LANGUAGE AND LINGUISTICS 11.3:527-551, 2010
2010-0-011-003-000291-1

On the Semantic Distinction between Classifiers and Measure Words in Chinese*

One-Soon Her and Chen-Tien Hsieh
National Chengchi University

Whether classifiers (C) and measure words (M) can be meaningfully distinguished in Chinese has been a controversial issue, reflected also by the drastic discrepancy in the inventories of classifiers previously proposed. The two tests, i.e. *de*-insertion and adjectival modification, that proponents for the C/M distinction proposed have been shown to be unreliable and thus rejected. We re-examine these two tests closely and propose two sets of refined, reliable, and revealing tests. We further employ the Aristotelian distinction between essential and accidental properties as well as the Kantian distinction between analytic and synthetic propositions to characterize the C/M distinction. M is therefore semantically substantive and thus blocks numeral quantification and adjectival modification to the noun; C, in contrast, does not form such a barrier, for it is semantically null in the sense that it merely highlights a semantic aspect inherent to the noun and thus contributes no additional meaning.

Key words: classifier, measure word, essential property, analytic proposition

1. Introduction

The classifiers this paper is concerned with are often referred to as 'numeral classifiers' because they are required to co-occur with numerals (e.g. Aikhenvald 2003:2). More specifically, it focuses on the so-called sortal classifiers. Such classifiers have lexical meanings in that classifier selection is based on certain intrinsic properties of the noun referents (e.g. Allan 1977, C. Hsieh 2009). Greenberg (1990[1975]:227) in fact claims that, in classifier languages, a numeral first forms a unit with a classifier, which in turn forms a unit with the noun. The fixed word order in a nominal phrase in Mandarin Chinese, a typical classifier language, is [(D)-NUM-CL-(N)], where a

* We are sincerely grateful to the anonymous reviewers of *Language and Linguistics* for their careful and insightful comments, which led to significant improvements of the paper. The contribution of the two co-authors is 80% and 20%, respectively, and the first author is also the corresponding author.

semantically compatible classifier or measure word must be employed before the noun when a numeral is present (e.g. M. Hsieh 2008, Tang 1990).

Previous studies of Mandarin classifiers, however, have come up with very different inventories, ranging from as many as six hundred (Hu 1993), four hundred and twenty-seven (Huang & Ahrens 2003), two hundred (Hung 1996), to as few as just several dozen (Erbaugh 1986); as noted by Liang (2006:17), this drastic discrepancy arises primarily from the dispute over what counts as a 'classifier'. It is uncontroversial that besides sortal classifiers, as in (1a), there are mensural classifiers, as in (1b) (e.g. Chao 1968, Allan 1977, Loke 1983, C. Hsieh 2009). The former subcategorizes objects with reference to their intrinsic properties, while the latter measures the quantity (e.g. Liang 2006, C. Hsieh 2009). However, they do seem to occupy the same slot and are mutually exclusive, as shown in (1c-d).[1]

(1) a. 一 本 書 (Sortal classifier)
yi *ben* *shu*[2]
one C book
'one book'

b. 一 箱 書 (Mensural classifier)
yi *xiang* *shu*
one box book
'one box of books'

c. *一 箱 本 書
yi *xiang* *ben* *shu*
one box C book

d. * 一 本 箱 書
yi *ben* *xiang* *shu*
one C box book

However, measure words are a mundane part of all natural languages, but sortal classifiers uniquely set apart the small number of classifier languages like Chinese and Japanese.[3] Tai (1994) thus points out the significance of this distinction from a typological point of view. We shall hereafter reserve the term *classifiers*, or C for short,

[1] However, we shall demonstrate in §2.3 that this is not the real picture and that they are not mutually exclusive and thus do not occupy the same slot.

[2] Hanyu Pinyin is used in the paper for transcription.

[3] Such classifier languages include most Southeast Asian languages, indigenous languages of western Americas, and Sub-Saharan African languages (Nichols 1992:200).

for sortal classifiers only, and refer to mensural classifiers as *massifiers* or *measure words*, or M in short.[4] We shall argue in the paper that the C/M distinction is real and it is crucial. The paper is organized as follows. Section 2 first justifies two sets of tests that serve to distinguish C and M. Section 3 then offers a precise semantic characterization for C and M. Section 4 concludes the paper with a summary.

Note that, unlike most previous studies on Mandarin classifiers, this study uses only data from a single dialect, Taiwan Mandarin (cf. Cheng 1985, Hsu 2006, Her 2009). All grammaticality judgments are thus based on native speakers from Taiwan and corpus data are from the Sinica Corpus and Google searches in the Taiwan domain.

2. Tests for the distinction between C and M

In this section, we shall first briefly review some of the formal syntactic analyses for C/M and then establish the count/mass distinction in relation to the C/M distinction, followed by a review of the two previous tests for the distinction of C as a distinct category from M. We shall offer three precisely formulated tests for the C/M distinction.

2.1 Formal syntactic analyses

In terms of the formal syntactic structure involving C and M, there are two opposing views in recent studies. Tang (2005) and M. Hsieh (2008) consider C and M to be syntactically alike and offer the same structure analysis. Thus, in (2a), Tang (2005) has both C and M as the head of a classifier phrase (ClP), while M. Hsieh (2008) likewise has either C or M as #, the head of a #P.

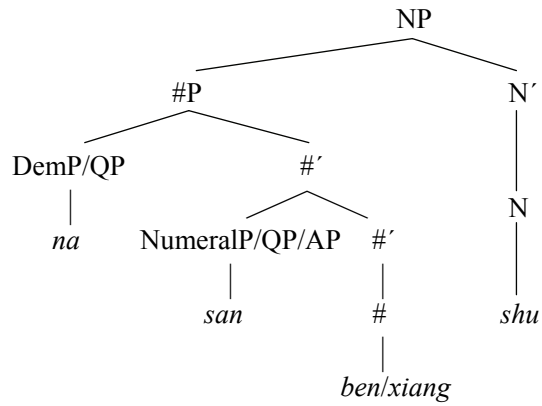
(2) a. $[_{DP}\ldots[_{NumP}\ldots[_{ClP}\ldots[_{NP}\ldots]]]]$

b.

