

Syntax of classifiers and measure words in Chinese and beyond

by

Hengliang Xia

Doctoral thesis submitted to the University of Ottawa
In partial fulfillment of the requirements for the PhD in Linguistics degree

Summer 2025

Department of Linguistics
Faculty of Arts
University of Ottawa

© Hengliang Xia, Ottawa, Canada, 2025

Abstract

This thesis deals with classifiers and measure words in Chinese, focusing specifically on Mandarin, Cantonese and Northern Wu. It provides a typological account of the variation observed in these languages and offers both a synchronic and a diachronic approach to syntactic variation.

Numeral classifiers are known to occur obligatorily in quantified noun phrases in Chinese (Chierchia 1998: 354-355). Different accounts have been proposed in the literature for these: some researchers propose a left-branching structure in which the classifier merges with the numeral (Greenberg 2013 [1974]; Her 2017), while others argue for a right-branching structure in which the classifier merges directly with the head noun (Cheng & Sybesma 1999; Simpson 2005). In addition, it has been observed that Chinese languages have measure words such as *shao* ‘spoonful’, and it has been proposed that they are syntactically identical to classifiers (Her 2017).

In this thesis, I propose that classifiers should be distinguished from measure words, and that both structures are needed for a complete analysis of the classifiers and measure words in Mandarin, Cantonese and Northern Wu. I hypothesize that classifiers and measure words can be syntactically different, and that they tend to undergo different types of diachronic changes, which may account for their synchronic syntactic differences.

These hypotheses are tested against data collected from native speakers of Northern Wu and Cantonese with a grammaticality judgement task; the historical data come from previous descriptive works and corpora of Old and Middle Chinese. The findings include a systematic syntactic difference between classifiers and measure words in Cantonese and Northern Wu (which is not found in Mandarin), as well as several types of diachronic changes that can be subsumed under grammaticalization.

Based on those data, I argue that classifiers differ syntactically and categorially from measure words in Cantonese and Northern Wu, and that they are syntactically distinct from Mandarin classifiers. They are associated with different structures: classifiers and measure words in Mandarin as well as Cantonese and Northern Wu measure words occur in a left-branching structure, while classifiers in Cantonese and Northern Wu are used in a right-branching structure. These differences between the three Chinese languages are attributable to the diachronic changes that they underwent, and I propose a correlation between the syntactic behaviour of classifiers and their level of grammaticalization.

This thesis on Chinese classifiers provides original work in that synchronic syntactic properties are directly linked with diachronic change and in proposing a universal left-branching structure for measure words and language-specific structures for classifiers. My theory on Chinese classifiers has several major theoretical consequences. First, classifiers and measure words are associated with different functional heads. Second, the classifiers that occur in right-branching structures in quantified noun phrases are expected to show other signs of grammaticalization (e.g. phonological reduction). Third, numeral classifiers and number morphology are not in complementary distribution (Bisang 2012), and their co-occurrence is costly but not intrinsically prohibited. Last, some properties of classifiers could be analyzed as emergent rather than fixed (Hopper 1998), and they explain why classifiers undergo some changes diachronically and bear certain vestigial similarities to measure words despite more systematic differences.

Table of Contents

ABSTRACT	II
-----------------	-----------

TABLE OF CONTENTS	III
--------------------------	------------

ACKNOWLEDGEMENTS	V
-------------------------	----------

CHAPTER 1

INTRODUCTION	1
---------------------	----------

1.1 The general context: an overview of the characteristics of Chinese classifiers	1
1.2 The puzzles and research questions	4
1.3 Hypotheses and relevant literature	6
1.4 Methodology	16
1.5 Outline of the thesis	18

CHAPTER 2

THE CLASSIFIERS AND MEASURE WORDS IN MANDARIN CHINESE	19
--	-----------

2.1 CL and MW as one and the same grammatical category	20
2.2 A continuum from individual classifiers to standard measures: phonological, morphological, semantic, and lexical differences between CL and MW	29
2.2.1 Phonological similarities and differences between classifiers and MWs	32
2.2.2 Morphological similarities and differences between classifiers and MWs	40
2.2.3 Semantic similarities and differences between classifiers and MWs	43
2.2.4 Lexical characteristics of classifiers and MWs	48
2.3 Evidence for [Num-CL/MW] constituency in Mandarin Chinese	49

CHAPTER 3

THE CLASSIFIERS AND MEASURE WORDS IN WU AND CANTONESE	62
--	-----------

3.1 Distribution of classifiers in Northern Wu and Cantonese	63
3.2 Results of the surveys about Cantonese classifiers and MWs	67
3.3 Results of the surveys about Northern Wu classifiers and MWs	82
3.4 Discussion of the results	96
3.5 Interim conclusion	104

CHAPTER 4

<u>THE EVOLUTION OF CHINESE CLASSIFIERS: AN ACCOUNT BASED ON</u>	
<u>GRAMMATICALIZATION</u>	106
4.1 A brief history of Chinese classifiers	107
4.1.1 OLD CHINESE	107
4.1.2 EARLY MIDDLE CHINESE	112
4.1.3 LATE MIDDLE CHINESE	114
4.1.4 MODERN CHINESE LANGUAGES: MANDARIN, NORTHERN WU AND CANTONESE	117
4.2 Changes in the classifier system as grammaticalization	122
4.3 A diachronic comparison of Chinese classifiers and measure words	125
4.4 Conclusion	128
 <u>CHAPTER 5</u>	
<u>THEORETICAL IMPLICATIONS AND RELATED ISSUES</u>	130
5.1 RELEVANT ISSUES	130
5.1.1 The mass-count distinction in Chinese and beyond	131
5.1.2 Grammatical number in Chinese	135
5.1.3 Optionality and omission of Chinese classifiers	142
5.2 Theoretical and typological implications	146
5.3 CONCLUSION	150
 <u>CHAPTER 6</u>	
<u>CONCLUSION</u>	152
 <u>BIBLIOGRAPHY</u>	154
<u>APPENDIX : CANTONESE AND NORTHERN WU QUESTIONNAIRES</u>	160

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Éric Mathieu for his constant support, great patience, and valuable guidance. Since 2019, I have benefited immensely from his advice and encouragement, without which my graduate studies would have been much more difficult.

I wish to thank the readers of my thesis, including Prof. Naomi Francis, Prof. Dennis Ott, Prof. Andrés Pablo Salanova and Prof. Niina Ning Zhang, whose suggestions, comments and questions have been enormously helpful. A sincere thank-you goes to Prof. Marc Brunelle for his advice and help when I worked on my second comprehensive exam.

At the University of Ottawa, I am deeply appreciative to Scarlet Li, Yubin Xing and Adam Cheng for their helpful comments on my surveys and presentations.

My thanks also go to the attendees of the MOTH Syntax Workshop in Ottawa (2024) in Toronto (2025) and the CLA conference in Montréal (2025) for their comments and questions on my presentations based on previous versions of my thesis.

I wish to express my appreciation for my parents' unceasing support. I would also like to thank my cousins Jeff Ying and Ellen Yu for their support and encouragement over the past years.

1. Introduction

1.1 The general context: an overview of the characteristics of Chinese classifiers

The modern Chinese languages, including Mandarin, Northern Wu, and Cantonese, have a group of morphemes known as classifiers.¹ In general, a classifier connects the head noun with another element (e.g. a numeral or a demonstrative), and it individuates² and classifies the noun. In most cases, there are no literal equivalents for classifiers in English, except for the “unit counters” such as the word *head* in *two heads of cabbage*, which are roughly comparable to Chinese classifiers (Li 2013: 34). In Mandarin, Northern Wu, and Cantonese, the classifiers are generally obligatory when employed with numerals, and they are very often used with demonstratives. The usual constituent order in a simple Mandarin NP is Num-Cl-N when it contains a numeral, or Dem-Cl-N when it contains a demonstrative. These two types of NPs are shown in the example below.

- (1) 那 个人 看到 三 头 牛。 [Mandarin]
Na ge ren kandao san tou niu
That CL person see three CL cattle
'That person saw three head of cattle.'

¹ Other terms for classifiers (CLs) include “count-classifiers” (Cheng & Sybesma 1999), “individual classifiers” (Li 2013), “sortal classifiers” (Bale et al. 2019). Other terms for measure words (MWs) include “mass-classifiers” (Cheng & Sybesma 1999) and “mensural classifiers” (Aikhenvald 2000; Bale & Coon 2014). Typologically, these three Chinese languages are multiple classifier languages, and the classifiers are at least numeral classifiers and deictic classifiers (with some additional functions in Cantonese and Northern Wu) (Aikhenvald 2000).

² Here I assume that all nouns in Mandarin, Northern Wu and Cantonese are mass, and that classifiers have the function of making nouns compatible with numerals in these languages. The topic of the mass-count distinction will be discussed in more detail in Chapter 5.

Example (1) shows that the classifier for *ren* ‘person’ is *ge*, and that the classifier for *niu* ‘cattle’ is *tou*. A classifier can also occur in bare classifier-noun (Cl-N) phrases in the post-verbal position in Mandarin. These constructions and the same constituent orders are found in Cantonese and Northern Wu, and there are several other contexts in which classifiers can occur (e.g. the possessive construction) in these languages.

In addition to the classifiers, the Sinitic languages have some measure words (MWs), which are comparable to English words such as *herd* and *bottle* and are used in the similar constructions. An example with Mandarin measure words is given below.

(2) 那 群 牛 喝 了 十 桶 水。
 Na qun niu he-le shi tong shui
 That MW_{-group} cattle drink-ASP ten MW_{-bucket} water
 ‘That herd of cattle drank ten buckets of water.’

In (2), the noun *niu* ‘cattle’ co-occurs with the measure word *qun* ‘herd’, and the noncount noun *shui* ‘water’ is used with the measure word *tong* ‘bucket’ and is compatible with a wide variety of measure words. Lexically, Mandarin has about 50 to 70 individual classifiers and some 180 measure words (Li 2013: 21). There are several dozen classifiers in Northern Wu and Cantonese (Qian et al. 2007; Matthews & Yip 2011), and the numbers of measure words seem comparable to that of Mandarin MWs. In all three languages, a single set of lexical items (i.e. classifiers or MWs) are used in different constructions. In general, each count noun selects a certain classifier³, hence the term “classifier”.

There is no overt number marking on nouns in these three Chinese languages, which makes bare nouns ambiguous in number (Li 2013).⁴ There are no overt determiners like *the* or *a* in these Chinese languages, although the classifiers in Cantonese and Northern Wu can mark definiteness (among other things) in some contexts (Matthews & Yip 2011).

³ Some nouns are compatible with two classifiers, the distribution of which may depend on register.

⁴ The morpheme *-men* in Mandarin is a collective suffix (for nouns denoting humans) rather than a plural suffix (Cheng & Sybesma 1999: 536-537). This analysis should also apply to its counterpart in Northern Wu, namely the suffix *-laq*, which is used in very similar contexts.

Other than the classifiers, the nouns in these languages are not categorized in a different way, such as genders or noun classes.

The classifiers in these three Chinese languages share several characteristics that are cross-linguistically variable. First, Mandarin, Cantonese and Wu classifiers cannot be omitted when they are used with numerals, which is not true of all classifier languages. For instance, classifiers are “often optional” in Tibetan (Her 2017A: 32), and the classifiers in Bouyei (Northern Tai; Tai-Kadai: China) are omitted if the head noun is a kinship term or denotes a body part (Wang 1987: 295). Second, when a classifier is used with a numeral, the normal constituent order in Mandarin, Cantonese and Northern Wu is Num-Cl-N.⁵ Three other constituent orders are attested in other classifier languages, namely N-Num-Cl (e.g. in Thai), Cl-Num-N (e.g. in Ibibio [Niger-Congo]) and N-Cl-Num (e.g. in Jingpo [Tibeto-Burman]) (Aikhenvald 2000: 104-105; Her 2017B: 267). Furthermore, the three Chinese languages are all “multiple classifier languages” since the same classifiers are used in some other constructions in addition to Num-Cl-N phrases (Aikhenvald 2000: 204). This is different from the situation in Palikur (North Arawak) and Malto (Dravidian) where different sets of classifiers coexist in the same language (Aikhenvald 2000), and there are classifier languages in which the classifiers are only used with numerals (e.g. Burmese) (Jenny & Hnin Tun 2016: 52). Lexical sets of numeral classifiers can vary greatly in size among different classifier languages, ranging from four classifiers (e.g. in Nung [Tai-Kadai]) to several hundred (e.g. in Tzeltal [Mayan]) (Aikhenvald 2000: 103). In terms of length in syllables, the individual classifiers are always monosyllabic in the three Chinese languages in question. Many classifier languages (e.g. Uzbek, Assamese) do not seem to have a similar constraint on the length of classifiers (Aikhenvald 2000).

⁵ The word order N⁰-Num-Cl is found in presentational situations (Simpson 2005: 816), but it is rare and usually unacceptable elsewhere. Note that Num-Cl-N⁰ is still the predominant constituent order in Chinese, which is unlike Japanese and Korean where two or three constituent orders are commonly used (Aikhenvald 2000: 106-107).

Broadly speaking, the classifiers in the Chinese languages all perform two major functions at the same time. First, they classify nouns through semantic selection which is largely arbitrary (Zhang 2013: 163-164). Second, they individuate the classified noun so that it becomes compatible with quantifying morphemes including numerals. They need to fulfil this function because the bare nouns are mass-like expressions that refer to kinds according to the Krifka-Chierchia hypothesis (Chierchia 1998: 353-355; Li 2013: 87-88). In other words, the use of classifiers in Chinese is arguably related to the generic meaning of bare nouns and the lack of plural marking (Li 2013: 89). This claim is supported by the fact that the optional classifiers in Armenian cannot co-occur with the optional plural marking (Borer 2005: 94-95). These two functions appear to be the most basic functions of numeral classifiers in some other classifier languages (e.g. Japanese) as well (Chierchia 1998: 354). Historically, the classifying and individuating functions are the main functions of the first classifiers in Chinese, even though the classifiers are later associated with some additional functions (e.g. marking definiteness) in Cantonese and Northern Wu (C. Zhang 2012).

The functions of measure words (or mensural classifiers) are somewhat different from those of sortal classifiers. While sortal classifiers tend to be selected in a rather rigid fashion, the selection of a specific measure word is generally much more flexible and is “often determined by the temporary state of an object” (Aikhenvald 2000: 115). In other words, measure words do not classify nouns or do so in a fundamentally different way from the classifier (Zhang 2013: 166-167). However, measure words do fulfil the individuating function, allowing speakers to count ontologically uncountable nouns such as *water* and *flour*. Since the classifiers and measure words have overlapping functions and share the same distribution in Mandarin, they are compared in chapters 2, 3 and 4.

1.2 The puzzles and research questions

In a Num-Cl-N phrase, the classifier is linearly adjacent to both the numeral and the head noun. As a result, it is not apparently obvious at first sight what syntactic structure such NPs could have in the three Chinese languages under review. Numeral-classifier constituency and classifier-noun constituency are both logically plausible, and both have

been proposed multiple times (Zhang 2013: 155-156; Her 2017A: 37-38). The two structures may also occur in different languages. A major goal of this thesis is to determine the syntactic structure of Chinese NPs containing classifiers.

A second puzzle is whether there are fundamental differences between classifiers and measure words despite the fact that they occur in similar contexts. If there is adequate evidence to show that they belong to two distinct grammatical categories, then one may expect to see some differences in the way they merge into NPs. It is also conceivable that classifiers are very similar to measure words in some languages, and that they are more different from each other in other languages.

The third puzzle is about the evolution of Chinese classifiers. Despite the prevalence of classifiers in modern Sinitic languages, Old Chinese (i.e. the earliest attested form of Chinese) was not a classifier language for at least a millennium before 2nd century BCE. At that stage, precursors of classifiers existed in Old Chinese, but they were used infrequently in NPs with numerals and were not easily distinguishable from their lexical sources (i.e. nouns or verbs) (Yang-Drocourt 2004: 1-2; Li 2017: 176-178). The development of a classifier system in the history of Chinese may have brought about some syntactic changes in Chinese NPs, which I aim to describe in a later part of this thesis.

Two major research questions will be investigated in this thesis. First, do the Num-Cl-N phrases in Mandarin, Northern Wu and Cantonese have a left-branching structure or a right-branching structure? In other words, does the classifier merge with the numeral first or with the head noun first? Related to this is the question of the syntactic structure of Num-MW-N phrases: do they have the same structure as Num-Cl-N phrases? If not, in what ways are they different? Are there any syntactic differences between the three Chinese languages in terms of the syntax of classifiers? If so, how do the NPs in the three languages differ structurally from one another and why have the differences arisen?

Second, I study the following research question to solve the third puzzle. How did a classifier system come into existence in the history of Chinese? This also comprises a few sub-questions. When did Chinese become a classifier language and what was the situation before then? In what type of NPs were the classifiers first used? How did the emergence of a classifier system affect the syntax of Chinese noun phrases? What were the changes from

Middle Chinese to the three modern languages in the use of classifiers? What is the historic relationship between classifiers and measure words? Is there a link between the syntax of classifiers and their level of grammaticalization (Hopper & Traugott 2003; Wu 2005)?

I aim to resolve these research questions in the next three chapters of this thesis. My hypotheses are summarized in the next section, which also provides an overview of the existing literature relevant to my research questions.

1.3 Hypotheses and relevant literature

Regarding the first research question, my hypothesis is that the Num-Cl-N phrases have a left-branching structure in the three Chinese languages under review. I also hypothesize that measure words and classifiers belong to the same syntactic category and that Num-MW-N phrases have the same left-branching structure as Num-Cl-N phrases.

The literature on the branching directions of quantified NPs in classifier languages is quite extensive, and there are three types of accounts: unified left-branching accounts (i.e. [Num-CL] constituency in all classifier languages), unified right-branching accounts (i.e. [CL-N] constituency in all classifier languages), and split accounts (Her 2017A). Greenberg (2013 [1974]) studied the use of numeral classifiers in 100 languages from a wide array of areas and families, and he argues that the classifier first merges with the numeral to form a constituent, which then “enters into a more remote construction with the enumerated noun” (Greenberg 2013 [1974]: 325). He offers five types of evidence for this claim, summarized as follows. First, the constituent orders that are logically possible but unattested (viz. Cl-N-Num and Num-N-Cl) both contain a head noun intervening between the numeral and the classifier. Second, when there is language-internal variation in that regard, it is usually the relative order of the head noun and the combination of numeral and classifier that varies (e.g. N-Cl-Num as a variant of Num-Cl-N), while variation between Num-Cl and Cl-Num word order is rare. Third, the classifier is prosodically closer to the numeral than to the noun, which can be observed in the fusion of the numeral and classifier into one prosodic word in many languages. Fourth, it is not uncommon for the head noun to

be separable from the Num-Cl combination⁶ at least in some constructions. Lastly, all languages in the sample allow the Num-Cl combination to be used anaphorically “without overt expression of the noun” (Greenberg 2013 [1974]: 326). Greenberg (ibid. 307-308) also mentions some contexts where classifiers are seldom employed or frequently optional, which will be discussed in more detail in chapter 5 of this thesis.

Some works focused mainly on Chinese while assuming or arguing that [Num-Cl] constituency or the left-branching structure is universal. According to Li & Thompson (1981: 105), numeral-classifier combinations and demonstrative-(numeral)-classifier combinations are “classifier phrases”, although it is the noun that (lexically) selects specific classifiers. In addition, Li & Thompson discussed several types of measure words (which they often termed “measure classifiers”) but did not differentiate between measure words and count classifiers in terms of syntactic structure (Li & Thompson 1981: 106-112).

Her (2017A) also proposes a left-branching structure for the quantified NPs in all classifier languages. Her (2017A) demonstrates that the order of numerals and classifiers in those NPs correlates to the order of the multiplier and base in complex numerals such as *two hundred*, and that the numeral-classifier sequence is a constituent just like complex numerals (Her 2017A: 61-62). He also points out that the left-branching structure (with [Num-Cl] being a constituent) makes correct predictions about which orders are attested in classifier languages (ibid. 62-63). More specifically, Her (2017A: 38-41) adduces several reasons why [Num-Cl] is a constituent in Mandarin Chinese. First, using a coordinating conjunction to link two [Num-Cl/MW] constituents is allowed in Mandarin. Second, head nouns in Num-Cl/MW-N phrases can undergo ellipsis and extraction, while the [Cl/MW-N] sequence cannot. Third, the modification marker *de* can (sometimes) be inserted after the [Num-Cl/MW] part of the phrase⁷, and “it is uncontroversial that pre-*de* modifiers are constituents” (Her 2017A: 40). Lastly, the morphemes *-ban* ‘half’ and *-duo* ‘more’ can be

⁶ Greenberg (2013 [1972]) is referring to the constituent [Num-Cl], and the surface word order can be Num-Cl or Cl-Num depending on the language. This is also true of the last generalization.

⁷ It is worth noting that *de*-insertion is only possible in some contexts. For example, group classifiers and partition classifiers are often incompatible with *de*-insertion (Li 2013: 63). Individual classifiers following a lower numeral that denote exact values also tend not to allow *de*-insertion (Li 2013: 65).

inserted after the [Num-Cl/MW] sequence, which are interpreted with the numeral rather than directly with the head noun, thereby supporting the left-branching account of [Num-Cl/MW-N] phrases. Her (2017A: 50-54) also examines the diachronic development of Chinese classifiers and the word orders of quantified NPs that are attested in the history of Chinese. There have been the four types of constituent orders, all of which are explicable with [Num-Cl/MW] constituency but not with [Cl/MW-N] constituency.

