markers (PMs), and as Peyraube (1998) observes, this is true for the rise of Cs in Chinese.

However, this generalization is noncommittal on the complimentary distribution of Cs and PMs, as it says nothing about the cases where either C or PM is optional. Borer (2005:94) and Her (2012a:1682) take this generalization further and claim that Cs and PMs are the same category. The *-s* suffix in English, for example, applicable to all count nouns, is seen as a general classifier, similar to the Chinese *ge* in (1a) (Her 2012a:1682); the two thus share the same constituent structure, as in (2).

- (1) a. 三 個 杯子  
*san ge beizi*  
3 C cup  
b. three cups



The C/PM unification predicts the two to be in complimentary distribution on N. Yet, it does not preclude the scenario where the two coexist in a language but do not co-occur on N. The first objective of this paper is purely empirical: to identify to what extent these two predictions are borne out. We then set out to account for the

general pattern of distribution between Cs and PMs across languages as well as the exceptional cases where C/PM do co-occur on N. The paper is organized as follows. Section 2 offers a mathematical interpretation of Cs and PMs as the functional basis for their unification. Section 3 then obtains from the WALS database 22 languages that employ Cs and PMs and examines the distribution of the two on N in each language. Section 4 consists of discussions of the empirical facts obtained in the previous section and offers a formal syntactic account of Mandarin C/PM co-occurrence. Section 5 examines the implication that C/PM unification has on the controversy of count/mass distinction in languages. Section 6 concludes the paper.

## 2 Unification of Cs and PMs

In classifier languages, the C position in relation to a numeral (Num) and N can also be occupied by a measure word (M). Her and Hsieh (2010) and Her (2012b) demonstrate that, while C/M belong to the same category, they differ semantically. Specifically, while M is semantically substantive, C is semantically null, in the sense that it does not contribute any additional semantic information to N that N does not already have. Thus, if the grammatically required C is omitted for stylistic considerations, the meaning is not affected, as in (3), taken from Her (2012b:1219 (16)).

- (3) 五(張) 餅 二(條) 魚  
*wu zhang bing er tiao yu*  
 5 C loaf 2 C fish  
 ‘5 loaves and 2 fish.’

Based on the insight from Greenberg (1990[1972]:172), Au Yeung (2005), and Yi (2011), Her (2012a) proposes to account for C’s semantic redundancy mathematically in seeing [Num C] as  $[n \times I]$ . In a multiplicative operation, for a multiplicand to be null, its value can only be  $I$ . This view unifies all Cs under the concept of a multiplicand with the precise value of  $I$ .<sup>1</sup> To illustrate, (3) can be seen mathematically as (4).

- (4)  $[[5 (\times I)] \text{ loaf}] (+) [[2 (\times I)] \text{ fish}]$

<sup>1</sup> The only mathematical difference between Cs and Ms is that the value of an M is anything but  $I$ , e.g., 2 in the case of 雙 *shuang* ‘pair’, 12 in the case of 打 *da* ‘dozen’, and kilo in the case of 公斤 *gongjin* ‘kilo’ (Her 2012a).

Having established Cs as the multiplicand  $I$  entering a multiplicative relation with Num as the multiplier in [Num C N], we now compare the Chinese example in (1a), repeated in (5), with its English counterpart in (6): the only difference is that Chinese employs a C and English uses a PM  $-s$ , which can also be seen as the multiplicand  $I$ .

- (5) Chinese:  $[[3 \times I] \text{ cup}] = [3 \text{ } ge \text{ } beizi]$   
 (6) English:  $[[3 \times I] \text{ cup}] = [3 \text{ } \overleftarrow{-s} \text{ } cup]$

Surface word orders set aside, the two languages are identical in their nominal expressions with numerals, if C *ge* and PM  $-s$  are taken to be the two sides of the same coin. Indeed, like Chinese C, which is generally required, the generally required  $-s$  can be omitted without affecting the meaning. Thus, though (7a) is ill-formed in an argument position, its meaning is unmistakable. Also, in some languages, e.g., Hungarian, Tibetan, Archaic Chinese, among others, the counterpart of (7a) is well-formed in argument positions. Likewise in (7b), *three-cup*, well-formed as a modifier, still has the plural reading. And then, there are cases like those in (7c), where the omission of  $-s$  is obligatory, but a plural reading still must obtain.

- (7) a. \*three cup  
 b. a three-cup bra  
 c. three fish/deer/sheep

Note also that PM  $-s$  is still required when Num’s value is smaller than  $I$ , and thus not  $I$ , e.g., *0.5 apples* and *0 apples* and not *\*0.5 apple* and *\*0 apple*, indicating that  $-s$  here has little to do with plurality. The PM  $-s$  thus serves the same function as a general C in highlighting the discreteness or countability of N. However, there is a caveat: PM  $-s$  is not allowed when Num has the value of  $I$ , as in (8), and yet, the counterpart C *ge* in Chinese is well-formed.