三	箱/本	書
san	*xiang*/*ben*	*shu*
one	box/C	book

'three (boxes of) books'

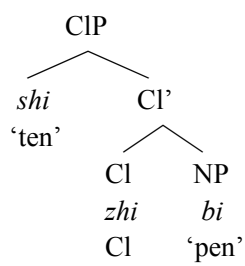
[4] Cheng & Sybesma (1998) is the first to use the term *massifier*, adopted also in M. Hsieh (2008).



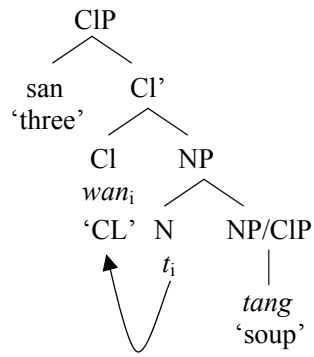
Such a unified analysis certainly explains why C and M occupy the same syntactic slot and are mutually exclusive. However, we shall demonstrate in §2.3 that this is not the real picture and that M in fact is structurally more prominent than C. Also, note crucially that even some of the proponents of the unified analysis, i.e. Tang (2005) and M. Hsieh (2008), acknowledge the count/mass distinction: C can only select count nouns, while M can occur with either count or mass nouns.

Though a unified analysis, even if justified, is not necessarily detrimental to our position that C and M are cognitively and semantically distinct, assigning C and M to different syntactic slots is certainly more conducive. This is precisely what Cheng & Sybesma (1998, 1999), Borer (2005), Watanabe (2006), and N. Zhang (2009) argue for. C, according to Cheng & Sybesma (1998, 1999), is base-generated as the head of ClP, as in (3a), while M, as shown in (3b), is based-generated under N and then moves upward to C. N. Zhang (2009), on the other hand, contends that C is base-generated as the head of SortP and can move up to Q, the head of QP and also where M is base-generated, as shown in (4).

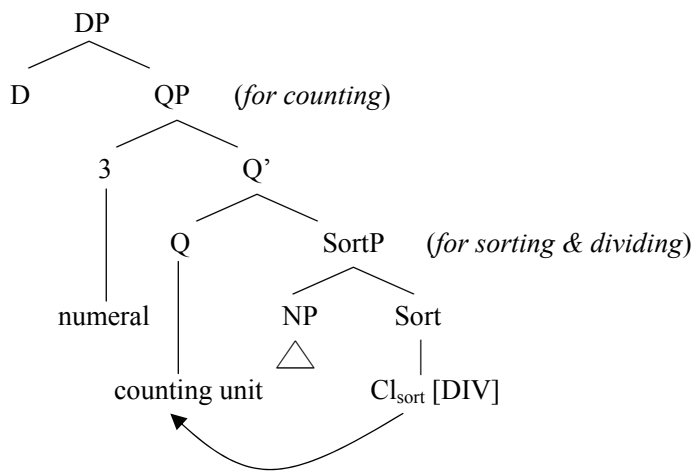
(3) a. 十 枝 筆
shi *zhi* *bi*
ten C pen
'ten pens'



b. 三 碗 湯
san *wan* *tang*
three bowl soup
'three bowls of soup'



(4)



In either account, C eventually ends up in the same position as M, and thus also explaining why C and M occupy the same syntactic slot and are mutually exclusive. Again, this may be incorrect, as we shall demonstrate in §2.3. We shall thus no longer be concerned with the debate over the formal syntactic analysis and move on to the count/mass distinction and syntactic tests for the C/M distinction.[5]

[5] For a critical review on the C/M distinction, see M. Hsieh (2008), and for a critical review of M. Hsieh (2008), see N. Zhang (2009).

2.2 Count/mass distinction

It has been widely observed that C is restricted to things perceived as inherently discrete, thus countable, as in (5), while measure words are not, as in (6) (e.g. Allan 1977, Loke 1983, Cheng & Sybesma 1998, 1999, W. Li 2000, Tang 2005, Liang 2006, M. Hsieh 2008, H. Zhang 2007, N. Zhang 2009, C. Hsieh 2009).

(5) 一 根 香蕉/*水
yi *gen* *xiangjiao*/*shui*
one C banana/water
'one banana/*water'

(6) 一 箱 香蕉/水
yi *xiang* *xiangjiao*/*shui*
one box banana/water
'one box of bananas/water'

In (5), the C *gen* denotes a discrete and elongated object, e.g. a *banana*, and thus rules out the mass noun *water*.[6] On the other hand, in (6), both the countable *banana* and the mass *water* can be contained in a box *xiang*, an M. Thus, the crucial difference is that the numeral *one* in effect quantifies the countable noun *banana* in (5), while it quantifies the countable M *box* in (6), not the noun *banana* or *water*. The count/mass distinction thus serves to distinguish C and M in that C, not M, fails to co-occur with mass nouns. Cheng & Sybesma (1999:515, 1998:403) are thus correct in stating that 'the count/mass distinction is clearly reflected in the classifier system'.

Tai (2003:312), however, contends that, in terms of the linguistic relativity hypothesis, there is no count/mass distinction in classifier languages, where all nouns can be treated as mass, and this is reflected in the fact that nouns in Chinese are not inflected for plurality,[7] cannot be counted without the accompaniment of classifiers, and can be either definite or indefinite when standing alone. He thus concludes that 'nouns in classifier languages denote materials or substances, non-discrete and unbounded, while in English and other European languages, they denote objects with discrete boundaries'. The weaker version of Tai's claim, i.e. that count nouns *can* be interpreted as mass nouns, is in fact not necessarily inconsistent with the conventional view, e.g. in Cheng & Sybesma (1998) and Tang (2005), which *can* be interpreted as: M does not

[6] Likewise, it should also rule out count nouns that do not have a compatible inherent feature.