While there is some support for the unified left-branching accounts, other linguists have proposed unified right-branching accounts. Simpson (2005) argues that the numeral and classifier are two separate heads and that the structure of quantified NPs is [Num [Cl NP]] in head-initial classifier languages. With NP-movement, this theory accounts for the word order variation among Southeast Asian languages such as Thai, Nung, Chinese and Vietnamese (Simpson 2005: 810-815). Simpson (2005: 808) points out that the function of classifiers and that of numerals are different, which are “individuation” and “number specification” respectively. The claim that the classifier head (Cl⁰) and the numeral head (Num⁰) merge with the noun separately is supported by the fact that “a classifier can occur alone without any numeral” in some classifier languages (e.g. Hmong, Vietnamese) and that some languages (e.g. Malay) allow classifiers to be omitted with vague numeral reference (i.e. when the individuation function is unnecessary) (Simpson 2005: 808-810). For the languages that allow bare classifier-NP sequences to occur without numerals and to be interpreted as definite NPs, Cl⁰-to-D⁰ movement is proposed (Simpson 2005: 823-826).

Y.-H. Audrey Li (1999) proposes a right-branching account and constructs a theory about grammatical number in (Mandarin) Chinese. According to Y.-H. A. Li (1999: 87, 91), the Classifier projection is structurally intermediate between Number and N, hence [Cl-N] constituency within the structure [D [Num [Cl [N]]]] of Mandarin Chinese NPs. This structure accounts for the distribution of the *-men* suffix, which is claimed to be a plural suffix in D: it can be suffixed to a common noun only when the DP contains no classifiers, because an intervening Cl⁰ will block the raising of the head noun to a higher position such as D⁰ (Y.-H. A. Li 1999: 81, 87-88). This theory also explains why the plural pronouns with

the *-men* suffix is compatible with numeral-classifier sequences in Mandarin⁸, namely that the pronouns are base-generated in D (ibid. 83). In addition, it accounts for the fact that nouns with the suffix *-men* are always definite in meaning, unlike bare nouns which can be definite or indefinite in Mandarin (ibid. 91). Curiously, Y.-H. A. Li (1999) treats the Num-Cl sequences in the examples as single words, presumably due to their prosodic proximity.

Likewise, Cheng and Sybesma (1999: 527) propose the right-branching structure [D [Numeral [Cl [NP]]]] for DPs with classifiers in Mandarin and Cantonese. After examining the distribution and interpretation of bare nouns, Cl-N phrases and Num-Cl-N phrases, they link the definite interpretation to the ι operator at Cl⁰, and the indefinite interpretation to “the presence of a [Numeral phrase]” (Cheng & Sybesma 1999: 521-523, 528). While bare Cl-N phrases can be definite in Cantonese, the Cl-N phrases can only have the indefinite reading in Mandarin, and definite bare nouns are used instead in Mandarin. Cheng & Sybesma (1999: 522-523) argue that the definite bare nouns in Mandarin are Classifier phrases as well, and that it is N-to-Cl movement that makes them definite in meaning despite the absence of any overt classifiers.⁹

It is worth noting that Cheng & Sybesma (1999) argue for the existence of a mass-count distinction in (Mandarin) Chinese, which is observable at the level of classifiers: the measure words (or “mass-classifier”, for mass nouns) are compatible with the modification marker *de* and with certain adjectives, while the (count) classifiers (for count nouns) are incompatible with these morphemes (Cheng & Sybesma 1999: 514-517). In other words, the difference between the count classifiers and measure words is both semantic (i.e. naming vs creating units) and syntactic in Mandarin according to them (Cheng & Sybesma 1999). However, it remains unclear whether the count classifiers and measure words differ from each other in terms of constituency and how common this distinction is among classifier languages.

⁸ Y.-H. Audrey Li (1999: 83) gives the same explanation for the use of DPs that contain a proper name and the suffix *-men*, which seems convincing for the dialect(s) of Mandarin in which this construction is acceptable. Personally, I find the [proper name – *men*] sequence infelicitous.

⁹ Note that this type of movement is unlikely (if not impossible) within the left-branching structure [[Num-Cl] NP].

While the unified accounts explain some patterns in Chinese and in other languages, according to some studies both types of the unified accounts are inadequate, and a few split accounts have been proposed, especially in the past two decades. According to the (typical) split accounts, the structure of quantified NPs is [[Num-Cl] N] under some conditions, and it can be [Num [Cl-N]] in other cases. There is no consensus on what factor affects the choice between the two structures, and different theories have been proposed: the split could be based on language, based on context, or based on sub-categories (Zhang 2013: 155-156; Her 2017A: 37-38). Some of these accounts are summarized below.

Little et al. (2022) argue that there are classifier-for-numeral languages in which the Num-Cl sequence is a constituent, and classifier-for-noun languages in which the Cl-N part of the phrase is a constituent. After an overview of the previous theories on the semantics of nouns, classifiers and numerals, they argue that classifiers can be required for different reasons in different theories. According to classifier-for-numeral theories, a numeral combines with a classifier first because the latter is the “extra semantic argument” (i.e. $\mu_{\#}$) that the numeral needs before it can merge with a noun (Little et al. 2022: 7-8). According to classifier-for-noun theories, a numeral “needs to combine with an atomic predicate”, but the nouns are not atomic; a classifier is therefore merged with the noun to “[generate] a set containing atoms” whose quantity can be specified (ibid. 8-9). Little et al. (2022) propose that the classifier-for-numeral theory works for classifier-for-numeral languages with [Num-Cl] constituents, and that the classifier-for-noun theory applies to classifier-for-noun languages with [Cl-N] constituents. In other words, there is no universal structure in Num-Cl-N phrases, and it is a parameter that different languages can set differently.

Specifically, they argue that Ch’ol (Mayan) is a classifier-for-numeral language, and that Shan (Kra-Dai) is a classifier-for-noun language. In Ch’ol, classifiers are obligatory with the lower numerals (from one to six) and are prohibited with numerals greater than six, which indicates that numerals can combine with classifiers in idiosyncratic ways that reflect lexical differences between numerals (Little et al. 2022: 20-22). Furthermore, Ch’ol classifiers are required in contexts where the noun is completely absent, such as counting and referring directly to the number, which shows that the numeral-classifier constituent occurs beyond nominal constructions (ibid. 25-26). On the other hand, the classifiers in Shan are obligatory with all numerals, but they can be omitted when combining with

certain nouns, which is evidence for the idiosyncratic merger of the classifier and the noun (ibid. 22-24). Moreover, Shan classifiers are employed in some contexts with no numerals (e.g. with demonstratives or relative clauses), and this indicates that the Cl-N combination occurs wherever the noun needs to be atomized, regardless of the absence of numerals (ibid. 24-25). In short, one can use the following criteria to determine if a classifier language has classifiers-for-numeral or classifiers-for-nouns: whether classifiers combine with numerals or nouns in idiosyncratic ways, and whether they are used in constructions containing no numerals or no nouns.

Li (2013) differentiates two types of interpretations of container measure words and argues that the syntactic structure of the NPs containing the MWs depend on their interpretation. The container MWs with the counting reading individuate nouns so that they can be counted, and the interpretation is often literal in the sense that the containers are physically present in the context. A container MW with the measure reading creates an abstract unit describing the quantity of the referent of the head noun¹⁰ (Li 2013: 129-132). In a Num-MW-N phrase with a counting reading, the measure word and the noun form a constituent, whilst the measure word merges with the numeral first in a phrase with a measure reading (Li 2013: 134). In Mandarin, there are a few constructions that require the counting reading to be available (e.g. the bare Cl-N phrases) and some contexts in which only the container MWs with a measure reading can occur (e.g. Num-MW-*de*-N¹¹ phrases) (ibid. 136-139). Using the features [$\pm$ C] (for the availability of the counting reading) and [$\pm$ M] (for the availability of the measure reading), Li (2013) distinguishes four types of classifiers (or MWs) in Mandarin. Of the [+C, -M] type are the individual classifiers, while the [-C, +M] type of MWs consists of “standard measures and temporary measures” (Li 2013: 144-148). The measure words of the container, partitive and group types are [+C,

¹⁰ There are a few tests that allow one to distinguish the counting reading from the measure reading. For example, distributive operator (e.g. *each*) should be compatible with NPs with the counting reading, but not with those that have a measure reading. There may also be morphological differences in some languages: English measure words are often compatible with the suffix *-ful* only when they have the measure reading (Li 2013: 131-132).

¹¹ The particle *de* is a modification marker in Mandarin, which links adjectives, possessor nouns and relative clauses to the noun they modify.

+M], and the kind classifiers have the [-C, -M] feature combination (ibid. 148-152). In other words, individual classifiers occur in a right-branching structure with [CL-N] constituency, and standard measures occur in left-branching DPs with [Num-CL] constituency, while the container MWs can be used in both structures (Li 2013).

There is some empirical evidence showing that these semantic differences that Li (2013) proposes are reflected in the syntax of measure words and classifiers. For instance, standard measures “must always be accompanied by a lexically realized Num”, which indicates a close connection between the numeral and the MW (cf. count classifiers which are often used without numerals) (Li 2013: 154). Likewise, since the modifier followed by *de* is arguably a constituent (Her 2017A), the fact that standard measures (unlike count classifiers) are compatible with this morpheme suggests that the feature [-C, +M] is linked to the left-branching structure (Li 2013: 155). However, according to Li (2013: 158-161), even individual classifiers could have the measure reading in certain contexts, and some standard measures could have a counting reading through coercion, because Mandarin has some type-shifting operations. This shows that the split account proposed by Li (2013) is fundamentally based on the interpretations of measure words and classifiers rather than based on categories.

Zhang (2013) proposes a split account based on the subcategories that the measure words and classifiers belong to. According to this account, the container MWs, standard measures, the collective MWs and partitive MWs are used in the left-branching structure, while the individual classifiers, the “individuating classifiers”¹² and kind classifiers occur in NPs with the right-branching structure (Zhang 2013: 170-171).

Four arguments in support of this claim are presented with Mandarin examples. First, a left-peripheral modifier (e.g. *chang* ‘long’) adjacent to the Num-MW sequence can co-occur with a modifier (e.g. *duan* ‘short’) in the core NP (which contains the head noun), even if the two modifiers are contradictory to each other (Zhang 2013: 158). This suggests that the two contradictory modifiers take “scope over separate constituents”, and that the left-peripheral modifier merges first with the Num-MW part of the phrase to the exclusion

¹² The “individuating classifier” is a special set of measure words that are used with a relatively small number of mass nouns (Zhang 2013).

of the core NP, hence the left-branching structure. Contradictory modifiers inserted into a Num-Cl-N¹³ phrase (e.g. with the count classifier *gen*) make the utterance ungrammatical, which indicates that classifiers occur in a different structure in Mandarin (Zhang 2013: 157-160). Second, a few adjectives can occur in the Num-Adj-MW/CL-N construction, and they semantically modify the head noun when the phrase contains a classifier. In a Num-Adj-MW-N phrase, however, the adjective modifies the measure word and does not affect the meaning of the noun. According to Zhang (2013: 160-163), the classifier c-commands the head noun in a Num-Adj-CL-N phrase, while the measure word in Num-Adj-MW-N phrases does not. Third, a noun in Mandarin s-selects the classifier(s) that can be used with it¹⁴, while these restrictions do not apply to measure words. Zhang (2013: 167) claims that the noun and the classifier that is s-selected are “in a local syntactic relation” (i.e. in the same constituent “excluding other elements”), and that nouns and measure words are not in this type of relation when they co-occur. Lastly, in a Num-Adj₁-CL-Adj₂-N phrase, if the first adjective is a shape adjective and the second one is a size adjective, then the utterance is infelicitous. However, if the two adjectives occur in that order in a Num-Adj₁-MW-Adj₂-N phrase, then the phrase is grammatical. An explanation for this difference is that Adj₁ c-commands Adj₂ in a Num-Adj₁-CL-Adj₂-N phrase, which has the right-branching structure, while Adj₁ does not c-command Adj₂ in a Num-Adj₁-MW-Adj₂-N phrase with Adj₁ being inside the [Num-Adj₁-MW] constituent (Zhang 2013: 167-169). Zhang (2013: 172-186) also summarizes and refutes some existing theories about constituency in quantified NPs (e.g. arguments based on Num-Cl adjacency).

¹³ Zhang’s (2013) choice of terms seems somewhat inconsistent, since the “collective classifiers” and “partitive classifiers” in fact behave like measure words rather than individual classifiers. The term “unit words” is also used in some places instead of *measure words*. These terms are changed here according to her theory, so that MWs are the ones occurring in the left-branching structure and that classifiers only occur in the right-branching structure.

¹⁴ The effect of semantic selection is the most obvious in individual classifiers since random Cl-N combinations are often completely unacceptable. Of the kind classifiers, some show a similar pattern of s-selection, while others (e.g. *zhong* ‘kind’ and *lei* ‘category’) do not. As for the “individuating classifiers”, some of them are only compatible with a single noun and seem idiomatic (Zhang 2013: 164-165).

Chinese classifiers and measure words have been categorized in a variety of ways. In the broadest sense, the “classifiers” include classifiers for verbs and compound classifiers (usually combinations of a classifier for noun and a classifier for verb) in addition to the classifiers for nouns (He 2008: 50). Classifiers for verbs and compound classifiers seem to behave rather differently from classifiers for nouns, and I focus on the classifiers for nouns in this thesis. Chao (1968: 584-608) divides the “measures” (i.e. classifiers for nouns) in Mandarin into six subcategories, namely “classifiers or individual measures” (such as the generic classifier *gè*), “group measures” (e.g. *shuāng* ‘pair’), “partitive measures” (e.g. *jié* ‘section’), “container measures” (e.g. *kuāng* ‘basket’), “temporary measures” (e.g. *shēn* ‘body’), and “standard measures” (e.g. *bàng* ‘pound’). He (2008) proposes similar classes of classifiers, with the additional category of “specific classifiers”¹⁵ that consists of individual classifiers that are only compatible with one or a very small number of nouns. The subcategories proposed in Chao’s (1968) book also form the basis of Li’s (2013) categorization of the classifiers (and MWs) in Mandarin. They are then re-classified into four categories according to their meaning: the classifiers that have a counting reading and no measure reading, the ones with a measure reading and no counting reading, the ones having both readings, and the ones with neither of the two readings (Li 2013: 143-152). Cheng and Sybesma (1999) distinguish two classes of classifiers, namely “count-classifiers” and “mass-classifiers”. The count-classifiers are used with count nouns such as *book* and *person*, while the mass-classifiers (or measure words) are mainly used with mass nouns (e.g. *soup*). They argue that these two types of classifiers are semantically different and that this difference “[reflects] the count/mass distinction” (Cheng and Sybesma 1999: 515). Zhang (2013) also proposes a dichotomy between measure words (occurring in the left-branching structure) and classifiers (occurring in the right-branching structure). According to this theory, MWs include standard measures and the unit words of the container, collective, and partitive types, while the individual, kind, and individuating classifiers are subtypes of classifiers. It is worth noting that the difference between measure words and classifiers is not linked to the count/mass distinction in nouns in Zhang (2013: 170-171).

¹⁵ The term is 专职量词 *zhuanzhi liangci* (literally ‘specific-role classifier’) in the original (He 2008: 36).

There are relatively few studies on the subcategories of Cantonese and Northern Wu unit words (i.e. classifiers and MWs). Matthews & Yip (2011: 114-123) classify Cantonese unit words into two groups, namely sortal classifiers and “measure (quantity) classifiers”, and the latter is further divided into collective classifiers, container classifiers and “generic classifiers” (i.e. kind classifiers). According to Zhu (2006: 71-76), there are four types of classifiers in Shanghai Wu, namely individual classifiers (e.g. the generic classifier *tsaq*), measures of containers, temporary classifiers (e.g. *mode* ‘sack’) and standard measures. Each of these four types of classifiers (or MWs) can be found in Ningbo Wu.¹⁶ Overall, the subtypes of the classifiers/MWs in Cantonese and Northern Wu seem comparable to those in Mandarin despite many lexical differences.

Regarding my research question about the history of Chinese classifiers, I propose that grammaticalization is a major factor both in the emergence of a classifier system and in the subsequent syntactic changes. I hypothesize that the individual classifiers in Middle Chinese, which are results of grammaticalization of the proto-classifiers in Old Chinese, were treated as a subtype of the pre-existing measure words and were used in the same way in the left-branching structure. Only later in the history did the classifiers start to differ from measure words in some contexts (esp. in Cantonese and Northern Wu), and the divergence is a sign of further grammaticalization of Chinese classifiers. I also hypothesize that the classifiers were initially used in the Num-Cl-NP construction, and that they spread to the other contexts later when they were further grammaticalized.

Aikhenvald (2000: 352-412) provides a typological overview of how classifiers arise, evolve and decay in many classifier languages, and a variety of topics are discussed, including the lexical sources of classifiers, grammaticalization, language contact and semantic changes. Grammaticalization (including those affecting classifiers) is considered unidirectional and can be accompanied by morphological and phonological changes (often reduction) (Aikhenvald 2000: 374-376). In addition, classifiers can undergo semantic extension, which makes the choice of classifiers less transparent (ibid. 407-408).

¹⁶ Shanghai Wu and Ningbo Wu are closely related dialects of Northern Wu. As a speaker of Ningbo Wu, I discuss patterns in this variety in more detail, although Shanghai Wu will also be mentioned in this thesis, because there is a larger literature about the classifiers in Shanghai Wu.

For the classifiers in Chinese, their historical development and relevant changes are summarized in Yang-Drocourt (2004) and in C. Zhang (2012). Yang-Drocourt (2004) shows that the orders Num-N, N-Num, and N-Num-Cl occurred in Old Chinese but later (mostly) fell out of use, and that the classifiers started to be used regularly in the demonstrative-Cl-N construction and in bare Cl-N phrases in Late Middle Chinese. C. Zhang (2012: 64-65, 68-69) points out that most Chinese classifiers did not emerge as repeater classifiers, and that a group of generic classifiers (including *měi*, *gè* and *zhī*) has existed in the Chinese classifier system throughout its history, even when some lexical items of this group were only used infrequently. C. Zhang (2012) treats lexical changes more thoroughly, but also discusses the history of the *měi*-Cl-N construction (with the meaning ‘every N’) and the classifiers for abstract nouns. Both books cover most of the recorded history of Chinese. A few other works focus more specifically on Chinese classifiers in certain time periods of the history, including J.-P. Li (2017) about the early classifiers before the 3rd century CE, Zheng (2016) covering the period from the 3rd to the 6th century, Wang (2018) covering the period from the late 6th century to mid-10th century, and Ye (2008) about the classifiers in an early form of Mandarin from the 14th century to the 17th century.

The hypotheses for my first research question will be discussed in chapters 2 and 3, and those for my second research question will be examined in chapter 4. We shall now proceed to a description of the methodology used in this study.

1.4 Methodology

The current thesis includes data from four types of sources, namely three corpora¹⁷ for Old Chinese and Middle Chinese, my own examples and judgments for Northern Wu and Mandarin, information elicited from native speakers for Cantonese and Northern Wu, and examples and comments found in previous works on these languages. This thesis does not include a comprehensive corpus study on Old Chinese and Middle Chinese, and I used the

¹⁷ The three corpora are *Academia Sinica Tagged Corpus of Old Chinese* (abbr. ASTC-OC), *Academia Sinica Tagged Corpus of Middle Chinese* (abbr. ASTC-MC) and *Academia Sinica Tagged Corpus of Early Modern Chinese* (abbr. ASTC-EMC). The last one includes a few texts in Late Middle Chinese (by my criteria).

corpora to find examples and to see if there are counterexamples to some generalizations. The elicitation process is explained in some more detail below.

For both Cantonese and Northern Wu, the main goal was to elicit grammaticality judgment data from native speakers. First, a questionnaire with 38 Cantonese sentences was created and made available online using SurveyMonkey, and the link was posted on social media together with a recruitment poster which also contains that link. Two friends helped me to circulate the poster so that it was visible to more Cantonese speakers. In total, 14 Cantonese speakers were recruited from February to June 2024, and seven of them offered valid responses to the questionnaire.¹⁸ A language background questionnaire and a consent form were incorporated into the online survey. In the main part of the online questionnaire, the participants were asked three types of questions, namely how acceptable a Cantonese sentence is, whether the English translation of it is accurate, and how it should be changed if the translation is inaccurate. For some sentences, the English translation was not directly provided, and the participants were asked to choose the meaning(s) of the sentence from a list of possible meanings. The 38 Cantonese sentences were shown in traditional Chinese characters that are usually used to write Hong Kong Cantonese, and the remaining parts of the survey, including the instructions and options, were in English. This was not problematic because the participants must have an intermediate-level knowledge of English to be recruited. When the participants had finished the survey, the incomplete responses were excluded, and the data were then exported for analysis.

The Northern Wu data were elicited in a similar way. A questionnaire for Ningbo Wu and a questionnaire for Shanghai Wu were created and were sent to the participants. The participants are one native speakers of Ningbo Wu and one native speaker of Shanghai Wu, and they were recruited and contacted using social media in May and June 2024.¹⁹ The

¹⁸ I consider a response valid when it satisfies two criteria. Firstly, the variety of Cantonese that the participant speaks natively is Hong Kong Cantonese or Guangzhou Cantonese (which are extremely similar). Secondly, the response is complete or virtually complete.

¹⁹ A reader commented that the sample sizes were small for Cantonese and Northern Wu. The number of Cantonese-speaking participants is in fact close to the number of participants that is often seen in studies on syntax. The sample size for Northern Wu is small since there are relatively few native speakers of that

participants received three files (i.e. a consent form, a language background survey and a questionnaire on the grammar of Ningbo/Shanghai Wu) by e-mail as I did not make online surveys for Northern Wu. The format of the Wu questionnaires is very similar to that of the Cantonese one, except for the language: the instructions are in Chinese, and the sentences in Northern Wu (containing classifiers or measure words) are shown in simplified Chinese characters.²⁰ The English translations of the sentences were not translated into Chinese, since the participants have adequate knowledge of English to understand them. They were invited to comment on the format of the survey at the end of it. The ethics application for this project was approved in June 2023. Further details of the data collection process will be mentioned in Chapter 3 when relevant. The Cantonese and Northern Wu questionnaires that were used for elicitation can be found in the appendix.

