

An HPSG approach for Chinese numeral classifiers

Chenyuan Deng
dengchen@hu-berlin.de
Humboldt-Universität zu Berlin

1 Introduction

This paper proposes a new HPSG analysis for numeral classifiers, including sortal classifiers (1a), measure words (1b) and kind classifiers (1c) in Num(eral)-CL(assifier)-NP in Mandarin Chinese. Based on the same syntactic behavior, sortal classifiers and measure words will be treated syntactically in the same way, while measure words and kind classifiers will be analyzed semantically unified because of their similar semantic nucleus. As the data suggest, it is the classifier that is the *HEAD* of a noun phrase in Mandarin Chinese. The combination of both NPs (CLs and NPs) is licensed by a *head-argument-phrase*.

- | | | | |
|-----|---|---|--|
| (1) | a. san ben shu
three CL book
'three books' | b. san xiang shu
three box book
'three boxes of books' | c. san zhong shu
three kind book
'three kinds of books' |
|-----|---|---|--|

2 Previous studies

There is not much discussion about the structure of Mandarin numeral classifiers in HPSG. However, all the analyses treat NP in Num-CL-NP as the head of the noun phrase, and Num-CL has the *SPEC* position of NPs (cf. Gao 1993; Xue & McFetridge 1995; Liu 1997; Ng 1997). Since Xue & McFetridge (1995) suggests that Dem(onstratives) and Num-CL do not form a constituent, and that the Dem is higher than Num-CL-NP, left-peripheral modifiers like *da de* 'big DE' in (2) challenge this kind of analysis. According to the [_{NP} [_{SPEC} Num-CL] NP] structure, the adjectival phrase at the very periphery should modify the head noun *hun-tun* 'wonton', but in fact, the adjectival phrase only has scope over the classifier *wan* 'bowl'.

- (2) da de na **wan** xiao hun-tun
big DE that bowl small wonton
'that big bowl of small wonton'

The binary features [+definite] and [+quantified] proposed by Ng (1997) are good ways to stop recursion, which will also be applied to my subsequent analysis. However, Ng does not consider kind classifiers such as (1c). Kind classifiers allow a multiple-classifier (Liao & Wang 2011) together with sortal classifiers or measure words, cf. (3).

- (3) wo chi-le qi **ge** na liang **zhong** bang-bang-tang
1.SG eat-ASP seven CL that two kind lollipop
'I ate seven of those two kinds of lollipops.'

This indicates that the existing HPSG analyses have problems in dealing with adjectival scope as well as recursion. Therefore, a new analysis is needed. In the following sections, I will first introduce the subclasses of classifiers, and then discuss the branching issue of a classifier phrase as well as the categories of each part. After presenting the syntactic analysis, I will use Minimal Recursion Semantics (Copestake et al. 2005) as the semantic framework for the semantic analysis.

3 Classifiers and their subclasses

In general, there are two kinds of counting expressions in Mandarin Chinese (excluding kind classifiers first), sortal classifiers (1a) that simply name the natural unit of nouns and measure words (1b) (also measure classifiers) that create a unit of measure (Cheng & Sybesma 1999: 515).

Clearly, there is a semantic distinction between sortal classifiers and measure words, which will be discussed in Section 6. However, sortal classifiers and measure words share a lot of syntactic similarities. They both

allow NP-ellipsis, cf. (4); Nums can be omitted in directly postverbal position when Num=1, cf. (5); Both have the same behavior regarding the topicalization of NP, cf. (6).

- (4) wo mai-le san **ben** / **xiang** shu, ta mai-le wu **ben** / **xiang** -
 1.SG buy-ASP three CL box book 3.SG buy-ASP five CL box
 'I bought three books and he bought five books / I bought three boxes of books, he bought five boxes of books.'
- (5) wo mai-le - **ben** / **xiang** shu
 1.SG buy-ASP CL box book
 'I bought a book / I bought a box of books.'
- (6) shu_i, wo mai-le san **ben** / **xiang** i
 book 1.SG buy-ASP three CL box
 'I bought three books / three boxes of books.'