[7] Y.-H. Li (1999) argues that the *-men* suffix in Chinese is indeed a plural morpheme; however, Iljic (1994) contends that it indicates collectivity. We thank an anonymous reviewer for the latter reference.

distinguish between count and mass nouns, while C requires count nouns. Take (6) for example: while the most natural reading of *yi xiang xiangjiao* is 'one box of bananas', where the boxful of bananas comes in the natural discrete units of banana, it is possible, though not probable, to have another reading where what is contained in the box is bits or pieces of banana or even mashed banana. In this second reading, *banana*, like *water*, is substance or material, and the measure word *xiang* serves to 'carve out', or 'parcel' (in Au Yeung's (2007) terminology) the volume of a *boxful* from the banana or water mass. Thus, conceptually, the count/mass distinction does not exist, but only for M under such a reading.

However, the stronger version would mean that nouns in Chinese *always* denote mass, never discrete units. This seems to be the position upheld in Chierchia (1998), who contends that Chinese is a *kind*-denoting language, where common nouns must denote mass properties and thus can only be counted with the aid of an imposing classifier. In *object*-denoting languages, e.g. English, common nouns can denote discrete units and can be quantified by a numeral without a classifier. Under this interpretation, the C/M distinction also does not exist as they now function exactly the same: both serve to 'carve out' discrete, bounded units from substance or material. This view cannot be sustained. Take (5) for example. Under this view, *xiangjiao* 'banana' can only refer to the banana mass, and the reading of a natural unit of banana with peel is only accidental and due to the classifier *gen*, which 'carves out' an elongated discrete unit. This view thus predicts that (5), besides this natural reading, can also mean an elongated unit of bits or pieces of the banana substance or mashed banana. Such a reading is simply impossible. Tai & Wang (1990:38) characterize the C/M distinction as follows:

> A classifier categorizes a class of nouns by picking out some salient perceptual properties, either physically or functionally based, which are permanently associated with entities named by the class of nouns; a measure word does not categorize but denotes the quantity of the entity named by noun.

According to this characterization, in Chinese there are natural classes of nouns that name entities which come in discrete countable units and such properties are indeed permanently associated with these nouns. The example in (7) should drive this point home.

(7) 一 個 新 汽車
yi *ge* *xin* *qiche*
one C new car
'one new car'

The generic classifier *ge*, unlike *gen*, only serves to highlight the intrinsic discreteness of the thing in question and does not indicate shape, function, animacy, or any other semantic feature. Cars, especially those newly produced by automobile manufacturers come in naturally countable units and a forced reading of one newly and randomly assembled unit out of the car mass is not available. Nouns in Chinese that similarly denote entities with 'inherent' and 'permanent' properties, to quote Tai (1994:3), of being discrete units are count nouns; otherwise, mass nouns.[8]

The count/mass distinction thus serves as an indication to the C/M distinction. Here is the rule of thumb. Given a noun denoting an intrinsically discrete object or, as we shall demonstrate later, something abstract but perceived to be intrinsically discrete (e.g. *jiaoyi* 'business transaction' and *guiding* 'regulation'), the single bare element required between a numeral and this noun is either C or M; if this element is only grammatically required in the counting of the natural unit of the noun and therefore contributes no additional semantic value to the phrase, it is a C; otherwise, it is an M. As seen in (7), the generic classifier *ge* contributes no semantic value to the compositional meaning of 'one' and 'car'.

2.3 Formal tests for the C/M distinction

There have been two well-known syntactic tests for the C/M distinction: adjective insertion and *de*-insertion. Given the on-going controversy, it is not surprising that both tests have been rejected by opponents to this distinction. However, we shall demonstrate that both tests can in fact be made much more accurate, and thus reliable. The two tests can ultimately be unified under the observation that M blocks numeral quantification and adjectival modification to the noun, while C does not. Finally, *ge*-substitution will be introduced as a heuristic for C/M distinction.

Test A: Adjectival Modification

Cheng & Sybesma (1998:390) claim that only M—and not C—can be modified by some bare adjectives, as shown in (8a) and (9) respectively. However, note that the bare adjectives allowed for M are in fact also strictly restricted to size (e.g. *da* 'big' and *xiao* 'small') and fullness or wholeness (e.g. *zheng* 'whole', as indicated by (8b)). We note further that even this highly restricted set of bare adjectives does not necessarily apply to all members of M, as shown in (8c). In other words, Cheng & Sybesma (1998) could

[8] Our position is thus more compatible with Borer's (2005) universalist view that all nouns in *all* languages are mass. In languages like Chinese, the use of classifiers turns a noun to a count noun; whereas in non-classifier languages like English, plural inflections are used for the same purpose.

thus only have meant for this to be a sufficient, but not necessary, condition for M.

(8) a. 一 小/大 箱 書
yi ***xiao/da*** ***xiang*** *shu*
one small/big box book
'one small/big box of books'

b. *一 紅/美 箱 書
yi ***hong/mei*** ***xiang*** *shu*
one red/pretty box book

c. *一 小/大 磅 書
yi ***xiao/da*** ***bang*** *shu*
one small/big pound book

(9) a. *一 小/大 隻 狗
yi ***xiao/da*** ***zhi*** *gou*
one small/big C dog

b. *一 紅/美 隻 狗
yi **hong/mei** ***zhi*** *gou*
one red/pretty C dog

However, though confirmed by some linguists, e.g. Liang (2006) and Tsai (2003), some of the native speakers we consulted do find (9a) acceptable and counter-examples to this test are acknowledged even by Cheng & Sybesma (1998) themselves, albeit in a footnote, and have been subsequently corroborated by many others, e.g. Tang (2005), M. Hsieh (2008), and N. Zhang (2009). Tang (2005) and M. Hsieh (2008) in fact utilize Taiwan Mandarin examples from the Sinica Corpus. Our own Google searches in the Taiwan domain have also come up with numerous [Adj-C] examples. For instance, there are 20 and 161 exact matches of (10a) and (10b) respectively.[9]

(10) a. 一 大 顆 蘋果
yi ***da*** ***ke*** *pingguo*
one big C apple
'one big apple'

b. 一 大 本 書
yi ***da*** ***ben*** *shu*
one big C book
'one big book'

[9] Data accessed on November 12, 2009.