1.5 Outline of the thesis

The rest of this thesis consists of five chapters. Chapter 2 is about Mandarin classifiers and measure words, and I argue that they belong to one and the same category because they are syntactically very similar, and that they merge with the numeral first in left-branching structure. In Chapter 3, I examine the differences between the classifiers and MWs in Northern Wu and Cantonese to determine whether they belong to different categories and whether they also occur in the left-branching structure. Chapter 4 consists of a brief history of the classifiers in Chinese and a diachronic account for the convergence of classifiers and measure words. The implications of my claims are discussed in Chapter 5, and I will mention a few relevant issues, such as the optionality of classifiers and number marking in Chinese. Chapter 6 concludes the thesis.

language in Canada. An ideal follow-up study will include data from about 15 participants for each language, with surveys presented orally rather than in the written form. The current set-up was due to geographical and pandemic-related limitations when this study was designed.

²⁰ The scales of grammaticality in the Wu and Cantonese questionnaires are somewhat different, which is partly caused by the translation process. However, it should have been clear to all participants that there are four levels of acceptability.

2. The classifiers and measure words in Mandarin Chinese

Mandarin is a Sinitic language natively spoken in mainland China, Taiwan (ROC) and Singapore. With over one billion speakers, Mandarin is by far the most spoken Sino-Tibetan language. Standard Mandarin is the official language of China and Taiwan (ROC) and one of the official languages of Singapore. It is normally written in Chinese characters, which is usually transliterated into Latin alphabet using the Pinyin Romanization system. The basic clausal word order of Mandarin is SVO (Li 2013: 246), and the head noun is usually phrase-final in DPs (Huang et al. 2009: 298). Morphologically, Mandarin is a rather typical analytic language, with (mostly) monosyllabic morphemes forming multisyllabic words (Fábregas & Scalise 2012: 40-41). The classifier is the only noun categorization device in Mandarin, and there are no noun classes comparable to gender.²¹ Mandarin nouns do not show overt case or number inflection.

In terms of distribution, the individual classifiers in Mandarin are principally used in three contexts. They can follow the numeral in a Num-CL-N phrase, the demonstrative in a Dem-CL-N phrase, or just the verb in a bare CL-N phrase.

Measure words are arguably less homogeneous than classifiers, since a few groups of morphemes are subsumed under this category, notably “group classifiers”, “container classifiers” and “standard measures” (Li 2013: 21). They appear in the same positions as the (individual) classifiers in Mandarin.

Phonologically, individual classifiers are monosyllabic in Mandarin (and in Northern Wu and Cantonese), while measure words can be longer (with multiple syllables), or even morphologically complex.¹ The nominal suffixes *zi* and *er* can be added to certain measure words in colloquial Mandarin, although they can change the meaning to some extent with a

²¹ When written in Chinese characters, the third-person pronouns in Mandarin exhibit a three-way gender distinction, with 他 *tā* meaning ‘he’, 她 *tā* meaning ‘she’, and 它 *tā* meaning ‘it’. Since these pronouns share the same pronunciation, this distinction does not exist in spoken Mandarin.

¹ Morphologically complex classifiers (such as *ren-ci* ‘per person-time’) are found in written Mandarin and are derived from simplex classifiers through compounding. Complex classifiers are rare in colloquial Mandarin.

focus on entirety or produce some stylistic effects (He 2008: 63-73). Overall, measure words (esp. those denoting groups or parts) are also frequently monosyllabic. The multi-syllabic MWs can be morphologically simplex (e.g. *gongjin* 'kilogram'), created through derivation (e.g. *guo-zi* 'potful'), or through compounding (e.g. *cha-chi* 'teaspoon').

Lexically, the classifiers and measure words are two separate sets of morphemes, with only a handful of overlapping items (e.g. *ba* 'handful', also the classifier for *dao* 'knife', *yaoshi* 'key' and some other nouns). In this thesis, I gloss these potentially ambiguous morphemes according to their actual usage in the NP, without further explanation.

The next section examines the question whether the classifiers and measure words in Mandarin belong to two distinct grammatical categories or just one category.

2.1 CL and MW as one and the same grammatical category

To determine whether Mandarin classifiers and measure words belong to the same grammatical category, I start by describing their distribution. Both classifiers and measure words are frequently used in three types of NPs, namely in Num-CL/MW-N phrases, in Dem-CL/MW-N phrases, and in post-verbal bare CL/MW-N phrases. Some examples of these types of NPs are shown below:

- (1) 三 棵 树 [Mandarin]
san ke shu
three CL tree
'three trees'
- (2) 这 棵 树
zhe ke shu
this CL tree
'this tree'
- (3) 他们 想 种 棵 树。
tamen xiang zhong ke shu

they want plant CL tree
'They want to plant a tree.'

(4) 三 杯 水
san bei shui
three MW_{=cup} water
'three cups of water'

(5) 这 杯 水
zhe bei shui
this MW_{=cup} water
'this cup of water'

(6) 他们 想 喝 杯 水。
tamen xiang he bei shui
they want drink MW_{=cup} water
'They want to drink a cup of water.'

The first three examples illustrate that the classifier *kē* can be used in each of the abovementioned positions (viz. Num-CL-N in (1), demonstrative-CL-N in (2) and the bare CL-N phrase in (3)). Examples (4) to (6) show that the measure word *bēi* can be used in the same positions as well (namely in a Num-MW-N phrase in (4), in a demonstrative-MW-N phrase in (5) and in the post-verbal bare MW-N phrase in (6)).² In addition, deleting the classifiers and measure words has similar effects in the examples above. Deleting the CL/MW in (1) and (4) makes both phrases unacceptable, while deleting the CL/MW in (3) and (6) leads to the same semantic change, namely making both NPs generic. It is worth noting that the sequences CL-MW and MW-CL are unattested in Mandarin, which indicates that classifiers and measure words are “mutually exclusive” (Her 2017A: 38).

Furthermore, the classifiers and measure words are also employed in some other constructions, and once again they appear in the same positions while being mutually

² Note that the demonstrative-CL/MW-N phrases and bare CL/MW-N phrases must be interpreted as singular. To make their meanings plural, a numeral or the quantifier *jǐ* 'several' can be inserted in front of the CL/MW.

exclusive. For instance, they can co-occur with the quantifying determiner *mei* ‘every’ and with the interrogative word *na* ‘which’, as shown in the examples below:

(7) 每 棵 树

mei *(ke) shu

every CL tree

‘every tree’

(8) 哪 棵 树

na *(ke) shu

which CL tree

‘Which tree?’

(9) 每 杯 水

mei *(bei) shui

every MW_{=cup} water

‘every cup of water’

(10) 哪 杯 水

na *(bei) shui

which MW_{=cup} water

‘Which cup of water?’

Examples (7) and (8) indicate that classifiers (*kē* in this case) are consistently used with *mei* ‘every’ and *na* ‘which’ respectively, in order for them to merge with NPs. Examples (9) and (10) show that measure words must be inserted for *mei* ‘every’ and *na* ‘which’ to modify uncountable nouns such as *shui* ‘water’.

In addition to having the same distribution in various constructions, Mandarin classifiers and measure words also share the same syntactic behaviour in three syntactic tests. These tests include extraction, ellipsis, and coordination. As Her (2017A: 39) points out, N-extraction from Num-Cl/MW-N phrases are acceptable, while the extraction of the Cl/MW-N part of the phrase is illicit. This is shown in the examples below:

- (11) 树 只 有 三 棵 吗?
 shu zhi you san ke (~~shu~~) ma
 tree only have three CL t Q
 ‘Are there only three trees?’³
- (12) 水 只 有 三 杯 吗?
 shui zhi you san bei (~~shui~~) ma
 water only have three MW=_{cup} t Q
 ‘Are there only three cups of water?’

Example (11) illustrates that the noun *shu* ‘tree’ can be fronted to the beginning of the sentence as the topic. Fronting the classifier *ke* with the noun would make the sentence ungrammatical. Example (12) shows that *shui* ‘water’ can also be topicalized and moved to the left edge of the sentence. Similarly, fronting the measure word *bei* ‘cup’ with the noun would render the sentence illicit. A similar pattern is found in Dem-Cl/MW-N phrases, from which the noun can be extracted but the Cl/MW-N part cannot. This is exemplified below:

- (13) 树，还 是 这 棵 最 高。
 Shu hai shi zhe ke zui gao
 tree PARTICLE COP this CL SUPERL tall
 ‘As for trees, this is the tallest one.’
- (14) 咖啡，还 是 这 杯 最 好喝。
 kafei hai shi zhe bei zui haohe
 coffee PARTICLE COP this MW=_{cup} SUPERL tasty
 ‘As for coffee, this cup of coffee is the tastiest.’

³ A reader asks whether this pattern of N-extraction could be the result of topicalization and deletion of the lower copy of the noun. That analysis is indeed possible for (11). However, if the classifier-noun combination is more idiom-like (e.g. *yi-tiao-haohan* ‘a brave man’, with the classifier *tiao* normally used with non-human nouns), N-extraction is the better analysis as no other noun can occur in the position from which the noun is extracted. Crucially, N-extraction can occur in *yi-tiao-haohan* ‘a brave man’, but N-ellipsis cannot, which indicates that they are not the same syntactic operation.

Examples (13) and (14) show that the head noun (*shu* ‘tree’ or *kafei* ‘coffee’) in Mandarin NPs can be topicalized and fronted. Neither the classifier or the measure word can follow the head noun and move the beginning of the sentence.

Interestingly, the head noun in a post-verbal bare Cl/MW-N phrase cannot undergo the same type of topic fronting as in examples (11) to (14). Two examples of this pattern are shown below:

- (15) 树，他们 种了 棵
*shu tamen zhong-le ke (~~shu~~)
tree they plant-ASP CL *t*
Intended: ‘As for trees, they planted one.’
- (16) 水，他们 喝了 杯
*shui tamen he-le bei (~~shui~~)
water they drink-ASP MW=_{cup} *t*
Intended: ‘As for water, they drank a cup.’

Example (15) illustrates that the noun cannot be extracted from a post-verbal bare Cl-N phrase. Example (16) shows that N-extraction from a bare MW-N phrase is also illicit. Here, we see a general tendency in Mandarin for classifiers and measure words to be linking elements rather than standalone morphemes stranded in VPs.

In short, head nouns can be extracted from Num-Cl/MW-N phrases and Dem-Cl/MW-N phrases, but they cannot be extracted from bare Cl/MW-N phrases. Extraction of the Cl-N or MW-N part from a larger NP is never acceptable. Crucially, Mandarin classifiers and measure words always share the same syntactic behaviour in the extraction test.

Likewise, Mandarin classifiers and measure words are virtually indistinguishable from each other in the ellipsis test. First, in Num-CL/MW-N phrases, N-ellipsis is felicitous with appropriate contexts, while the ellipsis of the Cl-N or MW-N segment is not (Her 2017: 39). This can be observed in the examples below:

- (17) 我 买了 三 本 书， 他 买了 五 本。
 wo mai-le san ben shu, ta mai-le wu *(ben)
 I buy-ASP three CL book, he buy-ASP five CL
 'I bought three books, and he bought five.'
- (18) 我 买了 三 瓶 水， 他 买了 五 瓶。
 wo mai-le san ping shui, ta mai-le wu *(ping)
 I buy-ASP three MW=_{bottle} water, he buy-ASP five MW=_{bottle}
 'I bought three bottles of water, and he bought five (bottles).'
- (Li 2013: 40: (45) and (46), with modifications)

In example (17), the noun *shū* 'book' undergoes ellipsis in the second clause, but the classifier *ben* in the same phrase cannot. In example (18), the noun *shui* 'water' is omitted in the second clause, while the measure word *ping* 'bottle' cannot be omitted. Note that Mandarin differs from English in this respect: in the corresponding sentence, dropping the word *bottle* is perfectly acceptable in English (shown in the translation of (18)), and is even the preferred option for some native speakers (Li 2013: 40).⁴ Despite the possibility of setting this parameter differently, Mandarin classifiers and measure words behave identically in this regard.

Second, N-ellipsis is allowed in Dem-CL/MW-N phrases, while the ellipsis of the CL-N or MW-N sequence is again unacceptable. Some examples of this are given below:

- (19) 我 买了 这 本 书， 他 买了 那 本。
 wo mai-le zhe ben shu, ta mai-le na *(ben)
 I buy-ASP this CL book he buy-ASP that CL
 'I bought this book, and he bought that one.'
- (20) 他 买了 那 瓶 酒， 我 买了 这 瓶。

⁴ I assume that phrasal modifiers can license N'-ellipsis. This assumption seems necessary because N'-ellipsis is possible even when the Num-MW-N phrase (e.g. *san-sheng-shui* 'three litres of water') has an unambiguous measure reading, and the [Num-MW] constituency in that context is largely undisputed (Li 2013, Zhang 2013, Her 2017A, among others).

ta mai-le na ping jiu wo mai-le zhe *(ping)
 he buy-ASP that MW=_{bottle} wine I buy-ASP this MW=_{bottle}
 'He bought that bottle of wine, and I bought this.'

Examples (19) and (20) illustrate that neither the classifier nor the measure word can be deleted with the noun in Mandarin NPs, even when there is a proper context for N-ellipsis. The demonstratives *zhè* 'this' and *nà* 'that' can be independent words in other contexts, and it is logically possible for the classifiers and measure words to undergo ellipsis here. However, that is not allowed in the syntax of Mandarin, and the classifiers and measure words share the same syntactic behaviour again.

The pattern of N-ellipsis is different and somewhat unexpected in bare Cl/MW-N phrases, because the head noun is not readily omissible even with proper contexts. While N-ellipsis does not make the sentences completely ungrammatical, they are considerably less acceptable with ellipsis. This is shown in the examples below:

- (21) ?我买了 本 书，他也 买了 本。
 ? wo mai-le ben shu, ta ye mai-le ben
 I buy-ASP CL book he also buy-ASP CL
 Intended: 'I bought a book, and he bought one too.'
- (22) ?我喝了 杯 水，他也 喝了 杯。
 ? wo he-le bei shui, ta ye he-le bei
 I drink-ASP MW=_{cup} water he also drink-ASP MW=_{cup}
 Intended: 'I drank a cup of water, and he drank a cup of water too.'

In example (21), the noun *shū* 'book' is deleted in the second CP, which makes the sentence only marginally acceptable. It is not very difficult for a native speaker to guess its meaning, but uttering this sentence creates a feeling of incompleteness. One could repair this sentence by inserting the numeral *yī* 'one' before the classifier, or simply by not deleting the noun. The situation is very similar in (22), where the omission of the noun *shuǐ* 'water' in the second clause makes the sentence awkward. This sentence also sounds

incomplete, and the same repair strategies can be used to make it fully grammatical. In both cases, the classifier or measure word could also be deleted together with the noun. Doing so, however, would change the meaning of the second CP, and the quantity of the noun (namely *shu* ‘book’ or *shui* ‘water’) would become unspecified.

In short, head nouns in Num-Cl/MW-N phrases and Dem-Cl/MW-N phrases can undergo ellipsis with appropriate contexts. N-ellipsis in bare Cl/MW-N phrases are much less acceptable. Ellipsis of the Cl-N or MW-N sequence either makes the utterance illicit or changes its meaning. Again, the classifiers and measure words behave identically in the ellipsis test.

Mandarin classifiers and measure words also exhibit very similar syntactic patterns when used with coordinating conjunctions. With N-ellipsis or N-extraction, more than one Num-Cl/MW phrase can merge into one phrase through conjunction, which is perfectly grammatical in Mandarin (Her 2017: 38). This can be seen in the examples below:

- (23) 约翰 和 玛丽 分别 有 三 本 和 五 本 书。
Yuehan he Mali fenbie you san ben he wu ben shu
John and Mary respectively have three CL and five CL book
‘John and Mary have three books and five books respectively.’
- (24) 有 四 桶 或者 五 桶 水 就 足够 了。
you si tong huozhe wu tong shui jiu zugou le
have four MW=_{bucket} or five MW=_{bucket} water then enough SFP
‘If there are four or five buckets of water, then it is enough.’
- (25) 苹果 就 买 十 个 或者 一 袋 吧。
pingguo jiu ma shi ge huozhe yi dai ba
apple just buy ten CL or one MW=_{bag} SFP
‘Just buy ten apples or a bag of them.’

Example (23) illustrates that two Num-Cl phrases can merge into a single phrase with the help of the conjunction *he* ‘and’. Example (24) shows that two Num-MW phrases can merge into a single phrase using appropriate conjunctions. Example (25) indicates the

possibility of mixing and matching Num-CL phrases and Num-MW phrases. It is worth noting that the first classifier *ben* in (23) cannot be omitted, and the same is true of the measure word *tong* in *si-tong* ‘four buckets’ in (24).

Two or more Dem-CL/MW phrases can also merge into a single phrase using suitable conjunctions. For instance, *zhè-CL-háishì-nà-CL* ‘this one or that one’ is an expression frequently used in context of shopping, and the noun in question tends to be omitted since it is often obvious from the discourse context. The classifier in this expression can also be replaced by a measure word, especially when the noun that is being discussed is uncountable, such as drinks.

The situation is quite different in post-verbal bare CL/MW-N phrases. Combining two CL-N phrases or two MW-N phrases using conjunctions such as ‘and’ and ‘but’ is in fact illicit. This pattern is shown in the examples below:

(26) 我 买了 本 小说 和 支 铅笔

*wo mai-le ben xiaoshuo he zhi qianbi

I buy-ASP CL novel and CL pencil

Intended: ‘I bought a novel and a pencil.’

(27) 他 喝了 杯 茶 和 碗 汤

*ta he-le bei cha he wan tang

he drink-ASP MW_{=cup} tea and MW_{=bowl} soup

Intended: ‘He drank a cup of tea and a bowl of soup.’

Examples (26) and (27) illustrates that neither the CL-N-and-CL-N sequence nor the MW-N-and-MW-N sequence is allowed, even though each of the bare CL-N phrases (or bare MW-N phrases) would be perfectly grammatical in that position if the other one did not co-occur with it. The simplest way to repair these ungrammatical would be to insert a numeral *yi* ‘one’ before the second classifier (or measure word). It is curious that the “lopsided” coordination CL-N-and-*yi*-CL-N is preferable to the alternative CL-N-and-CL-N, which is more balanced and logically plausible. Even though this pattern is largely unexpected, it affords

direct evidence that the classifiers and measure words share the same syntactic behaviour in this type of NPs as well.

Cheng and Sybesma (1998) argue in favour of distinguishing “count classifiers” from “mass classifiers” and propose two syntactic criteria for differentiating the two categories.⁵ They base their arguments on two syntactic rules that they generalized from some Mandarin data, namely their different behaviour in Num-Adj-Cl-N phrases and in Num-Cl-*de*-N phrases. Cheng and Sybesma (1998) claim that only the “mass classifiers” can appear in the Num-Adj-Cl-N construction, and that only the “mass classifiers” are compatible with the modification marker *de*, which is optional in Num-Cl-*de*-N phrases.

These arguments have been examined more thoroughly by Li (2013), who shows that the claims are not supported by convincing empirical evidence. According to Li (2013: 57-58), nearly half of the individual classifiers can appear in Num-Adj-Cl-N phrases, and probably even more of them can if appropriate contexts are available.⁶ Furthermore, at least one entire subcategories of the mass classifiers (standard measures) is incompatible with the Num-Adj-Cl-N construction. The test based on the compatibility between “mass classifiers” and *de* also turned out to be unreliable: some mass classifiers are incompatible with *de*, while count classifiers can occur in Num-Cl-*de*-N phrases if the numeral is large, approximative or a fraction (Li 2013: 63-66). Even more problematically, some classifiers are expected to be “mass” according to one criterion and “count” according to the other (ibid. 66). In short, neither of the two criteria proposed by Cheng and Sybesma (1998) is effective in distinguishing individual classifiers from measure words.

2.2 A continuum from individual classifiers to standard measures: phonological, morphological, semantic, and lexical differences between CL and MW

⁵ Broadly speaking, the term “count classifier” corresponds to “(individual) classifier”, while “mass classifier” corresponds to “measure word”.

⁶ The interpretation of Num-Adj-Cl-N phrases and Num-Adj-MW-N phrases can be quite different, but that has to do with the lexical semantics of measure words. This matter is discussed further in the next section.

In the previous section, I have shown that classifiers and measure words share the same distribution and behave similarly in three syntactic tests. In fact, arguments for making the distinction between classifiers and measure words are often based on phonological, morphological, semantic and/or lexical patterns rather than direct syntactic evidence. In this section, I argue that these non-syntactic differences between classifiers and measure words are either idiomatic/unsystematic or attributable to non-syntactic factors. In other words, those differences do not reflect a syntactic distinction between two grammatical categories.

Another problem with a simple dichotomy between classifiers and measure words is that it fails to account for differences among the different subcategories of measure words. In general, the “group classifiers”, “partition classifiers” and “container classifiers”⁷ share very similar distribution and syntactic behaviour and can be subsumed under one broader subcategory (Group-Partition-Container measure words, or G-P-C MWs), which is the bulk of measure words. The other two subcategories, namely the standard measures and “temporary classifiers” behave rather differently from G-P-C MWs and somewhat differently from each other. Some examples of these subcategories of MWs are given below:

- (28) 一 束 花 这 束 花
 yi shu hua zhe shu hua
 one MW=_{bouquet} flower this MW=_{bouquet} flower
 ‘a/one bouquet of flower’; ‘this bouquet of flower’
- (29) 两 片 面包 那 片 面包
 liang pian mianbao na pian mianbao
 two MW=_{slice} bread that MW=_{slice} bread
 ‘two slices of bread’; ‘that slice of bread’
- (30) 三 瓶 水 这 瓶 水
 san ping shui zhe ping shui
 three MW=_{bottle} water this MW=_{bottle} water

⁷ These are the terms used by Chao (1968), but they are “massifiers” (i.e. measure words) rather than “count-classifiers” by Cheng & Sybesma’s (1999) criteria.