Furthermore, both sortal classifiers and measure words can be modified by adjectives (7). The typical noun phrase marker *de* (Sun 2015) that usually occurs after modifiers and complements like (8) can also be inserted between sortal classifiers or measure words and head nouns, cf. (9).

- (7) a. yi da **ben** shu
 one large CL book
 'a large book'
- b. yi da **xiang** shu
 one large box book
 'a large box of books'
- (8) a. bai de zhi
 white DE paper
 'white paper'
- b. chengshi de huimie
 city DE destruction
 '(the) destruction of (the) city'
- (9) a. wo mai-le yi-bai **ben** de shu
 1.SG buy-ASP 100 CL DE book
 'I bought 100 books.'
- b. wo mai-le yi-bai **xiang** de shu
 1.SG buy-ASP 100 box DE book
 'I bought 100 boxes of books.'

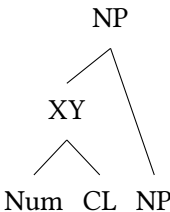
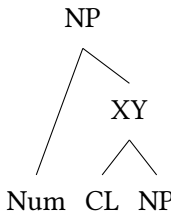
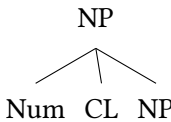
As mentioned above, there is another subcategory of classifiers, namely kind classifiers. Kind classifiers have all the above characteristics. But unlike sortal classifiers and measure words, kind classifiers can form a multiple-classifier together with sortal classifiers or measure words, cf. (3), whereas the recursion of classifier structures is generally not possible, cf. (10).

- (10) *wo chi-le liang **zhi** shi **ge** bang-bang-tang
 1.SG eat-ASP two CL ten CL lollipop

To summarize briefly, classifiers, measure words and kind words are all subclasses of classifiers. Classifiers are used to refer to all of this three in this study. It can be inferred from these syntactic similarities that measure words and sortal classifiers behave syntactically consistently. Following Tang (2005), Her (2012b) and Hsu (2015), I will treat measure words and sortal classifiers equally. Kind classifiers will be discussed separately because of their different behavior in the recursion problem.

4 Internal branching of a classifier phrase

As a noun phrase with three elements - Numeral (Num), classifier (CL), noun phrase (NP) - in a fixed order Num-CL-NP, there are only three branching possibilities for the structure, that is:

- (11) a. 
- b. 
- c. 

Since the presence of classifiers is always dependent on other words, namely on a Num, a Dem (demonstrative) or an NP, that is, a classifier must be combined with other elements to participate in the syntactic context, a ternary structure is actually ruled out.

Therefore, the question is: which element does a classifier combine with first? This has been a topic of debate in Chinese classifier researches for decades. There are generally three main approaches:

- Left branching: Classifiers are for numerals, i.e. [[Num CL] NP]
 - Tang (1990); Krifka (1995); Her (2012a); Bale & Coon (2014); Her & Tsai (2020); Tang et al. (2021)
- Right branching: Classifiers are for nouns, i.e. [Num [CL NP]]
 - Chierchia (1998); Cheng & Sybesma (1998, 1999); Borer (2005); Tang (2005); Huang et al. (2009); Hsu (2015)
- A split analysis: Based on different subclasses of CL, two structures exist.
 - Zhang (2011) and Li (2013)

The idea of a left branching structure can be attributed primarily to Chierchia (1998), namely the classifier/non-classifier distinction is dependent on the syntactic mass/count distinction. Cheng & Sybesma (1999) also argue that sortal classifiers are only used in conjunction with count nouns, whereas there is no such restriction for measure words. Different syntactic behaviors of sortal classifiers and measure words are presented as arguments, which are now widely accepted as no such distinction exists, cf. (7) and (9). Besides, sortal classifiers can also be combined with mass nouns. In (12), *zhang*, *li* and *ge* are typical classifiers for count nouns *ticket*, *cherry* and *cup*. But at the same time, they are also the typical classifiers of mass nouns *paper*, *rice* and *advice*.