This clear evidence is enough for even the proponents of C/M distinction, e.g. N. Zhang (2009), to write off this test as unreliable. However, we note a crucial difference here between C and M. As pointed out earlier, the pre-C numeral in effect quantifies the noun together with the C, while a pre-M numeral only quantifies the M itself, not the noun. Thus, concurrent numeral quantification at M as well as at the noun via a C, as in (11a), or stacking of M, as in (11b), is perfectly fine. However, comparable cases for C are entirely nonsensical, as in (12a-b). The concurrent M and C in (11a) also clearly indicate that C and M are not mutually exclusive as previously assumed, and M is in fact structurally more prominent than C.[10]

(11) a. 一 箱 十 顆 蘋果
yi ***xiang*** ***shi*** ***ke*** *pingguo*
one box ten C apple
'one box of ten apples'

b. 一 箱 十 包 蘋果
yi ***xiang*** ***shi*** ***bao*** *pingguo*
one box ten pack apple
'one box of ten packs of apples'

(12) a. *一 個 十 顆 蘋果
yi ***ge*** ***shi*** ***ke*** *pingguo*
one C ten C apple

b. *一 個 十 包 蘋果
yi ***ge*** ***shi*** ***bao*** *pingguo*
one C ten pack apple

Like the limited scope of quantification, a pre-M adjectival modification has only M as its scope, while a pre-C adjective transcends the C and also modifies the noun.

[10] The examples in (11a-b) are noun phrases, rather than a clause (i.e. topic-comment structure), because both can appear as the subject or object in a sentence, as shown below. We thank the reviewer who raised the question.

(i) 他 給了 我 一 箱 十 顆 蘋果
Ta *gei-le* *wo* *yi* *xiang* *shi* *ke* *pingguo*
he give-ASP I one box ten C apple
'He gave me one box of ten apples.'

(ii) 這 一 箱 十 顆 蘋果 夠 嗎？
Zhe *yi* *xiang* *shi* *ke* *pingguo* *gou* *ma*
this one box ten C apple enough Q
'Is this one box of ten apples enough?'

This fact is clearly illustrated in the English translations of (8) and (10), where the pre-C adjective in essence modifies the noun. Given the transcending modification of pre-C adjective, we can derive two precise predictions. First, adjectival modification on C is equivalent to that on N, as shown in (13b); M does not have this effect, as shown in (13a).

(13) a. 一 **大** 箱 蘋果 ≠ 一 箱 **大** 蘋果
yi ***da*** *xiang* *pingguo* *yi* *xiang* ***da*** *pingguo*
one big box apple one box big apple
'one big box of apples' 'one box of big apples'

b. 一 **大** 顆 蘋果 = 一 顆 **大** 蘋果
yi ***da*** *ke* *pingguo* *yi* *ke* ***da*** *pingguo*
one big C apple one C big apple
'one big apple' 'one big apple'

The second prediction is that stacking of antonymous adjectives is impossible for C, as shown in (15), as the apple cannot be big and small at the same time.[11] Yet, it is perfectly fine for M, as in (14), where the box is big while the apples are red/small.

(14) a. 一 **大** 箱 **紅/小** 蘋果
yi ***da*** *xiang* ***hong/xiao*** *pingguo*
one big box red/small apple
'one big box of red/small apples'

b. **大大的** 一 箱 **紅/小** 蘋果
dadade *yi* *xiang* ***hong/xiao*** *pingguo*
big one box red/small apple
'one big box of red/small apples'

(15) a. 一 **大** 顆 **紅/#小** 蘋果
yi *da* *C* ***hong/xiao*** *pingguo*
one big C red/small apple
'one big red/*small apple'

b. **大大的** 一 顆 **紅/#小** 蘋果
dadade *yi* *ke* ***hong/xiao*** *pingguo*
big one C red/small apple
'one big red/*small apple'

[11] Thus, the only coercible reading is that *xiao pingguo* is the name of a particular variety of apples and therefore the meaning of 'small' is opaque.

To conclude, while M blocks numeral quantification and adjectival modification to the noun, C does not. Thus, Adj-C is more restricted than Adj-M. Furthermore, given a well-formed Adj-C, the same adjective can be shifted to modify the noun without altering the meaning. Also, both numeral stacking and antonymous adjective stacking prove to be prudent tests. Test A is thus revised in much more accurate terms as follows.

Test A (revised): Numeral/Adjectival Stacking

Test 1: If [Num **X** Num **Y** N] is well-formed, then X = M, X ≠ C, and Y = C/M.

Test 2: If [Num A **X** N] = [Num **X** A N] semantically, then X = C and X ≠ M.

Test 3: Given antonyms A_1 and A_2, if [Num A_1 **X** A_2 N] is semantically incongruent, then X = C and X ≠ M; otherwise, X = M and X ≠ C.

Test B: *de*-insertion

Many linguists claim that *de* may be optionally inserted after M but not C, as shown in (16) (e.g. Chao 1968:555, Paris 1981:32, Zhu 1982:51, Tai & Wang 1990, Tai 1994, Cheng & Sybesma 1998:388, 1999:515, H. Zhang 2007:49).

(16) 一 箱/*本 的 書
yi *xiang*/*ben* *de* *shu*
one box/C DE book
'one box of/*C books'

Again, the real picture is far from being so clear-cut. Both opponents, e.g. Tang (2005) and M. Hsieh (2008), and proponents, e.g. N. Zhang (2009), to the C/M distinction have come up with plenty of well-formed C-*de*-N examples. M. Hsieh (2008) again cites examples from the Sinica Corpus, as shown in (17) and (18). Note that *de* does not change the meaning at all for either C or M.