- (8) He bought one cup(\*s).  
 (9) *Ta mai-le (yi) ge beizi.*  
 he bought 1 C cup  
 ‘He bought a cup.’

Her (2012a:1682) offers an explanation likewise based on mathematics. In  $[I \times I]$ , either the multiplier (Num) or the multiplicand (C/PM) can be omitted without changing the result. Both options are found in languages. As seen in (10a), in Chinese Num is *optionally* omitted, but only

when its value is *I*; in contrast, in Persian, when Num is *I*, it is *obligatorily* omitted, as in (10b) (Gebhardt 2009:212). In Khasi, an Austro-Asiatic language in India, when Num is *I*, it is C that is obligatorily omitted, as in (10c) (Temsen 2007: 6); the same is true in Tat (Greenberg, 1990[1972]:168), Amis, a Formosan language (Tang 2004:389), and Tetun, an Austronesian language (van Klinken, 1999). English, shown in (10d), is thus rather like Khasi; the only difference is that the multiplicand *I* is expressed as PM in English, C in Khasi. Incidentally, Indonesian is interesting in that the generally optional C is obligatory with the numeral *I* (Sneddon 1996).

- (10) Options of Num, Num=*I*
- a. Chinese  $[[I \times I] \text{ cup}] = [(1) \text{ C cup}]$
  - b. Persian  $[[I \times I] \text{ cup}] = [*1 \text{ C cup}]$
  - c. Khasi  $[[I \times I] \text{ cup}] = [1 *C \text{ cup}]$
  - d. English  $[[I \times I] \text{ cup}] = [1 *-s \text{ cup}]$

To summarize, Cs and PMs can be unified under the view that they both enter into a multiplicative relation with Num and function as a multiplicand with the precise value of *I*, which explains why both are semantically superfluous.

### 3 Potential Exceptions in 22 Languages

The unification of Cs and PMs as the same category means that they occupy the same syntactic position and share the same constituency structure. Consequently, Cs and PMs must be mutually exclusive on N. Yet, there has not been any serious attempt in finding out to what extent this prediction is borne out empirically. In the *World Atlas of Language Structures* (WALS) database, there are two studies that may shed light on this very issue, though indirectly: Gil (2008) looks at 400 languages and found Cs in 140, and Haspelmath (2008) examines 291 languages and 163 have PMs. What interests us is that 114 languages are covered in both studies, as shown in Table 1.

Table 1. C/PM Distribution in 114 languages

	PM×	PM√
C×	8	80
C√	4	<b>22</b>

Out of the 114, only 8 are without Cs and PMs, which will be examined in Section 5. The majority employs PMs only, while 4 employ Cs only. Cs and PMs thus do seem to be largely

complimentarily distributed in languages. However, 22 languages have both, as seen in Table 2.

Table 2. Cs and PMs in 22 languages

	Optional C	Obligatory C
Human Ns, optional	Hatam	Mandarin Japanese
All Ns, optional	Ainu Indonesian Khmer Tetun Chantyal	Garó Jacalteco Nivkh Teribe Ulithian Vietnamese
All Ns, optional in inanimates	None	Belhare
Human Ns, obligatory	None	Taba Kathmandu- Newar
All Ns obligatory	Hungarian Turkish Tuvaluan	Kham Mokilese

The 4 languages with obligatory Cs and PMs, if confirmed, are certain challenges, as C/PM co-occurring on N is certain. Yet, in fact all 22 languages may present problems for C/PM unification, if C and PM co-occur on N, whether optionally or obligatorily. In 3.1 are listed the 11 languages where Cs and PMs are found to be in complimentary distribution, and 3.2 presents the 11 languages that do allow Cs and PMs to co-occur on N.

#### 3.1 C/PM mutually exclusive

**Garó** (Tibeto-Burman): optional Cs and PMs, PMs not used where numerals denote plurality (Burling 1961, *p.c.*).

**Indonesian**: optional Cs and optional PM by way of reduplication (Sneddon 1996), does not allow C/PM co-occurrence (Johnny Lee, *p.c.*).

**Kham** (Tibetan): obligatory PM on all Ns, but the putative Cs are in fact ‘not true classifiers in the classical sense defined by Greenberg (1972) and others..’ (Watters, 2002:180).

**Jacalteco** (a Mayan language of Guatemala): obligatory Cs and an optional PM on all Ns. However, we suspect that the putative PM *heb*’ is an adjective or quantifier (see the discussions of Mokilese and Vietnamese below), not a morphological PM. See (11).

- (11) *ca-wan heb’ naj winnaj*  
 2-NumCL PL C man  
 ‘the 2 men’ (Craig, 1986:246)

**Mokilese** (Micronesian): plurality marked on the determiner, not N (Harrison, 1976), no C/PM co-occurrence on N (Doetjes 2012).