- | | |
|--|--|
| (12) a. san zhang piao
three CL ticket
‘three tickets’
b. san li ying-tao
three CL cherry
‘three cherries’
c. san ge bei-zi
three CL cup
‘three cups’ | (13) a. san zhang zhi
three CL paper
‘three pieces of paper’
b. san li mi
three CL rice
‘three grains of rice’
c. san ge jian-yi
three CL advice
‘three pieces of advice’ |
|--|--|

In Mandarin Chinese, a classifier is not chosen based on the mass/count distinction but, rather, ontologically on Chinese cognition of nouns: category, shape, function, etc. For example, *tiao* in (14) is a typical classifier for long-shape objects. Therefore, *fish* and *trousers* are usually paired with it. But *shorts* do not have a long shape, and *tiao* is still their typical classifier, since *shorts* are a subset of *trousers*.

- | | | |
|--|--|--|
| (14) a. yi tiao yu
one CL fish
‘1 fish’ | b. yi tiao ku-zi
one CL trousers
‘1 pair of trousers’ | c. yi tiao duan-ku
one CL shorts
‘1 pair of shorts’ |
|--|--|--|

The left branch’s view is more or less influenced by Krifka’s hypothesis (1995) that there are two different types of numeral interpretation in classifier and non-classifier languages. In Mandarin Chinese, an Num does not have an incorporated measure function and, thus, requires a classifier.

Considering that there is no mass/count distinction at the classifier level, I prefer a left branching structure based on the following facts:

First, Num-CL can be the answer of a question, while there is no such possibility for a CL-NP sequence, cf. (15). Second, when talking about the amount, a classifier always follows the numeral even if the noun does not appear or cannot be completed with a specific noun, cf. (16). In (17a), *san ge ban* ‘three CL half’ means *three and a half*. Because **ge ban* ‘CL half’ and **ban xi-gua* ‘half watermelon’ are ungrammatical, but *ban ge* ‘half CL’ is grammatical. It is reasonable to consider that two Num-CL structures, namely *san ge* ‘three CL’ and *ban ge* ‘half CL’ are combined, and the second classifier is deleted. Similarly, Her & Tsai (2020)’s interpretation

of the approximant numeral *lai* in (17b) that *lai* is uninterpretable without an immediately preceding numeral base, supports the left branching view.

- | | | |
|------|--|---|
| (15) | Q: ni mai-le duo-shao shu
2.SG buy-ASP how.many book
'How many books did you buy?' | A: *yi / yi ben / * ben shu
one one CL CL book
'One (book).' |
| (16) | a. xue-sheng de shu-liang shi 20 *(ming)
student DE number be 20 CL
'The number of students is twenty.' | b. shi yuan san *(ge)
ten dollar three CL
'ten dollar for three' |
| (17) | a. san ge ban xi-gua
three CL half watermelon
'three and a half watermelons' | b. shi lai ben shu
ten more CL book
'more than 10 books (10-19 books)' |

Further evidence can be found in various fields like phonology (left branching tone sandhi in Tang et al. (2021)), cross-linguistics (Nums and CLs are worldwide correlated in Her (2017)), historical linguistics (developmental Stages *horse 5 horse* → *horse 5 CL* → *5 CL horse* in history in Huang (1964)), etc.

5 Categorization and an HPSG approach

After identifying the branching issue of the structure, it is necessary to determine the categories of each element to propose an HPSG analysis. More precisely, should a classifier be treated as a new category or can it be integrated into an existing word class?

Classifiers can be considered as a subclass of nouns and have the HEAD *noun* in HPSG. First, as subclasses of classifiers, the vast majority of measure words and kinds are themselves nouns. Second, as demonstrated in (7), adjectives can be attached directly to the sortal classifiers, which is one of the basic features of nouns. Otherwise, a new rule allowing adjectives to attach to classifiers is needed. Third, in terms of the developmental stages of classifiers, the emergence of classifiers replaces the original noun position (Huang 1964). Forth, sortal classifiers can form compound nouns with their matching nouns. Reduplicative sortal classifiers can be used as a noun with context, cf. *duo-duo* 'CL-CL' in (19). And finally, in other typical numeral classifier languages like Japanese and Korean, classifiers are considered as a subclass of nouns (Bender & Siegel 2004; Kim & Yang 2007). There is no reason to accept a new part of speech in HPSG for Chinese classifiers.