(17) 五百萬 隻 的 鴨子
wubaiwan ***zhi*** ***de*** *yazi*
five-million C DE duck
'five million ducks'

(18) 幾百 條 的 海蛇
jibai ***tiao*** ***de*** *haishe*
several-hundred C DE sea-snake
'hundreds of sea snakes'

Again, the evidence seems rather obvious. Nonetheless, the intuition behind the observation that C-*de*-N is much more restricted than M-*de*-N is also reasonable: C and N merge as one unit semantically, while M and N form two separate entities; the C-N sequence is thus more resistant to *de*-insertion. Possessive *de* before a human noun reveals the same insight, as shown in (19) and (20). Note again the meanings remain the same with or without *de*.

(19) 我 (?的) 爸爸
wo *de* *baba*
I DE papa
'my papa'

(20) 我 *(的) 理髮師
wo *de* *lifashi*
I DE barber
'my barber'

The intimate kinship between me and my father is indicative to the resistance of *de* intervention; the much more distant relationship between me and my barber, on the contrary, much favors the separation of the two nouns by *de*. Thus, given this iconic value of *de*-insertion (cf. Chappell & Thompson 1992),[12] all is not lost for its use in C/M distinction. And indeed, as noted by Tang (2005), further corroborated by N. Zhang (2009), that, in a [Number-C-*de*-N] phrase, the lower the number, the less acceptable is the phrase. Thus, the higher the number, the more naturally *de* intervenes between C and N. This observation is certainly confirmed by M. Hsieh's two Sinica Corpus examples in (17) and (18), with *five million* and *several hundred* respectively. N. Zhang (2009) cites Croft (1994:163), Aikhenvald (2003:100), and Sato (2009:7) in noting that different properties between constructions with low and high numerals are observed in other languages as well and goes on to claim specifically that if the number is less than ten, then post-C *de* is ill-formed. This is confirmed by the examples cited in relevant literature and by our half dozen Taiwan Mandarin informants. However, our experiments also indicate that fractions of a number, including those with a value smaller than one, drastically increase acceptability. Google searches turned up 70 instances of 之一顆的 *zhi yi ke de* 'one fraction of', two of which are listed in (21).[13]

[12] Chappell & Thompson (1992) propose two principles governing the optional associative *de*.
(1) The closer the relationship between NP1 and NP2, the less likely *de* is to be used.
(2) The closer the relationship between NP1 and the *speaker*, the less likely *de* is to be used.
We thank the anonymous reviewer that provided this reference.

[13] Data accessed on November 12, 2009.

(21) a. 八分之一 顆 的 高麗菜
bafenzhiyi *ke* *de* *gaolicai*
one-eighth C DE cabbage
'one-eighth cabbage'

b. 四分之一 顆 的 洋蔥
sifenzhiyi *ke* *de* *yangcong*
one-eighth C DE onion
'one-eighth onion'

An explanation is attempted in Tang (2005:444), where numeral contrast is interpreted as a contrast in 'information weight', thus the higher the number, the higher its information weight. However, our data in (21) indicates that it is not the *absolute value* of the number that affects the 'weight'; rather it is the *computational complexity* that makes it 'heavy'.[14] Compare (22a) with (22b): *ban* 'half' and *yi* 'one' are similarly discrete and monosyllabic, but the latter is 'heavier' and thus better, with 20 exact Google matches, because it is computationally more complex than the former, with merely one single token found.[15]

(22) a. 半 顆 的 蘋果
ban *ke* *de* *pingguo*
half C DE apple
'half an apple'

b. *一 顆 的 蘋果
yi *ke* *de* *pingguo*
one C DE apple
'an apple'

The insight of information weight measured by computational complexity thus also predicts that any increased complexity of C itself should likewise increase the acceptability of *de*-insertion. The prediction is correct, as Adj-C-*de* with even the simplest number, one, is much better than bare C-*de*. Again, both examples in (23) are from Google searches, which gave 13 and 9 exact matches for each example respectively.[16]

[14] The term 'heavy' here is deliberate and relates nicely to Tang's (2005) use of 'information weight'. Also, the phenomenon under discussion here is surely reminiscent of the 'heavy NP shift' in English, where an ill-formed word order is allowed for a 'heavy' NP.
i. *I gave to Mary them/the flowers.
ii. I gave to Mary the flowers that I personally picked from the garden of my country cottage.

[15] Data accessed on February 22, 2010.

[16] Data accessed on November 12, 2009.

(23) a. 一 大 顆 的 高麗菜
yi *da* *ke* *de* *gaolicai*
one big C DE cabbage
'one big cabbage'

b. 一 大 條 的 魚
yi *da* *tiao* *de* *yu*
one big C DE fish
'one big fish'

Assuming further that Greenberg (1990[1975]:227), along with Paris (1981:105-117),[17] C.-T. Huang (1982), Lin (1997), and M. Hsieh (2005, 2008) (see the tree in (2b)), is correct that the numeral and C first form a constituent, say ClP, before merging with the noun, we can state with confidence that any increased computational complexity in ClP increases the acceptability of *de* intervention. But, why should it be any different for M, assuming that M likewise forms a constituent, say MP, first with the numeral? Recall our generalization earlier that M blocks numeral quantification and adjectival modification to the noun, but C does not. ClP is thus inherently more closely tied to N than MP and thus naturally more resistant to intervention by *de*. This insight therefore nicely unites Test A and Test B. Assuming that *one* is computationally the least complex number,[18] we can now restate Test B in much more restricted terms and much more accurately as follows.

Test B (revised): *De*-insertion
Test: [*yi* **M/*C** *de* N]

Test C: *Ge*-substitution

Tai & Wang (1990) and Tai (1994) propose that, for C/M distinction, if the element in question can be substituted by 個 *ge*, the generic classifier, without any loss in meaning, then it is C; otherwise, it is M.

(24) 三 **顆** 蘋果 = 三 **個** 蘋果
san *ke* *pingguo* *san* *ge* *pingguo*
three C apple three C apple
'three apples' 'three apples'

[17] We thank the anonymous reviewer that provided this reference.
[18] It certainly is not zero, the concept of which was developed much later in the number systems in all human languages.