- (31) ‘three bottles of water’; ‘this bottle of water’
 一 身 雪 那 身 雪
 yi shen xue *na shen xue
 one MW_{=body} snow that MW_{=body} snow
 ‘a body-ful of snow’; Intended: ‘that body-ful of snow’
 (Li 2013: 19: example (11a), modified and expanded)
- (32) 五 公斤 牛肉 这 公斤 牛肉
 wu gongjin niurou *zhe gongjin niurou
 five kilogram beef this kilogram beef
 ‘Five kilograms of beef’; Intended: ‘this kilogram of beef’

Examples (28) to (30) are instances of group MWs, partition MWs and container MWs respectively. These three examples show that they can all be used in Num-MW-N phrases as well as in the Dem-MW-N construction. In fact, measure words of these three subcategories can also be used in post-verbal bare MW-N phrases and in Num-Adj-MW-N phrases, and they also behave very similarly in various other construction (e.g. with N-ellipsis or N-fronting). I shall, therefore, assume that they are essentially one type of measure words (G-P-C MWs).

Example (31) shows that temporary MWs appear in Num-MW-N phrases, but they cannot appear in Dem-MW-N phrases. The same is true for standard measures, which can be observed in (32). Unlike the G-P-C MWs, temporary MWs and standard measures can only be used in Num-MW-N phrases. It is worth noting that temporary MWs are only compatible with the numeral *yi* ‘one’ but not with any other numeral (He 2008: 38), and that some temporary MWs are not actively used by all native speakers.⁸ The distribution of these five types of measure words is summarized in the table below and compared to the distribution of individual classifiers.

⁸ For instance, He (2008: 38) gives the example 一眼荒凉 *yi-yan-huangliang* ‘an eyeful of desolation’, with a temporary MW which I do not personally use. This indicates that temporary MWs are idiom-like.

Category construction	Num-Cl/MW-N	Dem-Cl/MW-N	Bare Cl/MW-N	Num-Adj-Cl/MW-N ⁹
Group MWs	G	G	G	G
Partition MWs	G	G	G	G
Container MWs	G	G	G	G
Temporary MWs	% (G if Num=1)	*	*	*
Standard measures	G	*	*	*
Individual classifiers	G	G	G	% (24 of 54 classifiers)

(Li 2013; He 2008)

Overall, the distribution of the group-partition-container measure words is very similar to that of the individual classifiers in Mandarin. The distribution of temporary MWs and standard measure are more different from the classifiers and the other measure words. However, these five subtypes of measure words also have some properties in common (though outside the domain of syntax), which will be discussed in the next subsections.

The rest of this section is divided into four subsections, about the characteristics of measure words and classifiers in the domains of phonology, morphology, semantics, and lexicology respectively.

2.2.1 Phonological similarities and differences between classifiers and MWs

Phonologically, the distinction between classifiers and measure words is not entirely groundless, because there is a clear constraint on the length of the former but not on the latter. Individual classifiers are monosyllabic in Mandarin, whilst measure words can be monosyllabic, disyllabic, or trisyllabic. In spoken Mandarin, over 60% of the words

⁹ Only a handful of adjectives denoting size or shape can appear in this construction. Abbreviations: G = grammatical; * = ungrammatical; % = grammatical under certain conditions.

are disyllabic (Myers 2022: 54). On the other hand, the most frequent words in Mandarin are mostly monosyllabic (including many high-frequency verbs and functional particles), and longer words usually consist of monosyllabic morphemes (Huang et al. 2022: 76). The fact that classifiers must be monosyllabic is illustrated in the sentence below:

- (33) 他有 两 栋 房子、一 辆 车 和 三 头 牛。
 ta you liang dong fangzi, yi liang che he san tou niu
 he have two CL house one CL car and three CL cattle
 ‘He has two houses, one car, and three head of cattle.’

The three individual classifiers in the sentence above (viz. *dong*, *liang* and *tou*) are all monosyllabic, and the same can be said of all the other classifiers listed in Chao (1968: 589-593). They are also usually (but not always) monomorphemic, and their morphology is discussed in the next subsection. Unlike individual classifiers, measure words can be disyllabic or longer, and this is true for each subtype of measure words. Some of the longer measure words are shown below:

- (34) 一 嘟噜 葡萄
 yi dulu putao
 one MW_{=bunch} grape
 ‘a bunch of grapes’
 (He 2008: 243)
- (35) 一 部份 人民
 yi bufen renmin
 one MW_{-part} people
 ‘a fraction of the people’¹⁰
 (Chao 1968: 598)

¹⁰ This partitive MW rarely co-occurs with other numerals and is thus comparable to the temporary MWs. Most other partitive MWs are monosyllabic (Chao 1968).

- (36) 一 卡车 苹果
 yi kache pingguo
 one MW_{-truck} apple
 ‘a truckload of apples’
- (37) 一 鼻子 灰
 yi bizi hui
 one MW_{=nose} dust
 Literally: ‘a nose of dust’; idiomatic meaning: ‘being refused’
 (Li 2013: 19: example (11b), modified)
- (38) 一 立方米 天然气
 yi lifangmi tianranqi
 one MW_{=cubic.metre} natural.gas
 ‘one cubic metre of natural gas’

Examples (34) to (36) illustrate that group-partition-container measure words can be disyllabic. In these subcategories, monosyllabic MWs are still more common than longer ones (Chao 1968). Example (37) shows that temporary measure words can also be longer than one syllable. In this subcategory, monosyllabic and longer MWs are about equally common, and the nominal suffix *-zi* is found in a few of the longer MWs (Chao 1968: 603; He 2008: 245). As for the standard measures (seen in (38)), they are also frequently disyllabic or polysyllabic. This is attributable to the fact that this subcategory of MWs contains many compounds and loanwords (He 2008: 247).

Although classifiers and measure words differ in terms of maximum length, one must not assume that they always behave differently in the domain of phonology. In fact, they exhibit very similar behaviour when they undergo tone sandhi in Mandarin.

There are four lexical tones in Standard Mandarin, including one level tone and three contour tones. Apart from some short and unstressed syllables which have the

neutral tone, each syllable has one of these four tones (Chao 1968: 25-26, 35). They are shown in the table below:

Tone number	Name	Description	Pitch	Pinyin
1 st	<i>Yinping</i>	High-level	55	ā
2 nd	<i>Yangping</i>	Mid-rising	35	á
3 rd	<i>Shangsheng</i>	Falling-rising	214	ǎ
4 th	<i>Qusheng</i>	High-falling	51	à

(Chao 1968: 26; modified)

Note that the 3rd tone (or Tone 3) has the most complex contour among the four tones in Mandarin. When appearing in the non-final position of a phrase, Tone 3 undergoes tone sandhi, which is a phonological process that turns the underlying lexical tone into a different tone at the surface level; in this case, the tone sandhi process is attributable to the “shorter durations [of nonfinal positions] that forbid full articulation” (Wee 2022: 313). The tone sandhi pattern of Standard Mandarin Tone 3 is summarized as follows:

- A. /214/ → [14] /__σ_{/214/} (An underlying Tone 3 becomes a rising tone when the syllable is immediately followed by another Tone-3 syllable.)
 - B. /214/ → [21] /__σ (When followed by a syllable with any other tones, Tone 3 turns into a low falling tone.)¹¹
- (Based on Wee 2022: 313)

Both tone sandhi patterns simplify the contour of Tone 3 in non-final positions. The effect of the tone sandhi process is shown in the examples below:

(39) 海 水
 hai shui

¹¹ Abbreviations: T3R = the rising variant of non-final Tone 3 = [14]; T3F = the falling variant of Tone 3 = [21].

- (40) 214→14 214
 sea water
 ‘Sea water’
 海 风
 hai feng
 214→21 55
 sea wind
 ‘Sea breeze’

In a phrase with three consecutive Tone-3 syllables, there are two tone sandhi patterns, which are then morpho-syntactically conditioned (Wee 2022). The two outcomes are [21-14-214] (or T3F-T3R-T3) and [14-14-214] (or T3R-T3R-T3), and they are illustrated in the examples below:

- (41) 小 水 井
 xiao shui jing
 21 14 214
 small water well
 ‘Small well’
- (42) 紫 薯 粉
 zi shu fen
 14 14 214
 purple sweet-potato powder
 ‘Powder of purple sweet potato’

In (41), the morphemes corresponding to the second syllable and the third syllable merge first to the exclusion of the first morpheme, and the tone pattern on the surface is [21-14-214]. In (42), the first two morphemes (corresponding to the first two syllables) merge first, and the tone pattern of the phrase becomes [14-14-214] instead. In other words, the surface tone pattern of a trisyllabic Tone-3 phrase is dependent on its morpho-

syntactic structure. Although the discussion and the examples here are about NPs, the same link between morphosyntactic structure and Tone 3 sandhi patterns is also found other types of phrases.

Being aware of this link between morphosyntax and Tone 3 sandhi patterns, we can use the surface tone patterns as an indicator or a diagnostic for the morphosyntactic structure of other NPs, such as those with classifiers or measure words. It is logically possible that Num-Cl-N phrases and Num-MW-N phrases behave differently in Tone 3 sandhi, if classifiers and measure words are two distinct categories associated with different syntactic structures. In fact, theories along those lines have been proposed for Japanese classifiers and measure words (i.e. Cl-MW “split account”) (Her 2017: 37-38). I shall start with some examples of Num-MW-N phrases, in which the morphemes all have lexical Tone 3 (underlyingly). The third line in the gloss shows the surface tone pattern.

- (43) 九 桶 水
 jiu tong shui
 14 14 214
 nine MW=_{bucket} water
 ‘nine buckets of water’
- (44) 两 捆 草
 liang kun cao
 14 14 214
 two MW=_{bundle} hay
 ‘two bundles of hay’
- (45) 五 斗 米
 wu dou mi
 14 14 214
 five Chinese.peck rice
 ‘five (Chinese) pecks of rice’¹²

¹² *Dou* is a traditional Chinese measure of capacity, and one *dǒu* is approximately ten litres.

Examples (43) to (45) show that Tone-3 Num-MW-N phrases always exhibit the tone pattern [14-14-214] (or T3R-T3R-T3) in their surface representation, which is the same as the tone pattern observed in (42). It is not affected by the subcategory of the MWs, since the same tone pattern is seen with the container MW *tǒng* ‘bucket’, the group MW *kǔn* ‘bundle’, and the standard measure *dǒu* ‘Chinese peck’. Now observe how tone sandhi unfolds in Num-Cl-N phrases with three lexically Tone-3 syllables:

- (46) 两 把 尺
 liang ba chi
 14 14 214
 two CL ruler
 ‘two rulers (stationery)’
- (47) 五 口 井
 wu kou jing
 14 14 214
 five CL well
 ‘five wells’
 (Xia 2021: (27) and (28), modified)

In (46) and (47), the underlying tone pattern (T3-T3-T3) undergoes the same kind of tone sandhi as in Tone-3 Num-MW-N phrases and becomes [14-14-214] (T3R-T3R-T3) on the surface. In other words, the tone sandhi pattern of Tone-3 Num-MW-N phrases and that of (most) Tone-3 Num-Cl-N phrases are the same as the tone sandhi pattern in example (42) *zi-shu-fen* ‘powder of purple sweet potato’, where the first two morphemes (corresponding to the first two syllables) merge first. This suggests that the numeral and the classifier or measure word in a Num-Cl/MW-N phrase also merge first to form a constituent to the exclusion of the noun. More concrete evidence supporting this claim is presented in the next section.

There is an interesting exception to the general Tone-3-sandhi rule in trisyllabic Tone-3 Num-Cl-N phrases, namely the special tone sandhi pattern of the classifier *pǐ*, which is mostly used for counting horses, and occasionally with mules and wolves (Chao 1968). A trisyllabic Tone-3 Num-Cl-N phrase is shown below:

- (48) 五 匹 马
 wu pi ma
 21 55 214
 Five CL horses
 ‘five horses’
 (Xia 2021: (30), modified)

In (48), the tone pattern is unusual not only because it is different from that of most other Tone-3 Num-Cl-N phrases, but also because the tone pattern [21-55-214] is not the regular outcome of any systematic tone sandhi process.¹³ In fact, this special pronunciation (*pǐ*) of the classifier *pǐ* is found in Num-Cl phrases (with N-ellipsis) where the phrase-final Tone 3 can normally be preserved as such. This special tone is also heard when the classifier *pǐ* follows non-Tone-3 numerals (e.g. *sān* ‘three’) and demonstratives (e.g. *zhè* ‘this’). In other words, this classifier has two variants differing in their tones, namely *pī* with Tone 1 and *pǐ* with Tone 3.¹⁴ The effect of this variation happens to manifest itself in trisyllabic Num-Cl-N phrases as well, but the variation per se does not originate from Tone-3 sandhi, or from any other synchronic phonological process.

The origins of the tone variation of this classifier seem to be the loss of Middle Chinese “Entering Tone” in the northern dialects (including Standard Mandarin); as the “Entering Tone” disappeared and merged with the other tones, irregular mergers created doublets and caused tone variation in many morphemes that had the “Entering Tone” (Chao 1968: 25). For instance, similar contextual or idiolectal tone variation is found in the

¹³ The regular tone pattern [14-14-214] is also heard, though it seems to be less common than the irregular tone pattern shown in (48). Note that [55] is Tone 1 in Standard Mandarin.

¹⁴ When pronounced in isolation or used in idioms, *pǐ* is the normal pronunciation of this morpheme.

morphemes meaning ‘one’, ‘eight’, ‘not’, ‘blood’, ‘bone’, ‘answer’ and ‘colour’¹⁵ in Standard Mandarin. The classifier *pǐ* is arguably one of these morphemes that used to be pronounced with the “Entering Tone”¹⁶ and developed tone variation due to sound change.

In short, Mandarin classifiers and measure words behave alike in Tone-3 sandhi, and there is some evidence that they merge with numerals first. The special tone sandhi pattern of the classifier *pǐ* has its origins elsewhere (in sound change rather than tone sandhi) and is not indicative of any difference between classifiers and measure words. To sum up this subsection, the classifiers and measure words in Mandarin differ from each other in terms of maximum length, but they exhibit the same tone pattern in trisyllabic Num-Cl/MW-N phrases that undergo Tone-3 sandhi.

2.2.2 Morphological similarities and differences between classifiers and MWs

Morphologically, the classifier and measure words in Mandarin can be classified into four categories, namely individual classifiers, group-partition-container measure words, temporary measure words and standard measures. Overall, the morphology of the classifiers is simpler than that of measure words.

Classifiers are usually employed in their bare form with no additional morphemes in Standard Mandarin. The suffix *-r*, which has a diminutive or affectionate meaning and often attaches to nouns, can be added to a handful of the individual classifiers. It is the only affix that can co-occur with classifiers, and it is always optional and non-syllabic when used with classifiers. Seven of the 51 classifiers listed in Chao (1968) are compatible with this suffix. A Num-Cl-N phrase containing this suffix is shown below:

- (49) 一 家儿 人家儿
 yi jia-r renjia-r
 one CL-DIM household-DIM

¹⁵ In Chinese characters, these morphemes are: 一, 八, 不, 血, 骨, 答, and 色 respectively.

¹⁶ In Cantonese and Northern Wu, the cognates of this classifier are still pronounced with short, checked tones.

‘a household’¹⁷

(Chao 1968: 590)

In (49), the diminutive suffix appears between the classifier and the head noun. This suffix is more frequently used in colloquial Beijing Mandarin than in Standard Mandarin. Note that the complex classifier *jia-r* in (49) is still monosyllabic.

The morphology of measure words tends to be somewhat more complex. Group MWs, partition MWs and container MWs have similar morphological behaviour and are subsumed under one subcategory (G-P-C MWs). The diminutive suffix *-r* can be added to many of these measure words, and for a dozen of them, this suffix is obligatory (Chao 1968: 599-602). In some MWs, the diminutive meaning of *-r* is transparent, while in others it can be quite vague (He 2008: 70-71). The suffix *-zi*, which is also a frequent nominal suffix in Mandarin, can also be added to some of the G-P-C MWs, and is sometimes obligatory (Chao 1968); this suffix is always unstressed with no lexical tone. When these two suffixes are optional, adding them to MWs often makes the MWs more vivid and colloquial (He 2008). Examples of morphologically complex measure words are given below:

- (50) 一 柜子 瓷器
yi gui-zi ciqi
one cabinet-ZI chinaware
‘a cabinet/cupboard of chinaware’

(Chao 1968: 601)

- (51) 一 伙子 贼
yi huo-zi zei
one band-ZI thief
‘a gang of thieves’

(He 2008: 67)

¹⁷ Here the meaning of the suffix *-r* is quite vague and not exactly diminutive.

The measure word *gui-zi* ‘cabinet’ in (50) is one of the six container MWs listed in Chao (1968: 601-602) in which the suffix *-zi* cannot be omitted.¹⁸ These complex MWs contrast sharply with classifiers, which are incompatible with the suffix *-zi*. In (51), the suffix *-zi* is technically optional, but the presence of this suffix strengthens the pejorative meaning of the phrase, and it has the same effect when added to a few other group MWs for human nouns (He 2008: 66-67). When co-occurring with measure words, the diminutive suffix *-r* can also have a stylistic effect, which is positive and affectionate (ibid. 67). It is worth noting that the suffix *-zi* adds an extra meaning of ‘entirety’ to container MWs (e.g. *yi guo-zi tang* ‘an entire pot of soup’) (ibid. 66), and that these complex container MWs with *-zi* are less acceptable in Dem-MW-N phrases and in bare MW-N phrases.

Temporary measure words, like G-P-C MWs, can be disyllabic and morphologically complex. A distinguishing characteristic of this subcategory is that most temporary MWs are morphologically stable, in the sense that they are either consistently complex (usually with the suffix *-zi*) or consistently simplex (Chao 1968). For instance, no suffix can be added to the temporary MW *lian* ‘face’, while the suffix in the MW *du-zi* ‘belly, abdomen’ is never omitted. This is true of all the temporary MWs listed in Chao (1968: 603) and in He (2008: 245), except for *shen* ‘body’, to which the suffix *-zi* can be added. The morphological stability of temporary MWs also accounts for their incompatibility with the Num-Adj-MW-N construction, and that there are hardly any compounds with temporary MWs. A typical temporary MW is shown in the example below:

- (52) 一 脸 汗
 yi lian han
 one face sweat
 ‘(having) sweat all over the face’

In (52), the numeral must be *yi* ‘one’, and its meaning is not transparent. The additional meaning of ‘all over’ is characteristic of this construction (namely *yi*-MW-N with

¹⁸ Although the morpheme *gui* ‘cabinet’ is not a measure word by itself, it can also become a measure word through compounding. For instance, *shū-gui* ‘bookcase’ is a measure word for books (e.g. *shū* ‘book’).

a temporary MW) (Chao 1968: 603). These measure words are “temporary” in the sense that the morphemes are most often used as nouns, whose referents are objects that are not designed as containers. While they are sometimes “treated as a subtype of container classifiers” (Li 2013: 19), their differences in distribution and in morphology support Chao’s (1968) classification, according to which container MWs and temporary MWs are two separate subcategories.

Standard measures are morphologically different from the other subcategories of measure words, as they do not allow any kind of affixation. They are not compatible with the suffixes *-r* or *-zi*, or with the size adjectives that can co-occur with other MWs (Chao 1968: 604-608). On the other hand, compounding is common in standard measures (He 2008). Compound standard measures may contain several types of morphemes in the attributive position in addition to the main morpheme, which can be observed in *pingfang-mi* ‘square metre’, *ying-cun* ‘inch’ (literally ‘English-*cun*’)¹⁹, and *qian-ke* ‘kilogram’ (literally ‘thousand-gram’). Complex standard measures are in general reasonably transparent in terms of morphological structure and meaning. Traditional Chinese weights and measures, which are still used in some situations, are usually simplex and monosyllabic.

In summary, classifiers generally have simpler morphology than measure words in Mandarin. Relatively few classifiers can be morphologically complex, when co-occurring the suffix *-r*. The group-partition-container measure words and temporary measure words are compatible with the suffix *-zi*, but for the former it is often optional, while for the latter subcategory it is either obligatory or prohibited in most cases. Standard measures could not co-occur with these suffixes, but they are often complex due to compounding.

2.2.3 Semantic similarities and differences between classifiers and MWs

Semantically, a major function of Mandarin classifiers and measure words is to individuate bare NPs, which refer to kinds instead of individuals in Mandarin (Chierchia 1998: 353-354). With the help of classifiers and measure words, nouns become compatible

¹⁹ *Cùn* is a traditional Chinese measure of length. The morpheme *yīng* (literally ‘English’) is usually added to imperial units.

with numerals and quantifiers such as *mei* ‘every’, and a difference could be made between *this* specific instance of a noun and *that* instance of it. When the unit for counting under discussion is clear from the context, measure words and classifiers may be placeholder-like in X-units-(of)-N expressions (Cheng & Sybesma 1999: 514), with few semantic roles to play other than individuation. Crucially, this function is shared by classifiers and MWs.

As Li (2013: 77-81) points out, some bare nouns can be highly ambiguous with “individual, stuff and partial-object readings”, at least in some contexts. The insertion of a classifier or a measure word in the bare NP generally disambiguates the sentence, which is illustrated in the examples below.

- (53) 桌 上 有 苹果。
 zhuo shang you pingguo
 table on there:be apple

- a. ‘There are apples on the table.’ OR ‘There is an apple on the table.’ [Individual reading]
- b. ‘There is apple on the table.’ [Stuff reading]
- c. ‘There is part of an apple on the table.’ [Partial-object reading]

(Li 2013: 79, example (16))

- (54) 桌 上 有 个 苹果。
 zhuo shang you ge pingguo
 table top there.be CL apple
 ‘There is an apple on the table.’