- | | | |
|------|---|--|
| (18) | a. yi duo hua
one CL flower
'one flower' | b. hua- duo
flower-CL
'flower(s)' |
| (19) | yuan-zi li you hen-duo hua, duo-duo dou hao-kan
yard in there-be a-lot-of flower CL-CL all beautiful
'There are a lot of flowers in the yard, and every one is beautiful.' | |

Since the HEAD of a classifier is *noun*, it is reasonable to consider making the classifier to be the head of the noun phrase Num-CL-NP. This is also consistent with the data in (2), where the left-peripheral modifiers can only have scope over the head of the noun phrase, in this case, the classifiers. Considering the inseparable relationship between Nums and CLs and the fact that in the position of a Num can appear demonstratives and interrogative determiners, the Num will be analyzed as a specifier of the classifier and selected by a classifier via SPR. A *head-specifier-phrase* licensed the Num-CL combination.

The feature descriptions in (21) represents the syntactic part of sortal classifiers and measure words. Sortal classifiers and measure words select an NP through the COMPS-list. A *head-argument-phrase* enables the two NPs to be combined. Following Ng (1997), I use two binary features QUANTIFIED+ and DEFINITE+ to stop the recursion of classifiers and demonstratives. All nouns (except classifiers) have the feature QUANTIFIED- and DEFINITE-. Sortal classifiers and measure words have the feature QUANTIFIED+ and can only select nouns with QUANTIFIED- and DEFINITE- or nouns with KIND+ and DEFINITE+. Once an NP has been selected by a

sortal classifier or measure word, this head feature will be preserved and therefore no other classifiers with *QUANTIFIED+* will be involved. Since sortal classifiers and measure words can choose a noun phrase with a kind classifier as the head, the *KIND+* feature is used. But in this case, a demonstrative in Dem-Num-KC-NP is required, so the chosen noun phrase must also have the *DEFINITE+* feature, cf. (20).

- (20) wo chi-le qi **ge** *(na) liang **zhong** bang-bang-tang
 1.SG eat-ASP seven CL that two kind lollipop
 ‘I ate seven of those two kinds of lollipops.’

- (21)
$$\left[\begin{array}{l} sc_mw_n \\ \text{HEAD } noun \\ \text{SPR } \left(\left[\text{HEAD } det \right] \right) \\ \text{COMPS } \left(\left[\begin{array}{l} \text{HEAD } noun \\ \text{DEFINITE-} \\ \text{QUANTIFIED-} \end{array} \right] \vee \left[\begin{array}{l} \text{HEAD } noun \\ \text{KIND+} \\ \text{DEFINITE+} \end{array} \right] \right) \\ \text{QUANTIFIED+} \end{array} \right]$$
- (22)
$$\left[\begin{array}{l} kc_n \\ \text{HEAD } noun \\ \text{SPR } \left(\left[\text{HEAD } det \right] \right) \\ \text{COMPS } \left(\left[\begin{array}{l} \text{HEAD } noun \\ \text{DEFINITE-} \\ \text{QUANTIFIED-} \end{array} \right] \right) \\ \text{KIND+} \end{array} \right]$$

Different from the feature descriptions of sortal classifiers and measure words, *KIND+* feature appears in the lexicon for kind classifiers in (22). Because the kind classifiers allow to be selected by other classifiers (but not kind classifiers themselves), they still select an NP with *QUANTIFIED-*. Similarly, the feature *DEFINITE+* is designed to ensure that only one Dem can appear within a noun phrase and that the Dem does not appear in the pre-NP position in the presence of a classifier, cf. (23b).