**Teribe** (a Chibchan language of Panama): obligatory Cs, optional PMs, do not co-occur (Quesada, 2000).

**Tetun** (an Austronesian language of Timor): optional Cs and an optional PM on all Ns (van Klinken *et al*, 2002), no examples of the two co-occurring (John Hajek, *p.c.*).

**Tuvaluan** (an Austronesian language of Tuvalu): some classifier-like elements, which led to its inclusion by Gil (2005). Yet, Besnier (2000:367) is emphatic that ‘Polynesian languages do not have classifier systems, and Tuvaluan is no exception.’

**Turkish**: optional Cs and PMs (Kornfilt, 1997; Göksel and Kerslake, 2011), no C/PM co-occurrence on N (Jaklin Kornfilt, *p.c.*).

**Ulithian** (Austronesian): obligatory Cs, but plurality is marked on demonstratives, not on Ns (Lynch *et al*, 2002).

**Vietnamese**: obligatory Cs, optional PMs on all Ns. Note, however, that the so-called ‘pluralizers’ or ‘plural markers’ are in fact quantifiers, not morphological PMs on N, and carry various explicit quantifier meanings and (in)definiteness (Thompson, 1965:180; Schachter, 1985:38). See (12).

- (12) *các con ngựa đen*  
 PL-def C horse black  
 ‘the black horses’ (Nguyen 2004:18)

### 3.2 C/PM not mutually exclusive

For each the 11 languages where Cs and PMs co-occur on N, a reference and an example are given.

**Ainu** (an indigenous language of Hokkaido): optional Cs and an optional PM for all Ns (Bugaeva, 2012)

- (13) *okkaypo utar tu-n*  
 young.man PL 2-C  
 ‘these 2 young men’ (Anna Bugaeva, *p.c.*)

**Belhare** (a Kiranti language of Nepal): obligatory Cs and optional PM on inanimate Ns.

- (14) *sip-paŋ maʔi-chi*  
 2-C person-PL  
 ‘2 people’ (Bickel 2003:563)

**Chantyal** (an endangered language of Nepal): optional Cs and an optional PM for all Ns.

- (15) *tin-ta jəmməy naku-ma*  
 3-C all dog-PL  
 ‘all 3 dogs’ (Noonan 2003:318)

**Hatam** (West Papuan): optional Cs and PMs.

- (16) *di-kindig-bat-nya i-bou can*  
 1sg-brother-COLL-PL 3PL-C 2  
 ‘my 2 brothers’ (Reesink 1999:83)

**Hungarian**: optional Cs, PMs obligatory on all Ns but do not co-occur with numerals.

- (17) *ex-ek a szem-ek rohad-t-ak*  
 this-PL the C-PL rotten-PL  
 ‘These rotten ones.’  
 (18) *három takaró-(\*k)*  
 3 blanket-PL  
 ‘3 blankets’ (Csirmaz and Dékány, 2010:13)

However, Csirmaz and Dékány (to appear) suggests that [Plural demonstrative + def. article + CL + N-PL] is not well-formed.

- (19) *??az-ok a fej salátá-k*  
 that-PL the C lettuce-PL  
 ‘those heads of lettuce’

**Japanese**: obligatory Cs and an optional PM on human Ns.

- (20) *Sono-gakusei-tati san-nin kita.*  
 that-student-PL 3-C came  
 ‘The 3 students came’ (Amazaki, 2005:224)

**Kathmandu Newar** (Tibeto-Burman): obligatory Cs and PMs on animate Ns.

- (21) *nya-mhə pasa-pī:*  
 5-C friend-PL  
 ‘5 friends.’ (Hale and Shrestha 2006:93)

**Khmer** (Austroasiatic and official language of Cambodia): optional Cs and an optional PM on all Ns (Gilbert, 2008; Gorgoniyev, 1966).

- (22) *proas (proas) bei nak*  
 man-man 3 C  
 ‘3 men’ (Soksan Ngoun, *p.c.*)

**Mandarin**: obligatory Cs and an optional PM for human Ns.

- (23) *san wei laoshi-men*  
 3 C teacher-PL  
 ‘3 teachers’ (Her 2012a)

**Nivk** (language isolate of Siberia): obligatory Cs and optional PMs on all Ns.

- (24) *ku-umguo vla-gu men*  
 that-girl-PL 2 C  
 ‘those 2 girls’ (Panfilov 1962:158)

**Taba** (Austronesian): obligatory Cs and an optional PM on human Ns.

- (25) *mapin-ci mat-tol*  
 woman-PL C-3  
 ‘3 women’ (Bowden 2001:256)