The sentence in (53) is ambiguous and has four possible interpretations. The bare noun *pingguo* ‘apple’ can refer to one or more individual apples, apple-stuff (such as cubes or slices of apples) or even “some part of an apple” (Li 2013: 79). By contrast, the sentence in (54) is unambiguous: only the individual reading of the noun *pingguo* ‘apple’ is available, and it refers to exactly one apple. In other words, the presence of the classifier *gè* prohibits the stuff reading and the partial object reading. In this and other similar situations, the

individuating function of classifiers has a disambiguating effect. The disambiguating effect is also found in NPs with measure words, an example of which is given below:

(55) 桌 上 有 水。

zhuo shang you shui
table on there:be water

- a. 'There is (splashed) water on the table.' [Stuff reading]
- b. 'There is a bottle of water on the table.' OR 'There are bottles of water on the table.' [Individual reading]²⁰

(Li 2013: 80, example (19); modified)

(56) 桌 上 有 瓶 水。

zhuo shang you ping shui
table top there.be MW_{-bottle} water

'There is a bottle of water on the table.'

In (55), the utterance is ambiguous with a stuff reading (i.e. splashed water) and an individual reading (i.e. bottled water), even though the stuff reading tends to be more salient than the individual reading. In (56), the bare MW-N phrase only has the individual reading and can only refer to exactly one bottle of water. Note that the disambiguating effect of the classifier *gè* in (54) and that of the measure word *píng* 'bottle' in (56) are remarkably similar.

Despite these similarities, classifiers and measure words are semantically different from each other in some ways. Importantly, a measure word "[creates] a unit of measure", while a classifier "[names] the unit in which the entity denoted by the noun naturally occurs" (Cheng & Sybesma 1999: 515). For instance, the natural units of trees and birds has existed long before they are named in any human language. On the other hand, measure words such as 'bottle', 'truckload' and 'kilogram' only exists after humans have created the

²⁰ Li (2013: 80) did not include the individual reading, but I add it as a second type of interpretation because the bare noun can be used this way in appropriate contexts.

container or have agreed upon the exact weight (or capacity, length, area, etc.) that a standard measure denotes.

This semantic difference between classifiers and measure words leads to other differences between them. Since a noun (with an ontologically discrete referent) generally corresponds to only one natural unit, there is very often only one name for that unit (i.e. the noun only selecting one classifier). However, a single noun can often be compatible with a wide range of measure words, which are man-made units for holding or measuring different kinds of objects. For instance, the noun *māo* ‘cat’ is only compatible with the classifier *zhī*, while the noun *jiu* ‘wine’ is compatible with many measure words, such as *ping* ‘bottle’, *bei* ‘glass’, *tong* ‘barrel’ and *sheng* ‘litre’.

Another semantic difference related to the “natural-artificial” difference between classifiers and measure words is found in the interpretation of Num-Adj-Cl/MW-N phrases. The adjective in a Num-Adj-Cl/MW-N phrase semantically modifies the noun, while the adjective in a Num-Adj-MW-N phrase can only modify the measure word. This difference is shown in examples below:

- (57) 一 大 箱 苹果
 yi da xiang pingguo
 1 big MW_{=box} apple
 ‘a big box of apples’
 (Her 2017: 31: (5a), slightly modified)
- (58) 一 大 颗 苹果
 yi da ke pingguo
 1 big CL apple
 ‘a big apple’
 (Her 2017: 31: (5b), slightly modified)

Examples (57) and (58) only differ in one morpheme, namely the measure word *xiang* ‘box’ in (57) and the classifier *ke* in (58).²¹ The meaning of (58) is simply ‘one big apple’, while the interpretation of (57) is different: it does not say anything about the size of the apples, but the box holding the apples has to be big for this utterance to be felicitous. In fact, the interpretation of (57) is expected, since the adjective is adjacent to the measure words, while that of (58) is unexpected. The semantic peculiarity of (58) is attributable to the fact that classifiers are “semantically and mathematically redundant in [Num-Cl/MW-N phrases]” (Her 2017A: 32). As the adjective cannot modify a semantically vacuous element, it bypasses it and modifies the head noun instead. In other words, the difference between (57) and (58) is purely semantic and should not be considered evidence against a unified account for the syntax of classifiers and measure words.

Syntactically, the use of pre-classifier adjectives in Mandarin is restricted in several ways. First, only a handful of dimensional adjectives (mainly *da* ‘big’, *xiao* ‘small’ and *chang* ‘long’) can occur in the pre-classifier position, and they can only co-occur with about half of the classifiers (Li 2013). Second, the pre-classifier adjectives cannot be modified with an intensifier, nor can they be followed by the modification marker *de*. Third, those adjectives cannot be reduplicated in the pre-classifier position, even though they can be reduplicated elsewhere (Luo et al. 2017).

All these restrictions can be explained by assuming that the adjective and the classifier merge morphologically (as two roots) to form a compound and that the “Lexical Integrity Hypothesis” is correct (Li 2013: 193-194). This Adj-CL compound analysis also accounts for a special semantic pattern: in certain contexts, the meaning of pre-classifier adjectives is evaluative rather than concrete (e.g. the meaning of *xiao* ‘small’ in Num-Adj-CL-N phrases can be closer to ‘insignificant’ than to ‘small’), while the meaning of the prenominal adjectives (e.g. in Num-CL-Adj-N phrases) remains concrete and literal (Li 2013: 191). Evaluative and abstract meaning tends to be found in compounds instead of phrases that are built in syntax (e.g. *wairen* ‘outsider’ vs *waimian-de-ren* ‘people outside’, literally ‘outside-DE-person’).

²¹ The classifier *kē* for apples is mainly used in Taiwanese Mandarin. Mainland Mandarin speakers usually prefer the generic classifier *gè* for apples, hence the lexical difference between (54) and (58).

In summary, both classifiers and measure words have the individuation function and a disambiguating effect in Mandarin NPs. They are interpreted differently in Num-Adj-Cl/MW-N phrases, and this difference is rooted in the fact that MWs creates new units while classifiers names naturally existing units. A pre-classifier adjective and a classifier form a compound before merging with other parts of the Num-Adj-Cl/MW-N phrase.

2.2.4 Lexical characteristics of classifiers and MWs

For much of the 20th century, *dānwèi-cí* “unit words” (an early name of classifiers, including individual classifiers and measure words) were widely believed to be a subtype of nouns (He 2008: 3-6). Li (2013: 23-28) has argued convincingly that classifiers form a distinct grammatical category and that they are not nominal. Historically, classifiers and measure words can be derived from verbs or adjectives, even though most of them were derived from nouns (Li 2013; Chao 1968).

Synchronically, hardly any (individual) classifiers can also be used as nouns in Mandarin.²² Some of them can be a bound morpheme in another noun, but even that is not true for many classifiers (Li 2013: 26-28; Chao 1968: 589-593). The same could be said of the group MWs and partitive MWs (Chao 1968). Container MWs and temporary MWs can usually be used as nouns as well, which is why they are sometimes considered to be borrowed from nouns (Chao 1968: 601-603; He 2008: 244). However, some container MWs are morphologically simpler than the corresponding nouns. For instance, the container MW *xiang* ‘box’ is a simplex, while *xiang-zi* ‘box’, the noun with the same morpheme and a directly related meaning, is morphologically complex.²³ As for standard measures, they are rarely used as nouns or as morphemes in other nouns (Chao 1968).

²² A morpheme is considered as such if the noun and the classifier are homophonous and etymologically related. A notable exception is the classifier *tou* (used with nouns denoting animals such as cows and donkeys), which means ‘head’ as a noun.

²³ The morphologically complex MW *xiang-zi* ‘box’ does exist, but it is much less frequent than *xiang*.

There are also a few cases of one morpheme being used as a classifier and as a measure word, or as two different subtypes of measure words. For instance, the morpheme *bǎ* is both a measure word meaning “handful” and the classifier for knives and chairs. The morpheme *jíé* is both a measure word meaning “section” (mainly used with bamboos and sugar cane) and the classifier for lessons. The morpheme *kuài* is both a partition MW meaning ‘lump, piece’ and a standard measure for currencies (Chao 1968). The occurrence of this phenomenon seems to be random, with no clear patterns in its distribution.

In summary, there are some non-syntactic reasons to distinguish classifiers and measure words as two subcategories of “unit words”, notably their differences in maximum length and morphological complexity, as well as their interpretation in Num-Adj-Cl/MW-N phrases. It is worth noting that even in these domains there is conflicting evidence: classifiers and measure words behave alike in trisyllabic Tone-3 Num-Cl/MW-N phrases, and they have very similar disambiguating effects when added to some bare nouns. Classifiers may also share some behaviour with some subtypes of measure words (esp. group MWs and partition MWs) but not with others, which is observed in their distribution in various constructions and in the rareness of lexical ambiguity.

2.3 Evidence for [Num-CL/MW] constituency in Mandarin Chinese

Since the syntactic behaviour of classifiers and that of measure words are extremely similar (if not identical) in Mandarin, I have demonstrated in section 2.1 that they belong to the same grammatical category. Then another question arises about their syntax, more precisely about the timing of merger with other parts of the NP, namely whether they merge with the numeral or with the head noun first. This is discussed in the current section, and I argue that a classifier or a measure word merges with the numeral first (to the exclusion of the head noun), hence the [Num-Cl/MW] constituency.

In this section, I start by examining the basic distribution of classifiers and MWs, and I argue that the existence of Demonstrative-CL/MW-N phrases and post-verbal bare CL/MW-N phrases is not conclusive evidence against my claim. Then I use the same syntactic tests as in section 2.1 (viz. extraction, ellipsis, and coordination) to show that

classifiers and MWs are indeed structurally closer to the numeral than the head noun in Mandarin. Lastly, I briefly discuss a possible link between [Num-Cl/MW] constituency and the morpheme order of complex numerals in Mandarin.

When classifiers or measure words co-occur with numerals in Mandarin, the word order is normally Num-Cl/MW-N and only occasionally N-Num-Cl/MW. In the latter case, the head noun is fronted for the purpose of presenting or listing things. Note that the usual constituent order Num-Cl/MW-N can be generated in two ways: the classifier or MW merging with the head noun first or with the numeral first. Both analyses are logically possible, but other syntactic patterns will show that the latter is correct.

Before going into the syntactic tests, I shall examine the behaviour of Dem-Cl/MW-N phrases and post-verbal bare Cl/MW-N phrases, which are also common in Mandarin and may seem to contradict my claim. While no overt numerals appear in these phrases, I argue that the Numeral head is null due to phonological reduction, and that the classifiers or the measure words merge with this null Numeral head.

Two pieces of evidence indicate that Dem-Cl/MW-N phrases are underlyingly Dem-Num-Cl/MW-N phrases with a null numeral head. First, in a Dem-Cl/MW-N phrase in Mandarin, the numeral *yi* ‘one’ can always be inserted between the demonstrative and the classifier (or measure word), with minimal change in the meaning.²⁴ Although Dem-*yi*-Cl/MW-N phrases are less common than Dem-Cl/MW-N phrases, these two types of phrases have the same distribution: in a Mandarin sentence, they can both be used as the topic, the subject, the post-verbal object or the fronted object in the disposal construction. This suggests that the difference between Dem-*yi*-Cl/MW-N phrases and Dem-Cl/MW-N phrases is phonological rather than syntactic. In Mandarin, the deletion of phrase-internal *yi* ‘one’ is by no means restricted to NPs. In complex verbs with the shape V-*yi*-V (e.g. *kàn-yi-kàn* ‘take a look’), the morpheme *yi* can be omitted. Likewise, in complex numerals such as *san-bai-yi-shi-wu* ‘three hundred and fifteen’ (literally ‘three hundred one ten five’), it is also common to drop the numeral *yi* in the phrase-internal position. It is hardly surprising that the numeral *yi* ‘one’ can be deleted in Dem-*yi*-Cl/MW-N phrases, which creates the

²⁴ A Dem-Cl/MW-N phrase is obviously definite in meaning. Unlike bare nouns, the noun in a Dem-Cl/MW-N phrase can only be singular.

Dem-Cl/MW-N construction. It is worth noting that the equivalent of Dem-*yi*-Cl/MW-N phrases is ungrammatical in Northern Wu (Li 2013: 280). The relevant structural differences between these two languages are discussed in the next chapter.

Second, there is empirical evidence for the phonological reduction of the numeral *yi* in the Dem-Cl/MW-N phrases in Mandarin. While the demonstratives in Mandarin are normally pronounced *zhè* ‘this’ and *nà* ‘that’, they have alternative forms in Dem-Cl/MW-N phrases, namely *zhèi* ‘this’ and *nèi* ‘that’. These alternative forms are the contractions of the demonstratives and the numeral *yi* ‘one’, and they mainly used in colloquial Mandarin. This phonetically reduced numeral only changes the vowel in the demonstratives. The Dem-Cl/MW-N phrases with *zhèi* ‘this’ or *nèi* ‘that’ indicate that the numeral *yi* can be reduced in Dem-*yi*-Cl/MW-N phrases as well.

Although phonologically the demonstrative and the numeral can contract to one syllable in a Dem-*yi*-Cl/MW-N phrase, results of syntactic tests suggest that they do not constitute a constituent. For instance, *zhe-yi* ‘this one’ and *na-yi* ‘that one’ are never acceptable in answers with N-ellipsis, which is shown in the example below:

- (59) 你 选 哪 一 幅 画 ? 这 一 幅 。
- ni xuan na yi fu hua zhe yi *(fu)
- you choose which one CL painting this one CL
- ‘Which painting do you choose?’ ‘This one.’

In (59), the classifier is obligatory both in the question and in the answer. The numeral *yi* ‘one’ could be phonetically reduced or deleted here, but the Mandarin translation of the English phrase *this one* must include a classifier. This holds true when a different numeral is used in this position. As this test indicates that the demonstrative and the numeral do not form a single constituent, the plausible constituency of the numeral and the classifier will be discussed later in this section. Since the Dem-Cl/MW-N phrases in Mandarin are underlyingly Dem-Num-Cl/MW-N phrases with the numeral *yi* ‘one’ that is deleted phonetically, the existence of Dem-Cl/MW-N phrases does not exclude the possibility of having [Num-Cl/MW] constituents in Mandarin NPs.

As for the post-verbal bare CL/MW-N phrase, I argue that there is also a Numeral head in its structure, which is not pronounced at the PF level. However, the fact that Num⁰ is not pronounced in this context is not due to its deletion in phonology. The differences between bare CL/MW-N phrases and *yi*-CL/MW-N phrases in distribution and interpretation exclude that possibility (Cheng & Sybesma 1999: 525-527; Li & Bisang 2012: 344-347). Instead, the silent Num⁰ is occupied by a special form of the numeral *yi* ‘one’, namely YI_{pro}, which is unpronounced and has syntactic properties different from overt numerals (Zhang 2019: 184). This special form is posited for two reasons. First, bare CL/MW-N phrases can only be interpreted as singular and indefinite, just like *yi*-CL/MW-N phrases and unlike bare nouns (which can be plural and definite) in Mandarin (Cheng & Sybesma 1999). As Zhang (2019: 194) points out, “neither the CL nor the NP encodes a singular meaning”, positing the null numeral YI_{pro} is therefore a simple way to account for the singular meaning of those phrases. Second, this account based on the silent numeral YI_{pro} is arguably more elegant than postulating that Mandarin has a separate set of deictic classifiers in bare CL/MW-N phrases, esp. when the set of deictic classifiers and that of numeral classifiers are lexically identical and semantically similar (Zhang 2019).

The distribution of bare CL/MW-N phrases is highly restricted in Mandarin, much more so than the Num-CL/MW-N phrases. Semantically, bare CL/MW-N phrases must be indefinite and non-specific. In terms of syntactic positions, these phrases are normally used as post-verbal objects (Li & Bisang 2012).²⁵ Furthermore, bare CL/MW-N phrases are not compatible with some verb forms, notably verbs with the progressive aspect marker *zài* and those with the experiential marker *guò*. By contrast, *yi*-CL/MW-N phrases (and other Num-CL/MW-N phrases) can be specific and are used in various positions and with various aspect markers (including the ones mentioned above) (Li 2013: 96). In other words, YI_{pro} is not simply equivalent to the weak form of numeral *yi* ‘one’²⁶, since there are other syntactic

²⁵ Some native speakers accept the use of bare CL/MW-N phrases in the disposal construction, where the object is fronted to a preverbal position following the object marker *ba* (Li & Bisang 2012: 339). To me, using bare CL/MW-N phrases in that context sounds non-native-like and causes difficulty in interpretation.

²⁶ The numeral *yi* can be weak when its meaning is indefinite without focus on quantity (Li & Bisang 2012). It is unstressed but cannot be deleted freely.

constraints on its occurrence. However, it seems safe to assume that bare CL/MW-N phrases and Num-CL/MW-N phrases differ in the lexical features of Num⁰ but not in their syntactic structure. As evidence points to the existence of this null numeral and a Num⁰ in bare CL/MW-N phrases, the use of bare CL/MW-N phrases is compatible with my theory of [Num-Cl/MW] constituency in Mandarin NPs.

In section 2.1, I showed that classifiers and measure words belong to one and the same category using three syntactic tests, namely extraction, ellipsis, and coordination. These tests also allow me to determine whether classifiers and measure words merge with numerals first or with nouns first. For instance, if the classifier (or measure word) can be extracted from a NP along with the noun, then it points to [CL/MW-N] constituency. If the classifier or measure word can never be extracted with the head noun from Mandarin NPs, then [Num-Cl/MW] constituency is more likely (Her 2017). I shall start by applying the extraction test to Mandarin NPs, and the other tests will be discussed later in the section.

N-extraction from Num-CL/MW-N phrases or Dem-CL/MW-N phrases is generally acceptable in Mandarin. In fact, fronting of the head noun is common if it is the topic or bears contrastive focus. This is illustrated in the examples below:

- (60) 料酒 嘛，有 一 瓶 就 足够 了。
liaojiu ma you yi ping jiu zugou le
cooking.wine TOP there.be one MW=_{bottle} then enough ASP
‘As for cooking wine, if there is one bottle of it, then that will be enough.’
- (61) 她的 小说 我 只 看过 这 本。
ta-de xiaoshuo wo zhi kan-guo zhe ben
she-POSS novel I only read-EXP this CL
‘Of her novels, I have only read this one.’

In (60), the head noun *liaojiu* ‘cooking wine’ is the topic and moves from the Num-MW-N phrase to the beginning of the sentence. The measure word *ping* ‘bottle’ cannot be fronted with it. The same is true of Num-Cl-N phrases: the head noun can be fronted to the left edge of the clause, while the [classifier-N] part of the phrase cannot. In (61), the NP *ta-*

de-xiaoshuo ‘her novels’ moves from the Dem-CL-NP phrase at the end of the sentence to the beginning of the sentence as the topic. The classifier *ben* cannot be fronted with that NP. The same holds true for Dem-MW-N phrases: the head noun can undergo topic fronting while the [MW-N] part of the phrase cannot. A simple explanation for these facts is that the CL/MW-N sequence is not a constituent in such phrases in Mandarin (Her 2017). This implies that the numeral and the classifier or measure word form a constituent. No evidence has been gathered from bare CL/MW-N phrases because N-extraction from these phrases is ungrammatical.

N-ellipsis in Num-CL/MW-N phrases or Dem-CL/MW-N phrases is grammatical in Mandarin when the phrase is in an appropriate context. The antecedent is generally overt and salient in the discourse, so that the ellipsis will not cause confusion. Two examples of N-ellipsis are shown below:

- (62) 张三 吃了 两 片 面包， 李四吃了 三 片。
 Zhangsan chi-le liang pian mianbao Lisi chi-le san *(pian)
 Zhangsan eat-ASP two MW=_{piece} bread Lisi eat-ASP three MW=_{piece}
 ‘Zhangsan ate two pieces of bread, and Lisi ate three pieces.’
- (63) 我 买了 这 本 书， 他 买了 那 本。
 wo mai-le zhe ben shu, ta mai-le na *(ben)
 I buy-ASP this CL book, he buy-ASP that CL
 ‘I bought this book, and he bought that one.’
 (Example (19) is repeated here.)

In (62), the head noun *mianbao* ‘bread’ of the Num-MW-N phrase *san-pian-mianbao* ‘three pieces of bread’ in the second clause has been elided. Unlike the head noun, the measure word *pian* ‘piece’ cannot be deleted in Mandarin, although ellipsis of measure words is acceptable in some languages including English (Li 2013: 40). The same pattern is found in Num-CL-N phrases, where the head noun can be deleted whilst the [CL-N] part of the phrase cannot. In (63), the head noun *shu* is deleted in the second Dem-CL-N phrase, since the discourse is obviously about books. Again, the classifier cannot be deleted in this

phrase, even though the same classifier appears in the antecedent. The same is true of Dem-MW-N phrases: the measure word cannot be deleted together with the head noun in this type of phrases in Mandarin. A straightforward way to explain this constraint is to posit a [Num-Cl/MW] constituent rather than a [Cl/MW-N] constituent in these phrases.

As for the coordination test, if two [Num-Cl/MW] sequences can be conjoined using coordinating conjunctions, then that part of the phrase should be considered a constituent. If two [Cl/MW-N] sequences can be connected using a coordinator, then it would be reasonable to treat the [Cl/MW-N] sequence as a constituent. In Mandarin, [Num-Cl/MW] sequences can be conjoined, while [Cl/MW-N] sequences cannot. This is exemplified below:

- (64) 约翰 和 玛丽 分别 有 三 本 和 五 本 书。
 Yuehan he Mali fenbie you san ben he wu ben shu
 John and Mary respectively have three CL and five CL book
 ‘John and Mary have three books and five books respectively.’
 (Example (23) is repeated here.)
- (65) 我 买了 (一) 本 书 和 支 笔
 *wo mai-le (yi) ben shu he zhi bi
 I buy-ASP one CL book and CL pen
 Intended: ‘I bought a book and a pen.’