- (23) a. nei shu
 that book
 ‘that book’
- b. (nei) wo xie de (nei) 3 **ben** (*nei) shu
 that 1.SG write DE that 3 CL that book
 ‘the three books I wrote’

6 Semantics

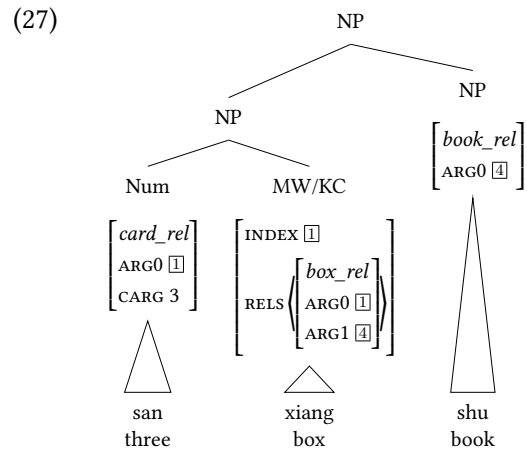
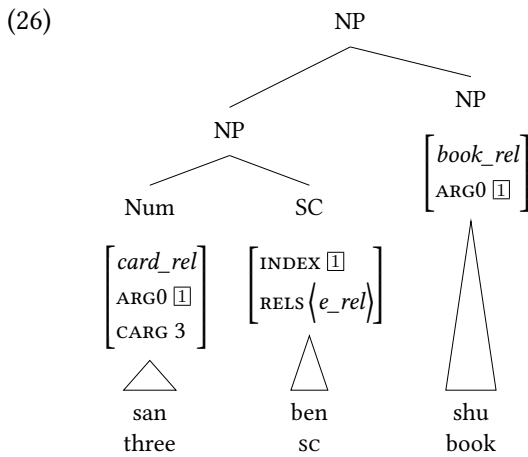
The discussion of semantics of classifiers begins with (7). In these counterexamples with pre-CL modifiers, sortal classifiers do not contribute to semantics of the phrase. But this does not mean that adjectives have scope over NPs. The examples (24) from Her (2012a) exhibit which sortal classifier is chosen depends on which characteristic of the noun we want to emphasize. *tiao*, *wei* and *zhi* are all possible sortal classifiers for *fish* because they present the features that the fish itself has. This is also known as the mapping between sortal classifiers and nouns.

- (24) a. yi **tiao** yu
 one CL fish
 ‘a fish’
 (SC: long shape)
- b. yi **wei** yu
 one CL fish
 ‘a fish’
 (SC: tail)
- c. yi **zhi** yu
 one CL fish
 ‘a fish’
 (SC: animacy)

In other words, sortal classifiers contribute no semantic content of their own, hence I denote them as *e_list*. The adjective attached to the sortal classifiers modifies the NP, and this modifying relationship is reflected by the shared *INDEX* value, i.e. [1] in (26).

MRS representations in (26) and (27) differ from each other in *RELS*. Measure words and kind classifiers themselves are nouns and have their own meanings, such as *box_rel* in (27). In *box_rel*, measure words constitute relations with an *ARG0* and *ARG1*, i.e. the semantics of NP, [4] in (27). All possible nouns can be transformed into a measure word by a lexical rule, which brings *SPR*, *COMPS* and other features of a classifier, cf. (25).

- (25) yi **wu-zi** / **che** shu
 one room car book
 ‘a room full of books / a carload of books’



7 Conclusions

This study discusses the numeral classifiers in Mandarin Chinese and proposes a new HPSG analysis for the Num-CL-NP structure. In this approach, three subclasses of classifiers are differentiated, i.e., sortal classifiers, measure words and kind classifiers. Sortal classifiers and measure words are treated syntactically equivalent based on their same syntactical behavior. The difference between these two subclasses is that measure words have their own semantics, while sortal classifiers contribute no semantic content. Kind classifiers are distinguished from sortal classifiers and measure words by taking the multiple-classifier structure into account, but semantically they are identical to the measure words. A noun phrase consisting of Num-CL-NP is a left-branching structure, i.e., a classifier selects its numeral via SPR to yield a unit and is combined with an NP via the *head-argument-phrase*. Such a structure is based on their own syntactic behaviors and is unrelated to the mass/count distinction.