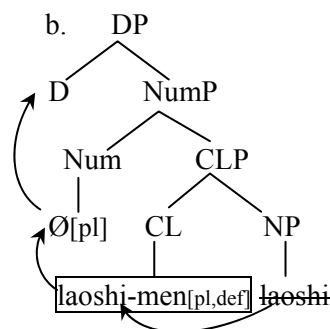
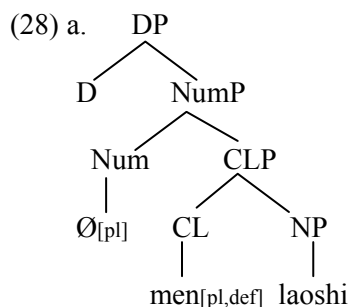
#### 4 A Formal Account for Mandarin

These 11 languages present a significant challenge to the unification of C/PM; however, a thorough and comprehensive examination of all 11 cases is clearly too wide in scope for the present paper. We thus focus on Mandarin, a quintessential classifier language believed to have the largest inventory of Cs (T’sou, 1976). Traditionally it has been claimed that C/PM do not co-occur on N in Mandarin. However, recent data from corpus and the Internet indicate that C and *-men*, a PM for human nouns, do co-occur, indicating variation in grammaticality judgment among Mandarin speakers (e.g., Her 2012a). To explain this C/PM co-occurrence, we propose a formal account with the following grammatical characterizations.

- (26) a. The category CL consists of the two subcategories: Cs and PMs.  
 b. The morpheme *-men* is a suffix that carries the feature [pl] and [def].  
 c. Cs are clitics and require a proper host (Yang 2002, Chen 2012).  
 d. Numerals project NumP and carry [pl], except *1*, which has [sg].  
 e. There are two null numerals,  $\emptyset$ [sg] and  $\emptyset$ [pl].  
 f.  $\emptyset$ [pl] subcategorizes for PM, all other numerals, C.

With that, we account for (27) with the structure and derivation in (28). The suffix *-men* attracts N to raise to CL. Given that *-men* carries a definite reading (e.g., Huang *et al*, 2009:8.4.1), the *N-men* phrase thus raises to Num and then to D to fill the empty heads.

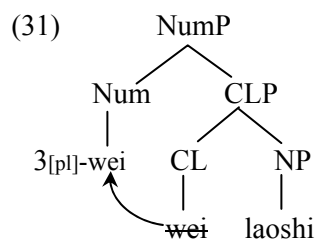
- (27) *laoshi-men*  
 teacher-PL



In contrast, numeral *san* ‘3’ in (29) is ill-formed, because overt numerals subcategorize for Cs, not *-men* (see (26f)). The example in (30) is thus well-formed; the clitic *wei* raises to an Num, as in (31). Following Huang *et al* (2009, chp.8), we assume (30) is ambiguous between a quantity reading with NumP, and an individual reading, thus with a null D projecting a DP and taking NumP as complement.

- (29) \**san laoshi-men*  
 3 teacher-PL

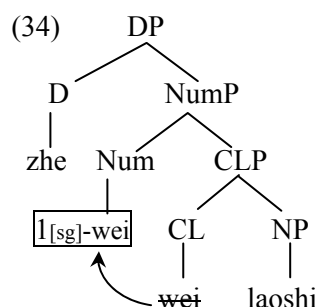
- (30) *san wei laoshi*  
 3 C teacher  
 ‘3 teachers’



The bare classifier phrase in (32) is ill-formed, for Cs, as clitics in Mandarin, require a proper host. The example in (33), where an overt numeral serves as the host for C, is thus well-formed with or without the overt D.

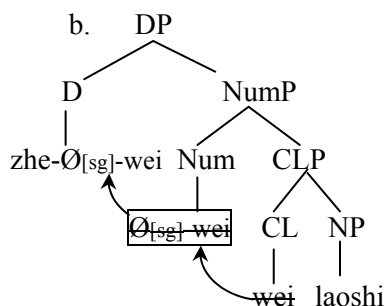
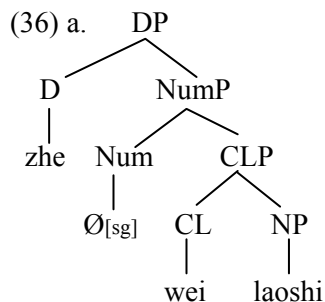
- (32) \**wei laoshi*  
 C teacher

- (33) (*zhe*) *yi wei laoshi*  
 the 1 C teacher  
 ‘(the) one teacher’