(25) 三 箱 蘋果 ≠ 三 個 蘋果
san *xiang* *pingguo* *san* *ge* *pingguo*
three box apple three C apple
'three boxes of apples' 'three apples'

Note that *ge*-substitution does not have the same status as the previous two tests. Logically, before it can be used as a test for C, the C/M distinction needs to be made valid first, and then *ge* needs to be independently proven to be C. Both premises should be uncontroversial at this point of discussion. The reader is welcome to run the previous two C/M distinction tests on *ge*. However, it has its limitations, as many nouns require specific classifiers and do not readily take *ge*, except perhaps in casual or even sloppy speech (e.g. Erbaugh 1986, Hu 1993). *Ge*-substitution thus may serve only as a heuristic for C and is formulated as follows.

Test C: *Ge*-substitution
Test: If [Num **X** N] = [Num ***ge*** N] semantically,
then X = C and X ≠ M.

3. Semantic characterization of the C/M distinction

Having established three reliable tests for the C/M distinction, we shall now attempt to crystallize the semantic distinction between C/M. C classifies or categorizes nouns by highlighting some salient or inherent properties of the noun and thus contributes no additional meaning. M, on the other hand, plays a substantive role in denoting the quantity of the entity named by the noun. In §3.2, we demonstrate that this distinction also implies that C is a closed set and M an open set.

3.1 Aristotle and Kant

In *Metaphysics* Aristotle distinguishes between the essential features of a thing and its merely accidental features (Bostock 1994). Robertson (2008) offers this precise definition.

> *P* is an *essential property* of an object *o* just in case it is necessary that *o* has *P* whereas *P* is an *accidental property* of an object *o* just in case *o* has *P* but it is possible that *o* lacks *P*.

This characterizes the C/M distinction perfectly. A classifier picks out an essential property of the entity the noun denotes; in other words, it does not impart any information to the noun that it does not already have. For example, having a tail is part of what necessarily makes a fish. Thus, in (26), the classifier 尾 *wei* and 魚 *yu* 'fish' are compatible in that the latter has a tail as an essential property. The classifier clearly adds no information to the phrase and merely identifies this essential property, tail.

(26) 一 尾 魚
yi *wei* *yu*
one C fish
'one fish'

A measure word does just the opposite: it provides an additional property to the noun, a property that is accidental and thus not a necessary part of the entity denoted by the noun. In (27), the measure word 桶 *tong* 'bucket' thus furnishes additional information to the phrase, indicating that the fish are inside the bucket and mass the bucketful quantity, both accidental properties.

(27) 一 桶 魚
yi *tong* *yu*
one bucket fish
'one bucket of fish'

Along this lineage of philosophizing, we further recruit the analytic/synthetic distinction, which Kant was the first to introduce in the Introduction to the *Critique of Pure Reason*, to illustrate the C/M distinction. Here is Kant's definition, cited in Rey (2003), from which the examples are also obtained.

Analytic proposition: a proposition whose predicate concept is contained in its subject concept; e.g. *all bachelors are unmarried*.
Synthetic proposition: a proposition whose predicate concept is not contained in its subject concept; e.g. *all bachelors are happy*.

According to this distinction, what C contributes to the noun can only be paraphrased into an analytic proposition with the noun as subject and C itself as the predicate concept, while what M contributes can only be restated as a synthetic proposition. To illustrate, the C and M in (26) and (27) are now restated as (28) and (29), respectively.

(28) The fish has a tail.
(29) The fish are in the bucket and fill the bucket.

Having a tail is an essential property of fish and this predicate concept is thus contained in the subject concept in (28), which is thus an analytic proposition. Being in a bucket or filling the bucket is an accidental property of the fish and is thus not contained in the subject concept of (29), which is therefore a synthetic proposition. The above two distinctions thus illuminate Adams & Conklin's (1973:2) insight that classifiers 'qualify rather than quantify the head noun' and also confirm W. Li's (2000:1117) insight that 'classifiers are semantically redundant'. Thus, C serves to classify or categorize nouns by highlighting certain properties inherent to the noun. C is therefore semantically null, or redundant, in the sense that it contributes no additional semantic value that the noun does not already have. M, on the other hand, serves as an integral part, together with the numeral, in the quantification of the noun. Consider the data in (30) and (31).

(30) 五 個 餅 二 條 魚 = 五 餅 二 魚
wu ge bing er tiao yu *wu bing er yu*
five C loaf two C fish five loaf two fish
'five loaves and two fish'

(31) 五 餅 二 魚 ≠ 五 箱 餅 二 箱 魚
wu bing er yu *wu xiang bing er xiang yu*
five loaf two fish five box loaf two box fish
'five loaves and two fish' 'five boxes of loaves and two boxes of fish'

The example refers to the familiar story of Jesus feeding five thousand people with five loaves and two fish, and is commonly referred to in Chinese simply as the story of 五餅二魚 *wu bing er yu* '*five loaves and two fish*'. The fact that C, not M, can be dropped due to stylistic pressure attests to their semantic distinction: M contributes additional meaning to the noun; C does not. This semantic characterization thus also explains the C/M contrast in blocking numeral quantification and adjectival modification to the noun.

Having formulated three sets of reliable tests for the C/M distinction in §2, we can now give it a more precise semantic description. 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.

3.2 C a closed set versus M an open set

We have thus far demonstrated that C and M are similar in that syntactically they both serve as the required link between the numeral and the noun; however, semantically they are drastically different in that, while M contributes additional and unique value and thus serves as an integral part in the total semantic composition of the phrase, C does not. C is thus more of a functional device, while M is semantically substantive. C thus should constitute a closed set, like other grammatical devices, e.g. case markers, prepositions, etc., while M should be more of an open set.