In (64), the sequence [three-CL-and-five-CL] shows that conjoining two [Num-CL] sequences is grammatical. In fact, if the noun *shu* ‘book’ were fronted to the beginning of the sentence as the topic, the sentence would still be grammatical. The same holds true for [Num-MW] sequences and [Dem-Cl/MW] sequences. On the other hand, example (65) indicates that conjoining two [CL-N] sequences is ungrammatical. The ungrammaticality persists even if a numeral is added at the beginning of the first [Cl-N] phrase. The same is true of [MW-N] sequences. In other words, conjoined [Cl/MW-N] sequences are not found in any contexts in Mandarin. The result of the coordination test thus affords telling evidence for my claim that the [Num-Cl/MW] sequence form a constituent.

In addition, the substitution test can also be applied to Num-CL-N phrases. Numeral-classifier sequences in Mandarin that consist of the numeral *liang* ‘two’ or *san* ‘three’ and the generic classifier *ge* can be replaced by a single morpheme, namely *lia* ‘two.CL’ or *sa* ‘three.CL’. A sentence containing the morpheme *lia* is shown below.

- (66) 我 就 买了 [两 个西瓜] or [俩 西瓜]。
 Wo jiu lai-le [liang ge xigua] or [lia xigua]
 I only buy-ASP two CL watermelon or two.CL watermelon
 ‘I only bought two watermelons.’

In (66), the numeral-classifier part of the NP meaning ‘two watermelons’ can be replaced by the morpheme *lia*, which is never followed by a classifier. The morphemes *lia* and *sa* are used mostly in colloquial Mandarin. The result of this test indicates that Num-CL sequences are constituents in Mandarin.

An examiner mentions that the CL-N part of a Num-CL-N phrase can be replaced by the pronoun *zhe* in Mandarin, for example in sentences such as (67).

- (67) 他 买了 一 个蛋糕 和 一 个西瓜。 二者 都 很 甜。
 ta mai-le yi ge dangao he yi ge xigua. er-zhe dou hen tian.
 he buy-ASP one CL cake and one CL watermelon two-ZHE both very sweet
 ‘He bought a cake and a watermelon. Both of them are very sweet.’

In (67), the morpheme *zhe* in the second sentence appears to replace the two CL-N sequences (i.e. CL-cake and CL-watermelon) in the NPs in the first sentence. However, there are two reasons for which this should not be considered a valid substitution test for Mandarin Num-CL-N phrases. First, this “modified pronoun” is best analyzed as “a residue of ancient Chinese” (Wang 2023: 33), as it is rarely used in colloquial Mandarin but is more frequent in formal or written Mandarin, as well as in Old Chinese texts such as the *Analects*

and *Zuozhuan* (ASTC-OC).²⁷ Second, numeral-*zhe* sequences are NPs that replace Num-CL-N phrases as a whole, but the pronoun *zhe* does not simply equal the CL-N part. Example (68) illustrates this point.

- (68) 有 三 个蛋糕 和 五 个瓜。 八者 / 二者 都 很 甜。
you san ge dangao he wu ge gua. ??ba-zhe / er-zhe dou hen tian.
have three CL cake and five CL melon eight-ZHE/two-ZHE all very sweet
'There are three cakes and five melons. They are all very sweet.'

In (68), the NP meaning 'three cakes and five melons' in the first sentence can be replaced by a numeral-*zhe* sequence, but the numeral that precedes the pronoun *zhe* is not simply the sum of the two numerals (i.e. *san* 'three' and *wu* 'five'). Instead, the numeral *er* 'two' is used with *zhe* at the beginning of the second sentence in (68). In other words, each of the two Num-CL-N phrases functions as an undivided unit when replaced by numeral-*zhe* sequences, and the pronoun *zhe* does not correspond to a specific part of Num-CL-N phrases. In short, substituting numeral-*zhe* phrases for Num-CL-N phrases is not a valid constituency test for CL-N sequences in Mandarin.

Her (2017: 42) points out that Num-Cl/MW order correlates with the multiplier-multiplicand order in complex numerals. For example, Mandarin numerals are base-final (e.g. *san-bai* literally 'three hundred') with multiplicands such as *shi* 'ten' and *bai* 'hundred' following the multiplier, and Mandarin classifiers and measure words follow the numeral. On the other hand, the Austronesian language Kilivila has a base-initial numeral system (where the literal translation of *two hundred* is 'hundred two') and the Cl/MW-Num order in NPs (Her 2017: 57-58). Both multiplicands and measure words (or classifiers) can be interpreted as units for counting, and the similarity between the complex numeral *three hundred* and the numeral-MW sequence *three dozen* is obvious (Her 2017). According to a survey based on 52 classifier languages, most classifier languages are either base-final and classifier-final or base-initial and classifier-initial, and inharmonious orders are found in

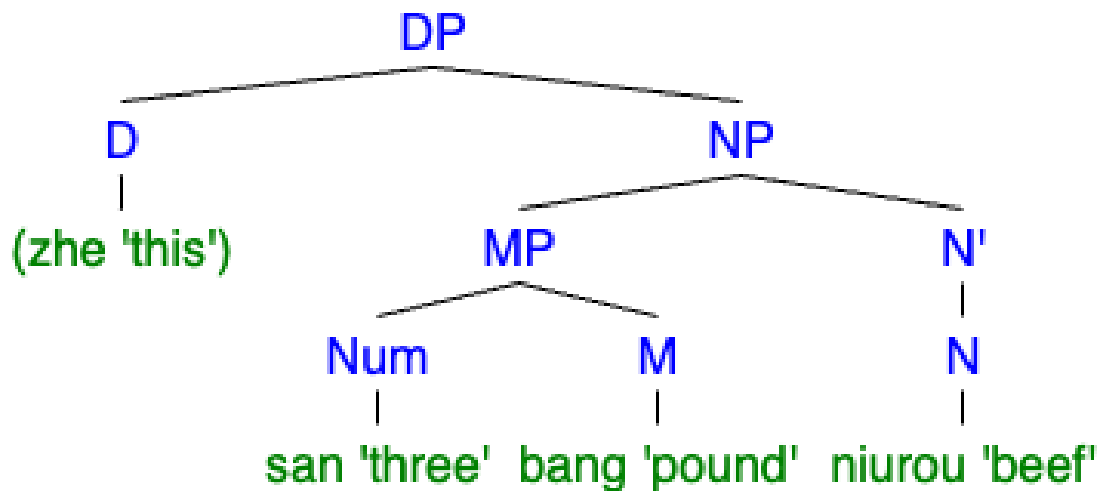
²⁷ This residual use of *zhe* is very restricted even in formal Mandarin. For example, the numeral *yi* 'one' and complex numerals are incompatible with the pronoun *zhe* (e.g. **ershi-zhe* 'twenty-ZHE').

four of the languages, which only represent 7.7% of the sample (Her 2017). Since the multiplier and multiplicand form a constituent to the exclusion of other words, it seems reasonable to assume the numeral and classifier (or MW) also form a constituent in the absence of compelling counterevidence. In other words, there is some cross-linguistic evidence for [Num-Cl/MW] constituency in Mandarin.

In some classifier languages such as Mi'gmaq and Chol, some numerals must be accompanied by classifiers, while others cannot (Bale & Coon 2014). This has been argued to be a piece of evidence for the claim that “[the] requirement for a classifier is dependent... on the syntax and semantics of the numeral”, as there may or may not be an incorporated measure function in the denotation of a numeral (Bale & Coon 2014: 702). A similar situation is found in Mandarin. While different numerals do not differ in terms of their compatibility with classifiers, two of the lower numerals have special forms that are mostly used in Num-Cl/MW-N phrases. The form of the numeral ‘one’ is *yī* if it is used in fractions, percentages, decimals and when used ordinally or in counting. When it combines with a noun, a classifier or measure word will be inserted, and its form changes to *yí* or *yì* (but never *yī*) depending on the tone of the next syllable. Likewise, the form of the numeral ‘two’ is *èr* if it is used in fractions, percentages, decimals and when used ordinally or in counting. When it is used in Num-Cl/MW-N phrases, the numeral ‘two’ takes the special form *liang*. In other words, the forms of these numerals change depending on whether they combine with a noun (with the help of classifiers or measure words) or not.

Based on the distribution of classifiers, the results of the syntactic tests and other types of evidence discussed in this chapter, I propose that Mandarin measure words and classifiers belong to the same category and occur in a left-branching structure which is represented in the syntactic tree below:

Tree 1: Structure of Mandarin quantified NPs with a classifier or measure word



(Based on Bale et al. 2019: 2 (1))

Tree 1 shows the structure of Mandarin Num-CL/MW-N phrases, more specifically the phrase *san-bang-niurou* ‘three pounds of beef’.²⁸ The Num-CL/MW constituent is a MP (Measure Phrase) at the specifier position of the NP, with the measure word *bang* ‘pound’ at Measure⁰. Classifiers such as *tou* for the noun *niu* ‘cattle’ also occur in the M⁰ position. As the numeral *yi* ‘one’ can sometimes be null, the Dem-CL/MW-N phrases and post-verbal bare CL/MW-N in Mandarin also have structures comparable to the one in Tree 1. This structure account for the selectional relation between the lower numerals (i.e. 1 and 2) and classifiers (or measure words), but the selection of classifiers by nouns may seem more problematic.

In fact, syntactic selection between classifiers and nouns is unnecessary, and semantic violations can also eliminate the incorrect combinations of nouns and classifiers. For instance, the classifier *tiao* is used with nouns for objects that are slender and pliable or winding (e.g. *she* ‘snake’, *xianglian* ‘necklace’), while the classifier *zhang* is used for nouns denoting flat and thin objects (e.g. *ditu* ‘map’, *zhaopian* ‘photo’), and the fact that neither of these two classifiers is used with the noun *xigua* ‘watermelon’ is expected in view of their

²⁸ An examiner asks about the possibility that the classifier is base generated in a position where NP is its complement and then head-moves to Num. The main problem with that hypothesis is that there does not seem to be any trace of head movement in Mandarin NPs, and the movement seems ad hoc as well.

semantic incompatibility (Wu 2022: 35, 98). For the infrequent classifiers (e.g. *mei*) that are not shape classifiers and are not used for a specific category of objects, they are used in an idiomatic way and can generally be replaced with the generic classifier *ge* (Chen 2024).

Saito et al. (2008) propose a right-branching structure for Mandarin NPs (which is therefore very different from the one I propose) based on N'-ellipsis patterns and the distribution of modification markers in Mandarin and Japanese. According to them, the modification marker *de* in Mandarin is in fact D⁰, while the modification marker *no* is a contextual case marker in Japanese (Saito et al. 2008: 252, 256).

I summarize below the major types of evidence for the [D [Num [CL N]]] structure of Mandarin DPs that they adduce. First, Mandarin *de* cannot be inserted between Num-CL and the head noun (since D is not between CL and N), while Japanese *no* can (Saito et al. 2008: 251). Second, Num-CL sequences can be extracted from the DP and fronted to the sentence-initial position in Japanese, while that is not possible in Mandarin because Num-CL is not a constituent according to them (ibid. 260-261). Third, N'-ellipsis is grammatical in Mandarin Num-CL-N phrases, while it is ungrammatical in Japanese Num-CL-*no*-N phrases (ibid. 261). Lastly, Mandarin *de* is obligatory after relative clauses (whose position is Spec DP), while Japanese *no* is prohibited after relative clauses (ibid. 262-264).²⁹

None of the above seems to be conclusive evidence that Mandarin DPs really have the [D [Num [CL N]]] structure, and I argue against each of these arguments. First, *de* are sometimes inserted in Num-CL-N phrases, for example when the numeral is approximative or “a high round number” (Li 2013: 64-65). Second, if one assumes that Japanese Num-CL sequences are adjuncts and that Mandarin Measure Phrases (containing a numeral and a classifier) are specifiers, then focus fronting of an adjunct is expected to be more acceptable than that of a specifier (cf. English *Great was...*). In other words, this difference in Num-CL extraction indicates that the Num-CL sequences in Mandarin are not adjuncts, not that the structure is right-branching. Third, N'-ellipsis occurs in Num-CL-N phrases in Mandarin because the Measure Phrase (i.e. [Num-CL]) c-commands the head noun as a specifier, while the Japanese Num-CL sequence, as an adjunct, does not c-command the head noun. Lastly, relative clauses can precede or follow Num-CL sequences in Mandarin, and *de* is

²⁹ There are other arguments in Saito et al. 2008, but they are rather tangential to the syntax of classifiers.

obligatory in both cases, but the structure they proposed (Saito et al. 2008: 263) appears to be incompatible with Num-CL-relative clause-*de*-N phrases.³⁰ In short, the differences between Mandarin and Japanese discussed in Saito et al. (2008) are attributable to the difference between Num-CL adjuncts and Num-CL specifiers, or to the difference between head-initial and head-final languages, and the facts that they adduced are not incompatible with my left-branching analysis of Mandarin [Num-CL]-N phrases.³¹

To conclude, the results of four syntactic tests and other types of evidence suggest that the numeral and the classifier or measure words form a constituent in Num-CL/MW-N phrases, and I label that constituent “Measure Phrase”. The existence of Dem-CL/MW-N phrases and bare CL/MW-N phrases do not constitute evidence against classifiers because they are underlyingly Dem-Num-CL/MW-N phrases (or Num-CL/MW-N phrases) with a phonetically null numeral.

³⁰ There are other relevant facts in Mandarin (not mentioned in Saito et al. 2008) that make their proposal more problematic. For instance, N'-ellipsis is ungrammatical in Num-CL/MW-*de*-N phrases, even though there is no reason why it could not occur if *de* was indeed the head of DP.

³¹ As for the status of *de*, the morpheme *de* in Num-CL/MW-*de*-N phrases should simply be a modification marker that places emphasis on the quantity (Li 2013: 217, 224). However, since *de* is extremely versatile in Mandarin, the possibility that it is D⁰ in certain types of NPs (e.g. possessor-*de*-N) cannot be excluded.

3. The classifiers and measure words in Wu and Cantonese

Individual classifiers are found in many other Sinitic languages, including Wu and Cantonese. Like their Mandarin counterparts, classifiers are obligatory in quantified NPs in Wu and Cantonese and are used in some other constructions (e.g. with demonstratives). There are numerous lexical similarities between the classifiers in Mandarin, Wu, and Cantonese. In all three languages, the compatibility between a classifier and a noun is partly dependent on the shape or size of the referent of the noun (Wu 2022). The classifiers are monosyllabic in these languages. Despite these superficial similarities, there are good reasons to believe that the NPs with classifiers in Wu and Cantonese are structurally different from their counterparts in Mandarin, and the structural differences among these languages will be discussed throughout this chapter.

Wu Chinese is spoken by over 70 million native speakers in eastern China, chiefly in Shanghai, Zhejiang, and the southern part of Jiangsu. Different dialects of Wu may not be readily mutually intelligible with one another. In this thesis, I focus mostly on the Ningbo dialect, a northern dialect of Wu Chinese which I speak as a mother tongue. All the data in (Northern) Wu will be from this dialect if not otherwise specified.

Cantonese is spoken by some 55 million native speakers mainly in Guangdong, Guangxi (mainland China), Hong Kong and Macau. When broadly defined, Cantonese is equivalent to the Yue dialect group, which includes several subgroups of dialects. Here I focus on the syntax of Hong Kong Cantonese, which is a highly influential variety and very similar to the Guangzhou dialect of Cantonese (sometimes considered to be the standard variety) (Matthews & Yip 2011).³²

This chapter is organized as follows: The distribution of the classifiers in Wu and Cantonese will be described concisely in the first section. Evidence in support of making a

³² Since modern written Chinese is based on Standard Mandarin, educated speakers of Wu and Cantonese can often write according to the grammar of Mandarin and read written texts out loud using Wu or Cantonese pronunciation (Matthews & Yip 2011: 7-8). In that situation, there can be loanwords and borrowed constructions that are not really part of the grammar of (spoken) Wu or Cantonese. In this chapter, I exclude all the constructions that are only used in writing and are not found in colloquial Wu or Cantonese.

distinction between classifiers and measure words in Wu and Cantonese will be presented and discussed in section 3.2. In section 3.3, I argue that measure words merge with the numeral first in a Num-MW-N phrase in Wu and Cantonese, and that classifiers merge with the head noun first in these languages. Section 3.4 serves as an interim conclusion where I summarize the major findings and claims of this and the previous chapter.

3.1 Distribution of classifiers in Northern Wu and Cantonese

As I have shown in chapter 2, the classifiers in Mandarin are used with numerals, with demonstratives, and in post-verbal bare CL-N phrases. The classifiers in Northern Wu also occur in those three types of NPs, and there are three additional contexts in which Wu classifiers can be used, namely in preverbal bare CL-N phrases, with a possessor NP, and in adjective-CL-N phrases. These types of NPs with classifiers are exemplified below.

- (1) 一 只 猫 [Ningbo Wu]
 iq tsaq mae
 one CL cat
 ‘a cat’
- (2) 该 部 汽车
 kiq bu chitsho
 that CL car
 ‘that car’
- (3) 渠拉 屋落 有 只 猫。
 jiqraq oqloq yeu tsaq mae
 they home have CL cat
 ‘They have a cat at home.’

Examples (1) to (3) show that Wu classifiers can also be used in Num-CL-N phrases, demonstrative-CL-N phrases, and post-verbal bare CL-N phrases. The classifier in a Num-CL-N phrase such as (1) is obligatory. In demonstrative-CL-N phrases, using the classifier is

preferable if the head noun denotes a human being. With other nouns, omission of the classifier makes the utterance unacceptable. The interpretation of the post-verbal bare CL-N phrases in Northern Wu is comparable to their counterparts in Mandarin since they can only be indefinite and singular in both languages (Li & Bisang 2012). In these phrases, the classifier is obligatory in the sense that its absence would change the meaning of the NP (which is then compatible with the plural reading). Unlike in Mandarin, however, the bare CL-N phrases in Northern Wu can also occur in pre-verbal positions as the subject or the topic. This is illustrated in example (4).

(4) 只 猫 欢喜 吃 鱼。

tsaq mae hunshi chioq ng

CL cat like eat fish

'The cat likes eating fish.'

Example (4) shows that pre-verbal bare CL-N phrases are grammatical in Northern Wu, and that they can only be interpreted as definite NPs. The classifier is obligatory because the bare noun would no longer have the definite reading if the classifier were deleted. The classifiers can be used in two other types of definite NPs in Northern Wu, which is shown in (5) and (6).

(5) 阿叔 部 汽车

ason bu chitsho

uncle CL car

'(my) uncle's car'

(6) 红红 部 汽车

ghonghon bu chitsho

red CL car

'the red car'

In (5), the classifier is employed between a possessor noun and the head noun, and the outcome is a singular definite NP. Possessor-CL-N phrases can be used as subjects, topics, or objects in Northern Wu. In (6), the classifier is used between an adjective and the head noun, and the entire phrase is also singular and definite. Adjective-CL-N phrases can also be subjects, topics, or objects. In both types of phrases, the classifier cannot be deleted but can be replaced with the linking particle *geq*, although the NPs with *geq* rather than a classifier can also be plural.

In Cantonese, the classifiers are often employed with numerals, with demonstratives and in post-verbal bare CL-N phrases, like their Mandarin and Northern Wu counterparts. In addition to those positions, Cantonese classifiers may also be used with a possessor NP, with an adjective, or in a pre-verbal bare CL-N phrase (Matthews & Yip 2011). Examples of these types of phrases are shown below:

(7) 三 隻 貓仔 [Hong Kong Cantonese]

saam zek maau-zai
three CL cat-DIMIN
'three kittens'

(8) 呢 個 病人

lī go behngyàhn
this CL patient
'this patient'

(Matthews & Yip 2011: 107)³³

(9) 我 聽日 會 搵 個 律師。

Ngóh tīngyaht wúih wán go leuhtsi
I tomorrow will contact CL lawyer
'I'll contact a/the lawyer tomorrow.'

³³ Matthews and Yip (2011) use a modified version of the Yale Romanization system for Cantonese, with which I am unfamiliar. I have not converted their transliterations into Jyutping, which I usually use.

(Matthews & Yip 2011: 106)

In example (7), the classifier *zek* occurs between the numeral ‘three’ and the head noun ‘kitten’. In (8), the classifier is between the demonstrative ‘this’ and the noun. In both (7) and (8), the classifiers are obligatory (Matthews & Yip 2011). Example (9) contains a post-verbal bare CL-N phrase, which is ambiguous as it has a definite reading and an indefinite reading.³⁴ Deleting the classifier in a post-verbal bare CL-N phrase such as (9) is possible, but it makes the NP ambiguous in number (with an additional plural meaning).

- (10) 隻 狗 今日 特別 聽話。
Zek gau gamjat dakbit tengwaa
CL dog today special obedient
‘The dog is especially obedient today.’
(Cheng & Sybesma 1999: 511, example (4b))

- (11) 你 個 朋友
Léih go pàhngyáuh
you CL friend
‘your friend’
(Matthews & Yip 2011: 128)

- (12) 我 咁 大 個 女 都 未 拍過拖。
Ngóh gam daai go léui dōu meih paak-gwo-tō
I so big CL girl still not-yet date-EXP
‘I’ve never been on a date, even at my age.’
(Matthews & Yip 2011: 112)

³⁴ Note that the post-verbal bare CL-N phrases in Mandarin and Northern Wu can only be indefinite.

Example (10) shows that Cantonese allows bare CL-N phrases to occur in the pre-verbal position as well, and that they must be interpreted as definite NPs in this position. In addition, Cantonese classifiers can also be with a possessor NP, which is shown in (11). This is one of the possessive constructions in colloquial Cantonese, and the function of the classifier in this position is comparable to that of the linking particle *ge*, although the latter is more formal and makes the phrase ambiguous in number (Matthews & Yip 2011: 129). In (12), the classifier *go* is used in the Adjective-CL-N construction. This construction is less frequent and more idiom-like, and only two adjectives can be employed in this construction (viz. *daai* ‘big’ and *sai* ‘small’)³⁵ (Matthews & Yip 2011: 112-113).