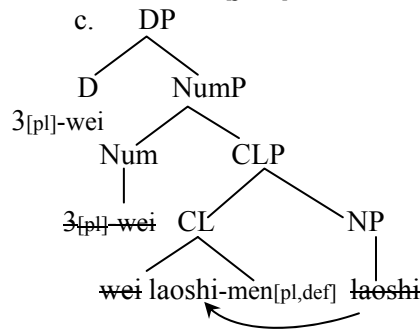
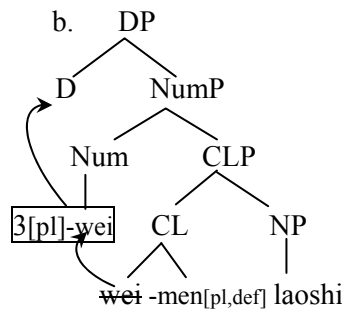
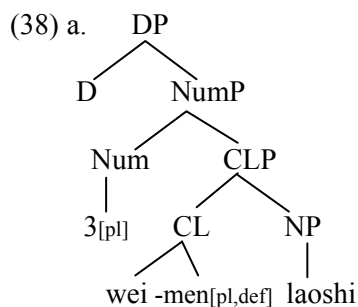
As mentioned earlier, the numeral *1* can be omitted. The example in (35) is thus well-formed with or without the numeral *1*, as long as there is an overt D serving as host for C.

(35) *zhe wei laoshi*  
 the C teacher  
 ‘this teacher’



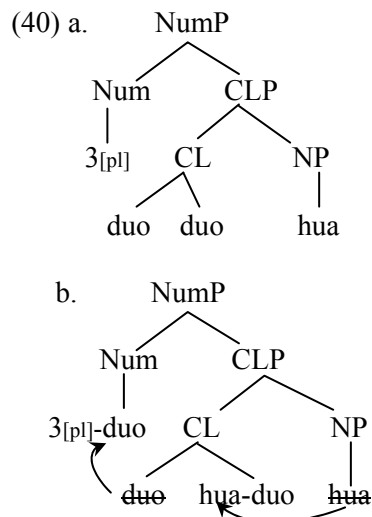
Finally, for the co-occurrence of C and *-men*, it should be noted that any analysis proposed should ideally reflect its marked nature, as C/PM co-occurring on N is clearly the exception, not the norm. An example is given in (37), the derivation of which is illustrated in (38a-c). Note the difference between (38) and (31); the latter is without *-men* and thus without DP.

(37) *san wei laoshi(-men)*  
 3 C teacher-PL  
 ‘(the) 3 teachers’



What’s marked about the structure is that CL is double-headed, with a C and a PM, each undergoing its normal derivation. In (38c), the Num-C phrase thus raises to D to fill its empty head for definiteness, the only compatible reading with the CLP due to *-men*. Thus, the DP structure of (38) remains the same with an overt D, e.g., *zhe* ‘the’. The double-headed CL is independently motivated by the so-called ‘CL copying construction’, coined by Zhang (2013:169). Two examples are given in (39), and the proposed derivation of (39a) in (40), which, like (30) and unlike (37) with *-men*, is a NumP.

(39) a. *san duo hua(-duo)*  
 3 C flower-C  
 ‘3 flowers’  
 b. *san pi ma(-pi)*  
 3 C horse-C  
 ‘3 horses’



There is also cross-linguistic evidence of double-headedness, e.g., appositional compounds (*woman doctors, women doctor*) (Bresnan, 2001), the double-headed VV structures in Classical Chinese (Feng 2002), and a double-headed verbal phrase structure for serial verb constructions in some African languages (e.g., Baker, 1989; Hiraiwa and Bodomo, 2008; Aboh, 2009).

## 5 Implication on the Count/Mass Issue

The discussions thus far indicate that Cs and PMs are two typological choices serving the same cognitive function of highlighting the discrete or count nature of a noun. In other words, the existence of C/PM in a language entails a count/mass distinction in that language. Thus, to the extent that C/PM is universal, so is the count/mass distinction. The C/PM unification thus supports Yi's (2009, 2011) and Her's (2012a) rejection of the thesis that classifier languages, unlike PM languages, have no count nouns, a thesis held prominently by Quine (1969b:35ff), Allan (1977), Krifka (1995), and Chierchia (1998), among others. This leaves us with two issues to explore further. First, is the count/mass distinction made universally at the lexical or syntactic level? Second, is there any evidence for the count/mass distinction in languages without C/PM?

### 5.1 Syntactic or lexical distinction

Borer (2005) contends that all nouns in all languages are mass at the lexical level and a count/mass distinction only exists at the syntactic level, i.e., a noun is count only when it appears as the complement in the syntactic configuration projected by C/PM. Her view is based primarily on data showing fairly robust convertibility between putative count nouns and mass nouns in English, as in (41) and (42).

- (41) A wine/wines, a love/loves, a salt/salts (on count reading)  
 (42) There is dog/stone/chicken on the floor (on mass reading)

Borer's view thus predicts that all putative mass nouns can be marked with a PM and be coerced into count, as in (41), and in the case of classifier languages, all putative mass nouns can appear with a C and thus turn into count. But this prediction is too strong to be true. Many putative

mass nouns in English cannot be quantified by numerals, unlike putative count nouns.