Thus, C is resistant to innovations, while M is quite the opposite. As Adams & Conklin (1973) point out, 'essentially anything can function as the unit of measurement'. This indicates that common nouns can easily function as measure words, e.g. 一屋子人 *yi wuzi ren* 'a houseful of people' and 一卡車垃圾 *yi kache lese* 'a truckload of trash'. Many are created with body parts, e. g., 一肚子壞主意 *yi duzi huai zhuyi* 'a stomachful of malicious intentions', 一臉不悅 *yi lian buyue* 'a faceful of displeasure', and 一頭白髮 *yi tou bai fa* 'a headful of gray hair'.[19] All these innovative or temporary uses of common nouns as measure words indicates that M is an open set. C, on the other hand, allows no such innovations. Thus, it is plausible to attempt a comprehensive inventory of C only, while such an effort for M is not meaningful.

4. Conclusion

It has been an ongoing debate as to whether classifiers can be meaningfully distinguished from measure words in Chinese. The two tests, i.e. *de*-insertion and adjectival modification, proposed previously by proponents of the C/M distinction have since been discredited and rejected by opponents. In this paper, we have first established

[19] An anonymous reviewer questioned if Test 3 of Test A (i.e. occurrence of antonyms) is used to distinguish M from C, then the fact that *一黑頭白髮 *yi hei tou bai fa* '*a black headful of gray hair' is bad would mean 頭 *tou* here is C, not M as we claim. We thank the reviewer for this keen observation and note that, as pointed out earlier in the discussion of Test A: Adjectival Modification, the bare adjectives allowed for M are also strictly restricted to size, e.g. *da* 'big' and *xiao* 'small', and fullness or wholeness, e.g. *zheng* 'whole'. Thus, the ill-formedness of *一黑頭白髮 *yi hei tou bai fa* '*a black headful of gray hair' is due to the fact that 黑 *hei* 'black' is banned as a modifier to M (or C). In addition, the semantic incongruence is due to the fact that the actual color of the head is invisible as the head is covered by (gray) hair. So, *一白頭白髮 *yi bai tou bai fa* '*a gray headful of gray hair' is likewise bad. We further note that 一整頭白髮 *yi zheng tou bai fa* 'a complete headful of gray hair' is much better and does have 4 tokens in Google Taiwan.

the relationship between the count/mass distinction in nouns and the C/M distinction and then re-examined the precious two tests more closely. Based on the insight that M is semantically substantive and C is semantically null and thus M—not C—blocks numeral quantification and adjectival modification to the noun, we have refined the previous two tests and come up with much more reliable and accurate formulations. Likewise, we have stated *ge*-substitution as a heuristic.

Test A (revised): Numeral/Adjectival Stacking
Test 1: If [Num **X** Num **Y** N] is well-formed, then X = M, X ≠ C, and Y = C/M.
Test 2: If [Num A **X** N] = [Num **X** A N] semantically, then X = C and X ≠ M.
Test 3: Given antonyms A_1 and A_2, if [Num A_1 **X** A_2 N] is semantically incongruent, then X = C and X ≠ M; otherwise, X = M and X ≠ C.

Test B (revised): *De*-insertion
Test: [*yi* **M/*C** *de* N]

Test C: *Ge*-substitution
Test: If [Num **X** N] = [Num ***ge*** N] semantically, then X = C and X ≠ M.

In terms of semantics, we employ the Aristotelian distinction between essential and accidental properties as well as the Kantian distinction between analytic and synthetic propositions to characterize the C/M distinction. Precisely, C indicates an essential property of the noun, and can be paraphrased as the predicate concept in an analytic proposition with the noun as the subject concept; M indicates an accidental property in terms of quantity, and can be restated as the predicate concept in a synthetic proposition with the noun as the subject concept. Finally, given the above characterization, M can be demonstrated to be more of a content word, thus open to innovations, while C is more a function word, thus forms a closed set and is resistant to innovations.

References

Adams, Karen L., and Nancy F. Conklin. 1973. Toward a theory of natural classification. *Chicago Linguistic Society* (*CLS*) 9:1-10.

Aikhenvald, Alexandra Y. 2003. *Classifier: A Typology of Noun Categorization Devices.* Oxford & New York: Oxford University Press.

Allan, Keith. 1977. Classifier. *Language* 53.2:285-311.

Au Yeung, Ben Wai Hoo. 2007. Multiplication basis of emergence of classifiers. *Language and Linguistics* 8.4:835-861.

Borer, Hagit. 2005. *Structuring Sense*, Vol. 1: *In Name Only*. Oxford: Oxford University Press.

Bostock, David. 1994. *Aristotle Metaphysics.* Oxford: Clarendon Press.

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chappell, Hilary, and Sandra A. Thompson. 1992. The semantics and pragmatics of associative *de*. *Cahiers de Linguistique – Asie Orientale* 21.2:199-229.

Cheng, Lisa L.-S., and 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., and Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30.4:509-542.

Cheng, Robert L. 1985. A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language* 61.2:352-377.

Chierchia, Gennaro. 1998. Plurality of mass nouns and the notion of "semantic parameter". *Events and Grammar*, ed. by Susan Rothstein, 53-103. Dordrecht & Boston: Kluwer.

Croft, William. 1994. Semantic universals in classifier systems. *Word* 45:145-171.

Erbaugh, Mary S. 1986. Taking stock: the development of Chinese noun classifiers historically and in young children. *Noun Classes and Categorization: Proceedings of a Symposium on Categorization and Noun Classification, Eugene, Oregon, October 1983*, ed. by Colette G. Craig, 399-436. Amsterdam & Philadelphia: John Benjamins.

Greenberg, Joseph H. 1990[1975]. Dynamic aspects of word order in the numeral classifier. *On Language: Selected Writings of Joseph H. Greenburg*, ed. by Keith M. Denning & Suzanne Kemmer, 227-240. Stanford: Stanford University Press. [First published 1975 in *Word Order and Word Order Change*, ed. by Charles N. Li, 27-43. Austin: University of Texas Press.]

Her, One-Soon. 2009. Language and group identity: on Taiwan Mainlanders' mother tongue and Taiwan Mandarin. *Language and Linguistics* 10.2:375-419. (In Chinese)

Hsieh, Chen-Tien. 2009. *A Frame-based Approach to Classifiers: A Case Study of Taiwan Mandarin*. Taipei: National Chengchi University MA thesis.