3.2 Results of the surveys about Cantonese classifiers and MWs

As described in section 1.4, I collected data on the use of Cantonese classifiers and measure words through elicitation using a questionnaire.³⁶ The main type of the data is grammaticality judgements, although the participants also offered alternative English translations when they thought that a Cantonese sentence should be translated differently. To elicit grammaticality judgements, the participants are asked to indicate how good a Cantonese sentence sound and to choose from four options, namely “excellent”, “good”, “poor” and “unacceptable”.³⁷ The Cantonese data consist of two parts. In the first part, I

³⁵ The Adjective-CL-N sequence usually follows a degree word, such as *hou* ‘very’ and *gam* ‘so’ (Matthews & Yip 2011: 112-113), which I incline to consider as part of this construction. Assuming this is correct, the construction should be written as the “Deg-Adj-CL-N construction” instead.

³⁶ Initially, 14 native speakers of Cantonese agreed to participate in this study. However, six of the responses were very incomplete (i.e. answering less than half of the questions) and were discarded. Another participant speaks Siyi Cantonese, which appears to be rather different from Hong Kong Cantonese (which the other participants speak), and was excluded for that reason. As a result, the number of valid responses was reduced to seven.

³⁷ Afterwards, these options were coded numerically, and a grammaticality point was calculated for each sentence. The procedure of this transformation is as follows: Excellent → 3, Good → 2, Poor → 1, Unacceptable → 0.

verify that Cantonese classifiers are used in the six constructions mentioned in section 3.1, and I determine whether the measure words can be used in these same constructions. In the second part, I examine the acceptability and interpretation of Cantonese classifiers and measure words in a few more complex constructions.

The results of the first part of the data indicate that there are consistent but subtle differences between Cantonese classifiers and measure words in their acceptability in pre-verbal bare Cl/MW-N phrases and in the modified Cl/MW-N phrases. Both classifiers and measure words are acceptable in post-verbal Cl/MW-N phrases and in Num-Cl/MW-N phrases. The acceptability of classifiers will be compared with that of measure words in each context, starting with the bare Cl/MW-N phrases.

It is uncontroversial that count classifiers can occur in bare Cl-N phrases in pre-verbal positions in Cantonese (Cheng & Sybesma 1999; Matthews & Yip 2011; Li & Bisang 2012), but it is less clear whether that applies to Cantonese measure words. The data that I collected suggest that the classifiers and measure words are somewhat different in this regard. Two examples with pre-verbal bare Cl/MW-N phrases are shown below:

- (13) 間 屋 好 靚 呀 !
Gaan uk hou leng aa³
CL house very beautiful SFP
'The house is so beautiful!'
- (14) 杯 茶 幾 香 嘅
Bui caa gei hoeng ge²
MW=_{cup} tea quite fragrant SFP
'That cup of tea smells quite good.'

In (13), the bare Cl-N phrase *gaan-uk* 'the house' is in the subject position and only has a definite reading. In terms of acceptability, all the participants considered it excellent (grammaticality point = 3.0). This and other similar sentences confirm that pre-verbal Cl-N phrases are highly acceptable in Cantonese. In (14), the MW-N phrase *bui-caa* 'that cup of

tea' consists of the measure word *bui* 'cup' and the noun *caa* 'tea', and it is also the subject of the sentence. As for its grammaticality, three participants found it "excellent", and four participants considered it "good" (grammaticality point = 2.43). Results of this part of the survey indicate that pre-verbal bare MW-N phrases are acceptable to a considerable extent in Cantonese, albeit somewhat less so than the ones with count classifiers such as (13).

The post-verbal Cl/MW-N phrases in Cantonese exhibit a different pattern. These NPs can be definite or indefinite (Matthews & Yip 2011: 106). The count classifiers and measure words are both commonly used in this construction in Cantonese (Mak & Taam 2011), which is confirmed by my data. Examples of these phrases are given below:

- (15) 你 問咗 個 律師 未 呀 ?
 Nei man-zo go leotsi mei aa³
 you ask-ASP CL lawyer NEG.PERF SFP
 'Have you consulted the lawyer?'
 (16) 食 完 碗 飯 未 呀 ?
 Sik jyun wun faan mei aa¹
 eat finish MW=_{bowl} rice NEG.PERF SFP
 'Have you finished that bowl of rice?'³⁸

In example (15), the Cl-N phrase consists of the classifier *go* and the noun *leotsi* 'lawyer', and it is the post-verbal object in the sentence. Semantically, it is the context that makes this NP definite, and the same NP can have an indefinite reading in the post-verbal position in a different context (Matthews & Yip 2011: 106). Regarding its grammaticality, five participants chose "excellent" and two chose "good" (grammaticality point = 2.71). In (16), the measure word *wun* 'bowl' is followed by the noun *faan* in the MW-N phrase, which is the object of the sentence. It is also definite in meaning due to the context and can be indefinite in other contexts. As for its acceptability, five participants selected "excellent" and two selected "good" (grammaticality point = 2.71).

³⁸ Two of the participants preferred to translate the Cantonese sentence in (16) as 'Have you finished your bowl of rice?', i.e. with the CL-N phrase being semantically possessive.

According to Matthews & Yip (2011: 128-9), there are four possessive constructions in Cantonese, including the Possessor-Cl-NP construction. Whether Cantonese has a corresponding possessive construction with measure words has been unclear. The data I collected confirm that the Possessor-Cl-NP construction is acceptable, whilst the Possessor-MW-NP construction seems somewhat less acceptable. This possessive construction with classifiers or measure words is illustrated below:

- (17) 王生 架車 整好咗 未 呀?
 Wong-saang ga ce zing-hou-zo mei aa³
 Wong-sir CL car fix-ASP-ASP NEG.PERF SFP
 ‘Has Mr. Wong’s car been repaired?’

- (18) 你 樽 汽水 係 乜嘢 味 㗎?
 Nei zeon heiseoi hai matje mei gaa³
 you MW=_{bottle} soft.drink COP what flavour SFP
 ‘What is the flavour of your drink (in the bottle)?’

The subject in (17) is a possessor-Cl-N phrase which simply means “Mr. Wong’s car”. Unlike possessor-*ge*-N phrases which is ambiguous in terms of number and can be definite or indefinite, the meaning of the possessor-Cl-N construction is unambiguously singular and definite (Matthews & Yip 2011: 128-129). For this reason, the possessor-Cl-N phrase in (17) cannot mean “Mr. Wong’s cars” or “a car of Mr. Wong’s”. Regarding its acceptability, four participants considered (17) to be excellent, and two participants chose “good”, while another chose “poor”, hence the average grammaticality point of 2.43.³⁹

In (18), the subject NP is a possessor-MW-N phrase meaning “your soft drink”. The measure word *zeon* ‘bottle’ is appropriate only when the soft drink is in a bottle. This

³⁹ The fact that a participant considered it “poor” is attributable to the stylistic effect of this construction. The possessor-CL-N construction is associated with the colloquial/informal register (Matthews & Yip 2011: 129). Since the data were collected with a questionnaire in the written form, some participants may have thought that it is less acceptable for this construction to occur in written Cantonese.

possessor-MW-N phrase is also definite and singular in meaning. Two participants found the sentence in (18) “excellent”, and five considered it “good”, hence the grammaticality point of 2.29, which is slightly lower than that of the possessor-Cl-N construction.

Another modified Cl-N construction in Cantonese is the Deg-Adj-Cl-N construction, which is relatively infrequent because the adjective in this construction must be one of two size adjectives (Matthews & Yip 2011: 112-113). I included some questions in the survey to determine how acceptable this and the corresponding Deg-Adj-MW-N construction is in Cantonese. Two examples are shown below:

- (19) 咁 大 間 房 幾多 個人 住 呀?
 Gam daai gaan fong geido go jan zyu aa³
 so big CL suite how.many CL person live SFP
 ‘How many people live in such a big suite?’
- (20) 佢 買咗 咁 大 堆 生果 畀 你 食。
 Keoi maai-zo gam daai deoi saanggwo bei nei sik
 3rd.SING buy-ASP such big MW=_{pile} fruit give you eat
 ‘He bought such a big pile of fruit for you to eat.’

The Deg-Adj-Cl-N phrase in (19) is *gam-daai-gaan-fong*, which means ‘such a big suite’ and refers to a specific suite that is obvious from the context. In respect of its acceptability, four participants considered the sentence excellent, one chose “good” and two chose “poor”, hence the grammaticality point of 2.29. In (20), the Deg-Adj-MW-N phrase is *gam-daai-deoi-saanggwo* meaning ‘such a big pile of fruit’. It also refers to a specific pile of fruit, and the adjective *daai* ‘big’ modifies the measure word *deoi* ‘pile’. As for its grammaticality, two participants found it excellent, two participants selected “good” and three chose “poor” (grammaticality point = 1.86). This use of classifiers is considered

idiom-like (Matthews & Yip 2011: 112), which might account for the relatively low grammaticality points of the Deg-Adj-Cl/MW-N construction.⁴⁰

Cantonese classifiers and measure words also co-occur with the demonstratives (i.e. *li* ‘this’ and *go* ‘that’). The classifier in the Dem-Cl/MW-NP construction in Cantonese is obligatory (Matthews & Yip 2011: 107), while the sortal classifier in the Dem-Cl/MW-NP construction in Mandarin is often optional.⁴¹ The interpretation of Dem-Cl/MW-N phrases is definite for obvious reasons. This use of classifiers and MWs is illustrated below.

- (21) 我 仲未 讀過 嗰 本 書。
Ngo zungmei duk-gwo go bun syu
1st.SING NEG read-EXP that CL book
‘I haven’t read that book.’

- (22) 佢 好 鍾意 呢 種 花 嘅。
Keoi hou zungji ni zung fa ge
3rd.SING very like this MW=_{type} flower SFP
‘She likes this kind of flower very much.’

In (21), the Dem-Cl-N phrase is *go-bun-syu* meaning ‘that book’, and it is the object of the sentence. Regarding its grammaticality, three participants selected “excellent” and four selected “good”, hence the grammaticality point of 2.43. In (22), the Dem-MW-N phrase *ni-zung-fa* meaning ‘this kind of flower’ is also the object of the sentence. One of the participants found it excellent, and five of them chose “good”, while one chose “poor”. The average grammaticality point of (22) is therefore 2.0.

⁴⁰ Two Cantonese-speaking attendees at the MOTH Syntax Workshop (April 2025) mentioned that Cantonese classifiers can also occur in two other nominal constructions, viz. Adj-CL-N phrases without degree words and Locative-CL-N phrases. For example, they accept phrases such as *laamsiq-kaa-ce* ‘the blue car’ (literally ‘blue-CL-car’) and *tsinbin-zoeng-toi* ‘the table in front’ (literally ‘in.front-CL-table’).

⁴¹ This construction is found in several other Sinitic languages. In some other classifier languages, such as Burmese, classifiers and demonstratives do not co-occur (Jenny & Hnin Tun 2016: 52).

The results of the first part of the survey indicate that Cantonese classifiers and measure words behave differently in the constructions in which they only have a definite reading. These constructions are bare Cl/MW-N phrases in the pre-verbal position, the possessor-Cl/MW-N construction, the Adj.-Cl/MW-N construction and the demonstrative-Cl/MW-N construction. The difference in acceptability between classifiers and measure words is not enormous, but it occurs consistently across different contexts. In post-verbal Cl/MW-N phrases, which can be indefinite or definite, classifiers and measure words are about equally acceptable; the same is arguably true of Num-Cl/MW-N phrases, which are semantically indefinite (Cheng & Sybesma 1999: 528). In other words, Cantonese classifiers and measure words exhibit somewhat different syntactic behaviour when the NP must be semantically definite. This could be considered evidence that the classifiers and measure words are two separate categories in Cantonese, which will be discussed in more detail in section 3.4. The results are summarized below in the form of grammaticality points.

Table 3: Average grammaticality points of five constructions with classifiers or MWs

Category construction	Pre-verbal Cl/MW-N	Post-verbal Cl/MW-N	Possessor-Cl/MW-N	(Deg)-Adj.-Cl/MW-N	Demonstrative-Cl/MW-N
Classifier	2.77 ⁴²	2.62 ⁴³	2.43	2.29	2.43
MW	2.48	2.64	2.29	1.86	2.0

In addition, there are some minor findings from the first part of the survey. The use of the collective (or plural) classifier *di* ‘some’ was examined, and its compatibility with both count nouns (e.g. *sailou* ‘child’) and uncountable nouns (e.g. *hyut* ‘blood’) is confirmed. Another finding is that the Dem-*jat*-Cl/MW-NP construction⁴⁴ is less grammatical than its simpler counterpart without the numeral *jat* ‘one’. Cantonese is more similar to Northern Wu than to Mandarin in this respect (Li 2013: 281-282).

⁴² An outlier has been excluded. With the outlier, the grammaticality point would be 2.64.

⁴³ An outlier has been excluded. With the outlier, the grammaticality point would be 2.43.

⁴⁴ Note that a demonstrative-Cl/MW-N phrase can only be singular in meaning even if it does not contain the numeral *jat* ‘one’.

In the second part of the survey, I examine some more complex constructions in which classifiers or measure words are expected to occur in Cantonese. I aim to determine how classifiers and MWs are used with relative clauses, whether the Num-Adj-Cl/MW-N construction is felicitous, and how Cl/MW-N-*tung*-Cl/MW-N⁴⁵ phrases are used.

First, classifiers occur in one of the relative clause constructions in Cantonese, co-occurring with a demonstrative between the relative clause and the head noun (Matthews & Yip 2011: 328-330). It is worth noting that the relative clauses in Cantonese are pre-nominal, hence the constituent order RelCP-Dem-CL-N, which is different from English NPs with relative clauses. My data confirm that the RelCP-Dem-MW-NP construction is mostly acceptable as well, like its counterpart with classifiers. The following examples show how classifiers and measure words are used with relative clauses in Cantonese:

(23) 佢 寫 嗰 部 小說 好 有意思。

Keoi se go bou siusyut hou jaujisi⁴⁶

3rd.SING write that CL novel very interesting

‘The novel that he wrote is very interesting.’

(24) 我 飲 嗰 樽 汽水 太 甜 啦。

Ngo jam go zeon heiseoi taai tim laa

1st.SING drink that MW=_{bottle} soft.drink too sweet SFP

‘That bottle of soft drink that I drank was too sweet.’

As we see in (23), a demonstrative-classifier sequence is inserted to connect the relative clause to the head noun *siusyut* ‘novel’ in this type of relative clause construction. Among the participants, three considered (23) excellent, three found it good, and one found it poor in acceptability, and its grammaticality point is therefore 2.29. In (24), the same relative clause construction contains a measure word instead of a classifier, and the demonstrative-MW sequence also connect the relative clause to the head noun. The

⁴⁵ The morpheme *tung* is the coordinating conjunction meaning ‘and’ in Cantonese.

⁴⁶ The NP with the relative clause is underlined in (23) and (24).

participants' grammaticality judgments are the same as those for (23), and the average grammaticality point is again 2.29. In short, the native speakers of Cantonese generally find RelCP-Dem-Cl/MW-N phrases acceptable.

Cantonese has another relative clause construction, which does not contain any classifiers or measure words. The linking particle *ge* is used in lieu of the demonstrative-classifier sequence to connect the relative clause to the head noun (Matthews & Yip 2011: 326-328). The interpretations of these two relative clause constructions are very similar, except that RelCP-*ge*-N phrases can be plural while RelCP-Dem-Cl/MW-N phrases cannot.⁴⁷ The use of this relative clause construction with the particle *ge* is illustrated below:

- (25) 識 我 嘅 人
 sīk ngóh ge yànn
 know me that people
 'People that know me'
 (Matthews & Yip 2011: 328)

In (25), the relative clause is again pre-nominal and is followed by the linking particle *ge*, which can also be used with possessor NPs and attributive adjectives (ibid. 328). Semantically, Cantonese RelCP-*ge*-N phrases can be singular, plural or mass with different nouns and in different contexts, and (25) is generally interpreted as plural. Note that this is different from the RelCP-Dem-Cl/MW-NP construction which must be singular in meaning. For instance, the NP in (23) refers to exactly one novel, and the NP in (24) refers to one bottle of soft drink.

Since relative clauses with the particle *ge* and bare Cl-N phrases are both found in Cantonese, I examined the possibility that they might combine to form another type of relative clause construction, namely the (hypothetical) RelCP-*ge*-Cl-NP construction. My data indicate that this construction is disfavoured, although there is some variation among the participants. Two sentences with this construction are shown below:

⁴⁷ Since Cantonese relative clauses are "essentially restrictive in function" (Matthews & Yip 2011: 327), these two constructions do not differ in that regard.

(26) 佢 寫 嘅 部 小說 好 有意思。
Keoi se ge bou siusyut hou jaujisi
 3rd.SING write LP CL novel very interesting
 ‘The novel that he wrote is very interesting.’

(27) 我 飲 嘅 樽 汽水 太 甜 啦。
Ngo jam ge zeon heiseoi taai tim laa
 1st.SING drink LP MW=_{bottle} soft.drink too sweet SFP
 ‘That bottle of soft drink that I drank was too sweet.’

In (26), the morphemes that occur between the pre-nominal relative clause and the head noun are the linking particle *ge* and the classifier *bou*. There is much variation in the participants judgements: one participant found it “excellent”, two considered it “good”, two chose “poor” and another two chose “unacceptable”. The grammaticality point is 1.29.

Likewise, the relative clause in (27) is connected to the head noun using the particle *ge* and measure word *zeon* ‘bottle’. Example (27) only differs from (24) in the use of the particle *ge* instead of the demonstrative *go* ‘that’. Three participants considered (27) “good”, three deemed it “poor”, and one found it “unacceptable”. The grammaticality point is, therefore, 1.29. In short, the hypothetical RelCP-*ge*-Cl/MW-NP construction is not quite grammatical in Cantonese and is considerably less so than the RelCP-Dem-Cl/MW-NP construction.⁴⁸ The choice between a classifier and a measure word does not seem to affect the acceptability (or unacceptability) of either construction.

⁴⁸ Another logically possible construction containing a relative clause is the RelCP-CL/MW-NP construction, which was briefly mentioned by Matthews & Yip (2011:329). In this case, only the classifier (or measure word) is inserted between the relative clause and the head noun. Based on the incomplete data that I have on this construction, the classifier *go* (for human nouns) and the plural classifier *di* are felicitous in this construction, while MWs and other classifiers seem to be poor in acceptability. This will need to be examined more thoroughly in the future.

Second, I investigate the possibility of inserting a size adjective in the pre-classifier position in Num-Cl/MW-N phrases. In Mandarin, a few dimensional adjectives can be used in the Num-Adj-Cl/MW-NP construction, and they are compatible with the group-partition-container MWs and nearly half of the individual classifiers⁴⁹ (Li 2013: 57-59). According to my data, the Num-Adj-MW-NP construction is generally grammatical in Cantonese,⁵⁰ while the Num-Adj-Cl-NP construction is somewhat less acceptable. The use of classifiers and measure words in this context is illustrated in the examples below:

- (28) 我 飲咗 一 大 碗 湯 啦。
 Ngo jam-zo jat daai wun tong laa
 1st.SG drink-PERF one big MW_{-bowl} soup SFP
 ‘I have eaten a big bowl of soup.’

- (29) 我 食咗 一 大 隻 雞髀 啦。
 Ngo sik-zo jat daai zek gaibei laa
 1st.SG eat-PERF one big CL chicken.thigh SFP
 ‘I have eaten a big chicken thigh.’

In (28), the object of the sentence is the Num-Adj-MW-N phrase *jat-daai-wun-tong* which means ‘one big bowl of soup’. Note that the adjective *daai* ‘big’ is in front of the MW *wun* ‘bowl’ rather than its usual position between the measure word and the noun. Four participants considered it excellent; two participants chose “good”, and one chose “poor”, hence the grammaticality point of 2.43.⁵¹

In (29), the Num-Adj-Cl-N phrase *jat-daai-zek-gaibei* ‘a big chicken thigh’ is also the object of the sentence, and it also contains the adjective *daai* ‘big’ in the same position. The main difference between (28) and (29) is that the latter contains the count classifier *zek*

⁴⁹ In addition, the head noun must be concrete and must not refer to humans.

⁵⁰ Conditions like the ones found in Mandarin also apply to this construction in Cantonese.

⁵¹ In addition to container MWs, group MWs are used in this construction as well, e.g. *jat-daai-pang-jan* ‘a big crowd of people’ (Mak & Taam 2011: 333).

rather than a measure word such as *wun* ‘bowl’. Regarding its grammaticality, three of the participants found (29) excellent, two found it good, and two considered it poor. The grammaticality point is 2.14. In short, the Num-Adj-Cl-NP construction is not as acceptable as its counterpart with a measure word in Cantonese.

The last construction that I examined in this survey is the “conjoined bare Cl/MW-N” construction that contains the coordinating conjunction *tung* ‘and’. The basic template of these conjoined Cl/MW-N phrases is, therefore, Cl/MW-N-*tung*-Cl/MW-N. I analyse the pre-verbal Cl/MW-N-*tung*-Cl/MW-N phrases and their post-verbal counterparts separately. As described in section 2.3, the corresponding construction in Mandarin is ungrammatical regardless of its position in the sentence.⁵² In Cantonese, the conjoined bare Cl/MW-N construction is largely acceptable, although the degree of acceptability depends on the category (i.e. classifier or MW) and position.

First, Cantonese Cl-N-*tung*-Cl-N phrases are highly acceptable in the pre-verbal position according to my data, and there seem to be contexts in which such NPs are used naturally in Cantonese. An example with this type of phrase is given below:

- (30) 隻 狗 同 隻 貓 喺度 打交。
 Zek gau tung zek maau haidou daagaau
 CL dog and CL cat PROG fight
 ‘The dog and the cat are fighting (with each other).’