- (43) a.\*one conspicuousness  
 b.\*one beautifulness  
 c.\*one precariousness  
 (44) a.\*three conspicuousnesses  
 b.\*three beautifulnesses  
 c.\*three precariousnesses

Likewise, many putative mass nouns in Mandarin cannot appear in the [Num C N] configuration either, again unlike putative count nouns in the language.

- (45) a.\*三個 空氣  
*san ge qi*  
 3 C air  
 b.\*三個 酒精  
*san ge jiujiing*  
 3 C alcohol  
 c.\*三個 不銹鋼  
*san ge buxiugang*  
 3 C stainless-steel

The problem is easily solved, however, if the traditional view is adopted, where a count/mass distinction is made at the lexical level.

### 5.2 Languages without C/PM

Given the lexical count/mass distinction in languages with C/PM and the fact that the majority of the world's languages have either Cs or PMs or both (again, see Table 1, repeated below), the implication is that the count/mass distinction is universal.

Table 1. C/PM Distribution of in 114 languages

	PM×	PM√
C×	8	80
C√	4	22

Out of the 114 languages covered by both Gil (2008) and Haspelmath (2008), only 8 are without C/PM. Early Archaic Chinese is another example. Since grammatically, count nouns, by definition, can be counted without the help a measure word, a language must logically have numerals, count quantifiers, e.g., *several* and *many*, or count determiners, e.g., *these* and *those*, for it to have count nouns. So, we shall have a closer look at the numeral systems in these 8 languages and Early Archaic Chinese. The nine languages are divided into two groups, those

with restricted numerals only and those with a (semi-)productive system.

### Group 1: restricted numerals

**Imonda**, a Papuan language, has numerals 1 to 5 only: 1, 2, 1+2, 2+2, and 2+2+1 (Chan, 2013).

**Pirahã**, an Amazonian language isolate, has no numerals (Frank *et al*, 2008).

**Yidiny**, a nearly extinct Australian language, has only numerals 1–5 (1991:224).

**Yingkarta**, an Australian language *a.k.a.* Yinggarda and Inggarda, has numerals 1-4 only: 1, 2, 3, and 2+2 (Chan, 2013).

### Group 2: (semi-)productive systems

**Early Archaic Chinese** already has a very mature decimal system.

**Chimariko**, a Hokan language of California, now extinct, has quinary and decimal system (Jany 2007:110).

**Kombai**, a Papuan language, has a semi-productive body tally system (Chan, 2013).

**Mapudungun**, an Araucanian language of Chile, has a decimal system (Chan, 2013).

**Salt-Yui**, a Papuan language, has a finger-and-toe tally system with a 2, 5, and 20-based cyclic pattern (Chan, 2013).

We will take Early Archaic Chinese as an example for Group 2, and Pirahã, for Group 1. Early Archaic Chinese in the Shang oracle-bone inscriptions, or Oracular Chinese, from 18<sup>th</sup>-12<sup>th</sup> centuries BC, is known to have neither Cs nor PMs (Xu, 2006). It does, however, have a well-developed decimal numeral system and also a number of plural quantifiers. Evidence of count/mass distinction comes from the fact that numerals can quantify an N directly, as in (46). Without exception, such Ns are all putative count nouns, indicating a lexical count/mass distinction.

(46) a. 五人 一牛 (Hu 1983 (01060))

*wu ren yi niu*  
5 person 1 ox  
'5 persons and 1 ox'

b. 鳥二百十二, 兔一 (Hu 1983 (41802))

*niao er-ba-shi-er tu yi*  
bird 2-hundred-ten-2 hare 1  
'212 birds and 1 hare'

Pirahã, on the other hand, is anumeric and also makes no distinction between singular and plural (Everett, 2005). More significantly, experiments conducted by Gordan (2005) and Everett and

Madora (2012) show that monolingual Pirahã speakers are only able to conceptualize an exact numerical quantity equal to or smaller than three. However, the notion of 'count' only requires the concept of *individual* via the notion of *one* (Yi, 2009:219). In other words, the notion of exact quantity above three is not a necessary condition for either the conceptual or the linguistic distinction between count and mass. Clear evidence for a count/mass distinction in Pirahã comes from the two different quantifiers in (47), both indicating a large quantity in approximation (Nevins *et al* 2009).

(47) a. xaíbái 'many' (count nouns only)

b. xapagí 'much' (non-count nouns only)

The fact that a language without numerals is still able to make a lexical distinction of count/mass, coupled with the fact that pre-linguistic infants are capable of representing precise numbers (1-3) as well as approximating numerical magnitudes (see Feigenson *et al*, 2004, for an excellent summary and review), suggests that the count/mass distinction is universal.