Hsieh, Miao-Ling. 2005. Two types of modifiers and parallelisms between DPs and TPs in Chinese. *Language and Linguistics* 6.3:397-429.

Hsieh, Miao-Ling. 2008. *The Internal Structure of Noun Phrases in Chinese*. Taipei: Crane.

Hsu, Hui-ju. 2006. *Some Aspects of Phonological Leveling in Taiwan Mandarin*. Taipei: National Taiwan Normal University dissertation.

Hu, Qian. 1993. *The Acquisition of Chinese Classifiers by Young Mandarin-speaking Children.* Boston: Boston University dissertation.

Huang, C.-T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Cambridge: MIT dissertation.

Huang, Chu-Ren, and Kathleen Ahrens. 2003. Individuals, kinds and events: classifier coercion of nouns. *Language Sciences* 25.4:353-373.

Hung, Feng-Sheng. 1996. *Prosody and the Acquisition of Grammatical Morphemes in Chinese Languages.* Bloomington: Indiana University Linguistics Club.

Iljic, Robert. 1994. Quantification in Mandarin Chinese: two markers of plurality. *Linguistics* 32.1:91-116.

Kant, Immanuel. 1998[1781]. *Critique of Pure Reason*. Translated by Paul Guyer & Allen W. Wood. Cambridge & New York: Cambridge University Press.

Li, Wendan. 2000. The pragmatic function of numeral-classifiers in Mandarin Chinese. *Journal of Pragmatics* 32.8:1113-1133.

Li, Y.-H. Audrey. 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 8.1:75-99.

Liang, Yu-Chang. 2006. *Nominal Phrases in English and Japanese Speakers' L2 Mandarin Grammars.* Cambridge: Cambridge University dissertation.

Lin, Jo-wang. 1997. Noun phrase structure in Mandarin Chinese: DP or NP? *Chinese Languages and Linguistics*, Vol. 3: *Morphology and Lexicon*, ed. by Feng-fu Tsao & H. Samuel Wang, 401-434. Taipei: Institute of History and Philology, Academia Sinica.

Loke, Kit-Ken. 1983. *A Psycholinguistic Study of Shape Features in Chinese* (*Mandarin*) *Sortal Classifiers.* York: The University of York dissertation.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Paris, Marie-Claude. 1981. *Problèmes de syntaxe et de sémantique en linguistique chinoise*. Paris: Collège de France, Institut des Hautes Etudes Chinoises.

Rey, Georges. 2003. The analytic/synthetic distinction. *Stanford Encyclopedia of Philosophy*, ed. by Edward Zalta. Accessed online, December 1, 2009. http://plato.stanford.edu/entries/analytic-synthetic/

Robertson, Teresa. 2008. Essential vs. accidental properties. *Stanford Encyclopedia of Philosophy*, ed. by Edward Zalta. Accessed online, December 1, 2009. http://plato.stanford.edu/entries/essential-accidental

Sato, Yosuke. 2009. Radical underspecification, general number, and nominal denotation in Indonesian: an exo-skeletal approach. Manuscript. Vancouver: University of British Columbia.

Tai, James H-Y. 1994. Chinese classifier systems and human categorization. *In Honor of William S-Y. Wang: Interdisciplinary Studies on Language and Language Change*, ed. by Matthew Y. Chen & Ovid J. L. Tzeng, 479-494. Taipei: Pyramid.

Tai, James H-Y. 2003. Cognitive relativism: resultative construction in Chinese. *Language and Linguistics* 4.2:301-316.

Tai, James H-Y., and Lianqing Wang. 1990. A semantic study of the classifier *tiao* (條). *Journal of the Chinese Language Teachers Association* 25.1:35-56.

Tang, Chih-Chen Jane. 1990. A note on the DP analysis of the Chinese noun phrase. *Linguistics* 28.2:337-354.

Tang, Chih-Chen Jane. 2005. Nouns or classifiers: a non-movement analysis of classifiers in Chinese. *Language and Linguistics* 6.3:431-472.

Tsai, Wei-Tien Dylan. 2003. Three types of existential quantification in Chinese. *Functional Structure*(*s*), *Form and Interpretation: Perspectives from East Asian Languages*, ed. by Y.-H. Audrey Li & Andrew Simpson, 161-179, London: RoutledgeCurzon.

Watanabe, Akira. 2006. Functional projections of nominals in Japanese: syntax of classifiers. *Natural Languages and Linguistic Theory* 24.1:241-306.

Zhang, Hong. 2007. Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16.1:43-59.

Zhang, Niina. 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/~lngnz/index. files/May%202009.pdf

Zhu, Dexi. 1982. *Yufa Jiangyi* [*Lectures on Grammar*]. Beijing: The Commercial Press.

[Received 3 December 2009; revised 4 March 2010; accepted 15 April 2010]

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

Chen-Tien Hsieh
Graduate Institute of Linguistics
National Chengchi University
64, Sec. 2, Zhinan Road
Taipei 116, Taiwan
faye2815@yahoo.com.tw

論華語中分類詞與量詞之語意區分

何萬順　　謝禎田
國立政治大學

分類詞與量詞是否能精準的區分一直存有爭議。贊成區分的學者所提出的兩個測試：「『的』插入法」與「形容詞修飾法」，已被證實缺乏準確性。本文深入檢視此二測試法，進而提出兩組精確且真實可靠之測試。並且運用亞里斯多德對於「本質特徵、偶然特徵」、以及康德對「分析命題、綜合命題」之區分，適切地描繪出「分類詞、量詞」之區辨。由於量詞具有實質之語義，因此阻絕了數詞及形容詞對名詞的修飾；相對的，分類詞僅彰顯名詞本身既有之某些語義特徵，並不貢獻任何額外的語義，因此數詞及形容詞可穿透分類詞而修飾名詞。

關鍵詞：類別詞，量詞，本質特徵，分析命題