References

- Bale, Alan & Jessica Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry* 45(4). 695–707.
- Bender, Emily M. & Melanie Siegel. 2004. Implementing the syntax of Japanese numeral classifiers. In *International Conference on Natural Language Processing*, Berlin, Heidelberg: Springer.
- Borer, Hagit. 2005. *In name only*, vol. 1. Oxford University Press on Demand.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 1998. yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies* 28(3). 385–412.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of np. *Linguistic inquiry* 30(4). 509–542.
- Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6. 339–405. doi:https://doi.org/10.1023/A:1008324218506.
- Copestake, Ann, Daniel P. Flickinger, Carl J. Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation* 3(4). 281–332. doi:10.1007/s11168-006-6327-9.
- Gao, Qian. 1993. Chinese np structure. In A. Pollard, C. and Kathol (ed.), *Ohio State University Working Papers in Linguistics 42*, Ohio State University. Department of Linguistics.
- 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). 1211–1511.
- Her, One-Soon. 2017. Structure of numerals and numeral classifiers in Chinese: Historical and typological perspectives and cross-linguistic implications. *Language and Linguistics* 18(1). 26–71.
- Her, One-Soon & Hui-Chin Tsai. 2020. Left is right, right is not: On the constituency of the classifier phrase in Chinese. *Language and Linguistics* 21(1). 1–32.
- Hsu, Yu-Yin. 2015. A unified account to measure words in Mandarin. *University of Pennsylvania Working Papers in Linguistics* 21(1). 14.
- Huang, Cheng-Teh James, Yen-hui Audrey Li & Yafei Li. 2009. *The syntax of Chinese*, vol. 10. Cambridge: Cambridge University Press.
- Huang, Zaijun. 1964. Cong jiawen jinwen liangci de yingyong kaocha hanyu liangci de qiuyuan yu fazhan (a study of the origin and development of classifiers in Chinese, beginning from their use in the oracle bone inscriptions and metal inscriptions). *Zhongguo Yuwen* 133(6). 432–441.
- Kim, Jong Bok & Jaehyung Yang. 2007. Syntax and semantics of Korean numeral classifier constructions. In Stefan Mueller (ed.), *Proceedings of the 14th international conference on head-driven phrase structure grammar*, 163–172. doi:10.21248/hpsg.2007.10.
- Krifka, Manfred. 1995. Common nouns: A contrastive analysis of Chinese and English. In Gregory N. Carlson & Francis Jeffrey Pelletier (eds.), *The generic book*, 398–411. Chicago: University of Chicago Press.
- Li, Xuping. 2013. *Numeral classifiers in Chinese: The syntax-semantics interface*. Beijing: De Gruyter Mouton.
- Liao, Wei-wen Roger & Yuyun Iris Wang. 2011. Multiple-classifier constructions and nominal expressions in Chinese. *Journal of East Asian Linguistics* 20(2). 145–168.
- Liu, Gang. 1997. *Eine unifikations-basierte Grammatik für das moderne Chinesisch – dargestellt in der HPSG*. Konstanz: Universität Konstanz PhD Thesis.
- Ng, Say Kiat. 1997. *A double-specifier account of Chinese nps using head-driven phrase structure grammar*: University of Edinburgh MSc thesis.
- Sun, Chaofen. 2015. The uses of de as a noun phrase marker. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 362–376. New York: Oxford University Press.
- Tang, Chih-Chen Jane. 1990. *Chinese phrase structure and the extended x'-theory*: Cornell University PhD thesis.
- Tang, Chih-Chen Jane. 2005. Nouns or classifiers: A non-movement analysis of classifiers in Chinese. *Language and Linguistics* 6(3). 431–472.
- Tang, Marc, Ying-Chun Chen, Nai-Shing Yen & One-Soon Her. 2021. Investigating the branching of Chinese classifier phrases: Evidence from speech perception and production. *Journal of Chinese Linguistics* 49(1). 71–105.
- Xue, ArtemisPing & Paul McFetridge. 1995. Dp structure, hpsg and the Chinese NP. In *Proceedings of the 14th Annual Conference of Canadian Linguistics Association*.
- Zhang, Niina Ning. 2011. The constituency of classifier constructions in Mandarin Chinese. *Taiwan Journal of Linguistics* 9(1). 1–50.