## 6 Conclusion

This study confirms the generalization that numeral classifiers (Cs) and plural markers (PMs) are largely complementarily distributed in languages. We concur with Her (2012a) that this generalization exists because it reflects C/PM's identical mathematical function as a multiplicand with the value of 1 and the cognitive function of highlighting the discreteness, or the count nature, of the noun. However, genuine exceptions, where C/PM co-occur on N in 11 languages out of 114 examined, do pose a challenge to the unification of Cs and PMs. These 11 languages are Ainu, Belhare, Chantyal, Hatam, Hungarian, Japanese, Kathmandu Newar, Khmer, Mandarin, Nivk, and Taba. We take Mandarin as an example and account for its [D Num CL N] construction, where the C/PM co-occurrence involves a marked structure with a double-headed CL.

Furthermore, the unification of Cs and PMs also has significant implications on the debate over the count/mass distinction in languages. Our preliminary survey of 9 languages without C/PM, with special attention on Pirahã, indicates that the existence of a numeral system in a language is in fact not a prerequisite for the count/mass distinction. Thus, to the extent that the unification of Cs and PMs is on the right track,



the implication is that the count/mass distinction is made on the lexical level and it is universal.

## References

- Aboh, Enoch. 2009. Clause structure and verb series. *Linguistic Inquiry* 40(1):1-33.
- Allan, K. 1977. Classifier. *Language* 53: 285-311.
- Amazaki, Osamu. 2005. *A Functional Analysis of Numeral Quantifier Constructions in Japanese*. PhD dissertation, State University of New York at Buffalo.
- Au Yeung, W.-H. Ben, 2005. *An interface program for parameterization of classifiers in Chinese*. PhD Dissertation, Hong Kong University of Science and Technology.
- Baker, Mark. 1989. Object sharing and projection in serial verb constructions. *Linguistic Inquiry* 20(4):513-553.
- Besnier, Niko. 2000. *Tuvaluan: A Polynesian Languages of the Central Pacific*. Routledge, London and New York.
- Bickel, Balthasar. 2003. Belhare. In *The Sino-Tibetan languages*, eds., Graham Thurgood and Randy LaPolla, 546-569. Routledge, London.
- Bowden, John. 2001. *Taba: Description of a South Halmahera Language*. Pacific Linguistics, Canberra.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell, Oxford.
- Bugaeva, Anna. 2012. Southern Hokkaido Ainu. In *The Languages of Japan and Korea*, ed., Nicolas Tranter. Routledge.
- Burling, Robbins. 1961. *A Garo Grammar*. Deccan College Postgraduate and Research Institute Poona.
- Chan, Eugene. 2013. *Numeral systems of the world's languages*. Available online at <http://lingweb.eva.mpg.de/numeral/>. Accessed on 2012/03/01.
- Chen, Ching Perng. 2012. *On the Bare Classifier Phrase in Mandarin Chinese*. MA thesis, Graduate Institute of Linguistics, National Chengchi University
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339-405.
- Craig, Colette. 1986. Jacaltec noun classifiers: A study in grammaticalization. *Lingua* 70:241-284.
- Csirmaz, Aniko and Éva Dékány. To appear. Hungarian is a classifier language. In *Word Classes*, eds., Simone Raffaele and Francesca Masini. John Benjamins, Amsterdam.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press.
- Doetjes, Jenny. 2012. Count/mass distinction across languages. In *Semantics: An International Handbook of Natural Language Meaning*, Part III, eds., Claudia Maienborn, Klaus von Stechow and Paul Portner, 2559-2581. De Gruyter, Berlin.
- Dixon, Robert. 1991. *Words of Our Country*. University of Queensland Press.
- Everett, Caleb and Keren Madora. 2012. Quantity recognition among speakers of an anumeric language. *Cognitive Science* 36(1):130-141.
- Everett, Daniel 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46:621-646.
- Feigenson, Lisa, Stanislas Dehaene, and Elizabeth Spelke. 2004. Core systems of number. *Trends in Cognitive Science* 8(7):307-314.
- Feng, Shengli. 2002. A formal analysis of the origin of VR-constructions in Chinese. *Yuyanxue Luncong* 26:178-208. Commercial Press, Beijing.
- Frank, Michael C., Daniel L. Everett, Evelina Fedorenko, and Edward Gibson. 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, Volume 108(3):819-824.
- Gebhardt, Lewis. 2009. *Numeral Classifiers and the Structure of DP*. PhD Dissertation, Northwestern University.
- Gil, David. 2008. Numeral classifiers. In *The World Atlas of Language Structures Online*, eds., Martin Haspelmath, Mathew Dryer, David Gil, and Bernard Comrie, chapter 55. Max Planck Digital Library. Available online at <http://wals.info/feature/55>. Accessed on 2012/5/21.
- Göksel, Asli and Celia Kerslake. 2005. *Turkish: A Comprehensive Grammar*. Routledge, London and New York.
- Gordon, Peter. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306:496-499.
- Greenberg, Joseph. 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., Keith Denning and Suzanne Kemmer, 166-193. Stanford University Press. [First published 1972 in *Working Papers on Language Universals* 9:1-39.]
- Hale, Austin and Iswaranand Shresthachrya. 1973. Is Newari a Classifier Language? *Nepalese studies* 1(1):1-21.

- Harrison, Shelly. 1976. *Mokilese Reference Grammar*. University Press of Hawaii.
- Haspelmath, Martin. 2008. Occurrence of nominal plurality. In *The World Atlas of Language Structures Online*, eds., Martin Haspelmath, Mathew Dryer, David Gil, and Bernard Comrie, chapter 34. Max Planck Digital Library. Available online at <http://wals.info/feature/55>. Accessed on 2012/5/21.
- 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 and Chen-Tian Hsieh. 2010. On the semantic distinction between classifiers and measure words in Chinese. *Language and linguistics* 11(3):527-551.
- Hiraiwa, Ken and Adams Bodomo. 2008. Object-sharing as symmetric sharing. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, eds. Charles B. Chang and Hannah J. Haynie, 243-251. Cascadilla Proceedings Project, Somerville, MA.
- Hu, Houxuan (ed.). 1983. *Jiaguwen heji* 'The great collection of the oracle inscriptions'. China Social Sciences Publishing House, Beijing.
- Huang, C.-T. James, Audrey Y.-H. Li, and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge University Press.
- Jany, Carmen. 2007. *Chimariko in Areal and Typological Perspective*. PhD Dissertation, University of California, Santa Barbara.
- Krifka, Manfred. 1995. Common nouns: a contrastive analysis of Chinese and English. In *The Generic Book*, eds., Gregory N. Carlson and Francis Jeffry Pelletier, 398-411. University of Chicago Press.
- Nevins, Andrew, David Pesetsky, and Cilene Rodrigues. 2009. Piraha Exceptionality: a Reassessment. *Language* 85(2):355-404.
- Nguyen, Tuong-Hung. 2004. *The Structure of the Vietnamese Noun Phrase*. PhD dissertation, Boston University.
- Noonan, Michael. 2003. Recent Language Contact in the Nepal Himalaya. In *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*, eds., David Bradley, Randy LaPolla, Boyd Michailovsky, and Graham Thurgood, 35-51. Pacific Linguistics, Canberra.
- Panfilov, Vladimir. 1962. *Grammatika nivkhskogo iazyka* [Grammar of Nivkh] 1. Leningrad: Izdatel'stvo akademii nauk.
- Peyraube, Alain. 1998. On the history of classifiers in Archaic and Medieval Chinese. In *Studia Linguistica Serica*, ed. Benjamin K. T'sou, 131-145. Language Information Sciences Research Centre, City University of Hong Kong.
- Reesink, Ger P. 1999. *A grammar of Hatam: Bird's Head Peninsula, Irian Jaya*. Pacific linguistics, Canberra.
- Quine, Willard van Orman. 1969. *Ontological Relativity & Other Essays*. Columbia University Press, New York.
- Sanches, Mary and Linda Slobin. 1973. Numeral classifiers and plural marking: An implicational universal. *Working Papers in Language Universals* 11:1-22.
- Schachter, Paul. 1985. Parts-of-speech. In *Language Typology and Syntactic Description: Clause Structure Vol. 1*, ed., Timothy Shopen, 3-62. Cambridge University Press.
- Sneddon, James N. 1996. *Indonesian: A Comprehensive Grammar*. Routledge, New York.
- Tang, Chi-Chen J. 2004. Two types of classifier languages: A typological study of classification markers in Paiwan noun phrases. *Language and Linguistics* 5(2): 377-407
- Thompson, Laurence. 1965. *A Vietnamese Grammar*. University of Washington Press, Seattle, WA.
- T'sou, Benjamin K. 1976. The structure of nominal classifier systems. In *Austroasiatic Studies*, eds., Phillip N. Jenner, Stanley Starosta, and Laurence C. Thompson, 1215-1247. University of Hawaii Press.
- van Klinken, Catharina L. 1999. *A grammar of the Feban dialect of Tetun: An Austronesian language of West Timor*. Pacific Linguistics, Canberra.
- Watters, David E. 2002. *A Grammar of Kham*. Cambridge University Press
- Xu, Dan. 2006. *Typological Change in Chinese Syntax*. Oxford University Press.
- Yang, Rong. 2002. *Common nouns, Classifiers, and Quantification in Chinese*. PhD Dissertation, State University of New Jersey.
- 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, Niina Ning. 2013. *Classifier Structures in Mandarin Chinese*. Mouton de Gruyter, Berlin & New York.