In (30), the Cl-N-*tung*-Cl-N phrase meaning ‘the dog and the cat’ is the subject of the sentence. Six of the participants consider (30) excellent, and one find it “good”. Thus, the grammaticality point is 2.86.

Second, I tested the acceptability of MW-N-*tung*-MW-N phrases in the pre-verbal position in Cantonese. (31) is an example of a sentence with this construction.

⁵² In Mandarin, pre-verbal Cl/MW-N-*he*-Cl/MW-N phrases (*he* means ‘and’) are categorically unacceptable. This is expected since pre-verbal bare Cl/MW-N phrases are infelicitous in Mandarin. Post-verbal Cl/MW-N-*he*-Cl/MW-N phrases are arguably more interpretable but are still hardly acceptable.

- (31) 碗 炒飯 同 碟 牛腩 都 好 好味 呀！
Wun caaufaan tung dip ngaunaam dou hou houmei aa
MW_{=bowl} fried.rice and MW_{=saucer} beef.brisket all very tasty SFP
‘The bowl of fried rice and the dish of beef brisket are both very tasty.’

The subject in (31) is the MW-N-*tung*-MW-N phrase which means ‘the bowl of fried rice and the dish of beef brisket’. I only include container measure words in this example, thereby avoiding the problem of a mismatch between different sub-categories. Three of the participants find it “excellent”, two participants consider it “good”, and two think that (31) is “poor” in acceptability. The grammaticality point is 2.14.

Third, I determine the possibility of Cl-N-*tung*-Cl-N phrases occurring in the post-verbal position. My data indicate that such phrases are also grammatical in that position. In addition to its grammaticality, the participants were asked about the meaning(s) of the example as well. The use of post-verbal Cl-N-*tung*-Cl-N phrases is illustrated below:

- (32) 我 買咗 支 筆 同 本 雜誌。
Ngo maai-zo zi bat tung bun zaapzi
1st.SG buy-PERF CL pen and CL magazine
‘I bought a pen and a magazine.’

In (32), the Cl-N-*tung*-Cl-N phrase which has the meaning ‘a pen and a magazine’ and contains the classifiers for *bat* ‘pen’ and *zaapzi* ‘magazine’, is the direct object in the sentence. Five participants find it excellent and two find it good, hence the grammaticality point of 2.71. Regarding the meaning of (32), the participants were not originally provided with the English translation, and they were asked to select all possible meanings from a list of four translations.⁵³ The four possible translations are listed below, each followed by the number of participants that selected the meaning and the corresponding percentage.

⁵³ In fact, there are two more options, namely “This sentence is unintelligible” and “Other (please specify)”. None of the participants chose either of these options.

- ① ‘I bought a pen and a magazine.’ – 7 (100%)
- ② ‘I bought a pen and the magazine.’ – 3 (42.86%)
- ③ ‘I bought the pen and the magazine.’ – 3 (42.86%)
- ④ ‘I bought the pen and a magazine.’ – 1 (14.29%)

Every participant accepted the first translation ‘I bought a pen and a magazine’, in which both NPs that are conjoined by *tung* ‘and’ are indefinite. The second translation, which contains the indefinite NP *zi-bat* ‘a pen’ followed by the definite NP *bun-zaapzi* ‘the magazine’ is considered possible by three participants. The third translation includes two definite NPs connected by *tung* ‘and’, and three participants consider it to be a possible translation of (32). The fourth translation contains the definite NP *zi-bat* ‘the pen’ followed by the indefinite NP *bun-zaapzi* ‘a magazine’, and only one participant think that (32) can be interpreted this way. In short, the participants have a clear preference to interpret the two NPs in (32) as indefinite NPs, and they tend not to interpret the first NP as a definite NP, especially when the second NP is indefinite.

Lastly, I collected data on the use of MW-N-*tung*-MW-N phrases in the post-verbal position. My data suggest that this construction is largely acceptable in Cantonese, although slightly less so than CL-N-*tung*-CL-N phrases in the same position. A sentence containing a MW-N-*tung*-MW-N phrase is shown below in (33).

- (33) 佢 飲咗 杯 茶 同 樽 汽水。
 Keoi jam-zo bui caa tung zeon heiseoi
 3rd.SING drink-ASP MW=_{cup} tea and MW=_{bottle} soda
 ‘He drank a cup of tea and a bottle of soft drink.’

In (33), the object of the sentence is the MW-N-*tung*-MW-N phrase *bui-caa-tung-zeon-heiseoi*, which is often interpreted as ‘a cup of tea and a bottle of soft drink’. This sentence is considered “excellent” by three participants and is “good” according to four. The grammaticality point is therefore 2.43. The English translation of the sentence in (33) was not originally provided to the participants. Instead, they were asked to choose all the

possible meanings of the Cantonese sentence from a list of four translations. Those four translations are shown below, each followed by the number of participants choosing that meaning and the percentage corresponding to that number.

- ① ‘He drank a cup of tea and a bottle of soft drink.’ – 6 (85.71%)
- ② ‘He drank a cup of tea and that bottle of soft drink.’ – 3 (42.86%)
- ③ ‘He drank that cup of tea and that bottle of soft drink.’ – 2 (28.57%)
- ④ ‘He drank that cup of tea and a bottle of soft drink.’ – 1 (14.29%)

Most (i.e. 6 out of 7) participants accepted the first translation where both MW-N phrases are interpreted as indefinite NPs. Three of the participants considered the 2nd translation (where the first NP is indefinite and the second is definite) to be a possible translation of (33). The third translation, in which both of the conjoined NPs are definite, is accepted by two participants. Only one participant thinks that the Cantonese sentence in (33) can be translated with the 4th sentence, in which the first MW-N phrase is definite while the second one is indefinite. In short, the participants prefer to interpret the first NP as an indefinite NP (i.e. ‘a cup of tea’), and most participants do not interpret the first NP as a definite NP and the second NP as indefinite. This pattern is very similar to what we have seen in the previous paragraphs on the interpretation of (32).

A minor finding from the second part of the survey is that the acceptability of Num-Cl-*tung*-Num-Cl phrases is not particularly high in Cantonese. The grammaticality point is no greater than 2.14, even when the two conjoined [Num-Cl] phrases quantify the same noun, and the meaning is made clear with the adverb *fanbit* ‘respectively’. Note that the Num-Cl-*tung*-Num-Cl phrases are less acceptable than the Cl-N-*tung*-Cl-N phrases.

To conclude, three major findings emerge from the second part of the survey. First, the RelCP-Dem-Cl/MW-NP construction can be used to form a complex NP with a relative clause in Cantonese, and the RelCP-*ge*-Cl/MW-NP construction is much less acceptable. Second, Num-Adj-MW-N phrases with a size adjective is generally considered grammatical, and Num-Adj-Cl-N phrases are somewhat less acceptable. Lastly, both pre-verbal and post-verbal Cl-N-*tung*-Cl-N phrases are highly acceptable in Cantonese, while MW-N-*tung*-MW-N

phrases are less acceptable, esp. those in the pre-verbal position. Both types of conjoined NPs are semantically definite in the pre-verbal position, while in the post-verbal position they are most often interpreted as indefinite NPs. Discussion of these results can be found in section 3.4, after I present the Northern Wu data in the next section.

3.3 Results of the surveys about Northern Wu classifiers and MWs

I also designed a questionnaire to collect grammaticality data from native speakers of Northern Wu. The methodology of data collection is essentially the same as the one used for collecting the Cantonese data. One notable difference is that the data are about two dialects, namely Ningbo Wu and Shanghai Wu, which are closely related to each other and are mutually intelligible. The data come from two native speakers of Ningbo Wu (i.e. a participant and me) and one native speaker of Shanghai Wu. Data from previous works on Northern Wu will be also discussed when appropriate. The focus will be on Ningbo Wu, and other dialects of Northern Wu will be discussed mainly for comparison purposes.

As I described in section 3.1, the classifiers or measure words in Num-CL/MW-N phrases are obligatory in Ningbo Wu. This is also the case in the dialects of Shanghai and Fuyang⁵⁴ (Zhu 2006; Li 2013: 236), and they do not differ from Mandarin or Cantonese in this regard. In addition, Northern Wu classifiers and measure words are used in several other contexts. I shall start with the bare CL/MW-N phrases and their plural forms.

The post-verbal (bare) Cl/MW-N phrases are known to be grammatical in Shanghai Wu and in Fuyang Wu (Zhu 2006: 77; Li 2013: 242). My data confirm that this construction is felicitous in Shanghai Wu as well as Ningbo Wu. The following examples in Ningbo Wu illustrate the use of classifiers and measure words in this position.

- (34) 我 昨末 买了 梗 鱼。 [Ningbo Wu]
 Ngo zoqmaq ma-leq kan ng
 I yesterday buy-ASP CL fish

⁵⁴ Fuyang Wu is a dialect of Northern Wu which has been described in Li (2013). Its relation to Ningbo Wu seems to be somewhat more distant than that between Ningbo Wu and Shanghai Wu.

‘I bought a fish yesterday.’

- (35) 其 有 事体， 吃了 杯 茶 就 去 嘞。
Ji yeu zythi, chioq-leq pei dzo jiu chi le
3rd.SING have business eat-ASP MW=_{cup} tea then go.away SFP
‘He had other things to do, so he had a cup of tea and went away.’

In (34), the bare Cl-N phrase *kan-ng* ‘a fish’ follows the verb *ma* ‘to buy’ as its direct object, and it can only be interpreted as an indefinite NP. Both native speakers of Ningbo Wu consider (34) to be perfectly acceptable, and the grammaticality point⁵⁵ is therefore 3.0. In (35), the bare MW-N phrase *pei-dzo* ‘a cup of tea’ follows the verb *chioq* ‘to eat or drink’, and it is indefinite in meaning as well. Both native speakers find (35) grammatical, and the grammaticality point is 3.0. The Shanghai Wu sentences corresponding to (34) and (35) are perfectly acceptable according to the Shanghainese-speaking participant. In other words, post-verbal Cl/MW-N phrases are felicitous in Northern Wu and are semantically indefinite. Classifiers and measure words behave similarly in this construction.

It is uncontroversial that pre-verbal bare Cl-N phrases are grammatical in Northern Wu (Qian 1997; Zhu 2006; Li 2013).⁵⁶ However, whether the pre-verbal MW-N phrases are acceptable is less clear in previous works. The data that I collected confirm that pre-verbal Cl-N phrases are highly acceptable in Ningbo Wu and Shanghai Wu. On the other hand, the pre-verbal MW-N phrases are overall less acceptable, and its acceptability varies according to speaker and sub-category. This construction is shown in the examples below:

- (36) 梗 鱼 死掉 嘞。 [Ningbo Wu]
kan ng shi-dio le

⁵⁵ The (average) grammaticality point is a value between 0 and 3 calculated using the following formula: Perfect → 3, Good → 2, Poor → 1, Completely unacceptable → 0. Grammaticality points are calculated to facilitate analysis and comparison between different dialects of Wu as well as between Cantonese and Northern Wu.

⁵⁶ This construction is also found in some dialects of Southern Wu (e.g. Jinhua Wu) (Qian 1992: 1011-1012).

CL fish die-ASP SFP

‘The fish is dead.’

(37) 杯 茶 冷掉 嘞。

pei dzo lan-dio le

MW_{=cup} tea cold-ASP SFP

‘The cup of tea is cold.’

In (36), the bare Cl-N phrase *kan-ng* ‘the fish’ is the subject of the sentence and is in the pre-verbal position. Semantically, it can only be interpreted as a definite NP, and it cannot mean ‘a fish’ or ‘fish (in general)’. The participants found (36) perfectly acceptable, and the average grammaticality point is 3.0 for both dialects of Northern Wu. Strictly speaking, pre-verbal Cl-N phrases can also be the topic of a sentence or the fronted object in “disposal constructions”⁵⁷ (Li & Bisang 2012). It has been shown that the exact function in the sentence (i.e. topic, subject or fronted object) does not affect the acceptability or interpretation of Northern Wu pre-verbal Cl-N phrases (Li & Bisang 2012: 339-340; Zhu 2006: 77), and my data confirms this for Ningbo Wu.

In (37), the MW-N phrase *pei-dzo* ‘the cup of tea’ precedes the predicate since it is the subject of the sentence. The grammaticality of (37) varies greatly from speaker to speaker: one speaker of Ningbo Wu considered it perfect, while the other found it poor; the Shanghainese version of this sentence is unacceptable according to the Shanghai-Wu-speaking participant. In total, five sentences with pre-verbal MW-N phrases are examined in this survey, and the average grammatical point is 1.4 (for Ningbo Wu). There seems to be some inter-speaker variation in the acceptability of this construction: the grammatical point is 1.8 for one speaker of Ningbo Wu, and it is 1.0 for the other; the grammatical point is 1.0 for the Shanghainese-speaking participant. Since the measure words can be divided

⁵⁷ A disposal construction in a Chinese language allows the speaker to use the OV word order rather than the basic VO order, and it consists of three parts: particle (i.e. object marker) – object – verb (or VP) (Li & Bisang 2012: 339). The basic template of the disposal constructions in the two dialects of Northern Wu are *taq*-O-V (Ningbo) and *nae/nau*-O-V (Shanghai) (cf. *ba*-O-V in Mandarin) (Zhu 2006).

into different subcategories (such as container MWs, group MWs, and partition MWs), the results have also been analysed according to that criterion. The pre-verbal MW-N phrases with group MWs (e.g. *tei* 'group, pile') appear to be the most acceptable in Ningbo Wu, with a grammaticality point of 2.0.⁵⁸ Those with container MWs have a grammaticality point of 1.5. The grammaticality point for the pre-verbal MW-N phrases with partition MWs is 0.5. However, the subcategory does not seem to be a major factor in Shanghai Wu, and it is the exact function of the phrase that plays an important role. If the MW-N phrase is the fronted object in a disposal construction, then it is considered felicitous. This is illustrated below:

- (38) 伊 拿 瓶 饮料 攢脱 了。 [Shanghai Wu]
 Yi nau bin inliau guaetheq leq
 3rd.SING OM MW=_{bottle} beverage throw.away ASP
 ‘He threw away the bottle of drink.’

In (38), the MW-N phrase *bin-inliau* ‘the bottle of drink’ is the fronted object, and it follows the object marker *nau* (literally ‘take’ or ‘bring’) in the disposal construction. It is worth noting that the clausal word order in (38) is SOV, and that the thematic role of this MW-N phrase is the same as that of the post-verbal object MW-N phrase in (35). According to the Shanghai-Wu-speaking participant, (38) is grammatical, while (37) and other similar sentences (with MW-N phrases in the subject position) are unacceptable. The bare MW-N phrase in (38) is interpreted as a definite NP and does not mean ‘a bottle of drink’.

In short, the acceptability of pre-verbal MW-N phrases vary according to different factors in Ningbo Wu and Shanghai Wu. In Ningbo Wu, the acceptability is mainly affected by the subcategory of the MW and by inter-speaker variation. In Shanghai Wu, the main factor is the function of the phrase: bare MW-N phrases that are fronted objects are acceptable, while those that are subjects are illicit. Overall, the level of acceptability of pre-verbal MW-N phrases is considerably lower than that of post-verbal MW-N phrases.

⁵⁸ Note that the only example with a pre-verbal MW-N phrase in Qian (1997: 99) contains the group MW *deq* ‘pile’.

All the bare CL/MW-N phrases discussed thus far are unambiguously singular in number. They can be pluralized in a few ways in Northern Wu. The quantifier *lian* ‘a few’ or the quantifier *ci* ‘several’⁵⁹ can be inserted to the left of the classifier (or measure word), or the classifier (or measure word) can be replaced by the plural classifier *ngae*. The use of these plural forms is shown in the following example.

- (39) 我 买了 两 件 衬衫 / 几 件 衬衫 / 眼 衬衫。
 Ngo ma-leq lian ji tshensae / ci ji tshensae / ngae tshensae
 I buy-ASP QUANT CL shirt QUANT CL shirt CL.PL shirt
 ‘I bought some shirts.’ [Ningbo Wu]

In (39), the Quant-CL-N phrase (or *ngae*-N phrase) is in the post-verbal position, and all three are felicitous. The quantifiers still have some semantic content, and the number of the first two NPs should probably be labelled “paucal” rather than “plural”. The *lian*-CL-N⁶⁰ phrase in (39) refers to two, three or four shirts, while the *ci*-CL-N phrase can be used if the number of shirts is between two and ten. The true plural form of the CL-N phrase *ji-tshensae* ‘a shirt’ is the NP *ngae-tshensae*, with the plural classifier *ngae* followed by the noun ‘shirt’, and it can refer to any number of shirts above one.

In the pre-verbal position, the *ci*-CL/MW-N phrases and *ngae*-N phrases are always accompanied by a demonstrative⁶¹ in Ningbo Wu, and those NPs with demonstratives correspond to both *those shirts* and *the shirts* in English. However, this does not seem to be true of *lian*-CL/MW-N phrases, even though Demonstrative-*lian*-CL/MW-N phrases are felicitous in Northern Wu. Li (2013: 258) and Qian (1997: 98-99) point out that *lian*-CL/MW-N phrases are used in the pre-verbal position with a definite meaning (in Fuyang

⁵⁹ The quantifier *ci* ‘several’ can also be used as an interrogative word ‘how many’. The interrogative use of this morpheme is not discussed here.

⁶⁰ The quantifier *lian* ‘a few’ is etymologically related to the numeral *lian* ‘two’, and it can be construed as a less exact version of the latter. The same change happened to their cognate *liang* ‘two’ in Mandarin.

⁶¹ The main demonstrative in Ningbo Wu is *kiq* ‘this, that’ (cf. French *ce* ‘this, that’). The demonstrative *daon* ‘this’ is mostly used by elderly speakers and is rare in the speech of the younger speakers of Ningbo Wu.

Wu and Shanghai Wu respectively). My data indicate that this use of pre-verbal *lian*-CL/MW-N phrases is also found in Ningbo Wu. An example in Ningbo Wu with a definite *lian*-CL-N phrase is given below:

- (40) 两 件 衬 衫 还 没 燥。
Lian ji tshensae waq maq sau
Quant CL shirt PARTICLE NEG dry
'The shirts are not dry yet.'
- [Ningbo Wu]

In (40), the *lian*-CL-N phrase ‘the shirts’ is the subject and precedes the predicate. It is the most precise translation of the English NP *the shirts* in that position, if the number of shirts is no greater than five. The participants all found this construction grammatical, and the grammaticality point of (40) and that of a similar sentence in Shanghai Wu are both 3.0. Regarding the meaning of this *lian*-CL-N phrase, one speaker of Ningbo Wu believes that it could refer to five shirts, while four is the maximum according to the other speaker. The Shanghainese-speaking participant strongly prefers the exact (numeral) meaning of *lian* (i.e. ‘two’), which might be due to a misplaced stress on the quantifier *lian* rather than true dialectal difference. According to my data, the pre-verbal *lian*-MW-N phrases are equally acceptable as those with classifiers, and their interpretations are also very similar.

Northern Wu classifiers are also employed in a few more complex constructions, known as “modified Cl+N” or “non-bare Cl+N” phrases (Li 2013: 275). The modifier in a modified Cl-N phrase in Northern Wu can be a possessor noun, an adjective, a relative clause or a demonstrative, and the modifier is followed by the classifier that it co-occurs with (Li 2013; Qian 1997). On the other hand, it seems unclear whether measure words can occur in similar constructions (i.e. modified MW-N phrases). With this part of the data, I verify whether all these modified Cl-N constructions occur in Ningbo Wu, and I examine the acceptability of modified MW-N phrases. My data about the modified Cl/MW-N phrases with possessor NPs will be presented first.

There are five possessive constructions⁶² in Ningbo Wu, and classifiers are used in two of them, namely possessor-Cl-N and possessor-Dem-Cl-N. Since the latter is relatively infrequent and is semantically closer to *that object of someone's* than to *someone's object*, I focus on the possessor-Cl-N construction here. Sentences with possessor-Cl-N phrases and possessor-MW-N phrases were included in the survey, and two examples are shown below:

- (41) 阿姐 部 汽车 是 白颜色 喔。
[Ningbo Wu]
Acia bu chitsho zy baq-ngaeseq gho
Elder.sister CL car COP white-colour SFP
'My elder sister's car is white.'
- (42) 依 瓶 饮料 是 啥 味道 啦?
Nou bin inlio zy so midau la
you MW=bottle soft.drink COP what flavour SFP
'What is the flavour of your drink (in the bottle)?'

In (41), the possessor noun is *acia* ‘elder sister’, and the possessed noun is *chitsho* ‘car’. These two nouns are linked together using the classifier *bu*, which is the classifier for cars in this and other contexts. In Northern Wu, the classifier in a possessive construction only categorizes the possessed noun and not the possessor or the relation between the two nouns, and the term for this type of classifiers is “possessed classifiers” (Aikhenvald 2000: 125-126). This possessor-Cl-N phrase is definite and singular in meaning, just like its English counterpart in the translation. According to my data, possessor-Cl-N phrases such as (41) are highly acceptable in both dialects of Northern Wu: the grammaticality point is 2.75 in Ningbo Wu and 3.0 in Shanghai Wu.

⁶² The five possessive constructions are as follows: Possessor-N, Possessor-*laq*-N, Possessor-LP-N, Possessor-Cl-N, Possessor-Dem-Cl-N. The Possessor-N construction is mostly used with nouns denoting family members or body parts. The possessive marker *laq* is used with singular pronouns (with some further phonetic changes) and with some nouns denoting humans. Possessor-LP-N phrases (with the linking particle *ghoq*) and Possessor-Cl-N are the most frequent possessive constructions in Ningbo Wu.

In (42), the possessor is the pronoun *nou* ‘you’, and the possessed noun is *inlio* ‘soft drink’. The measure word *bin* ‘bottle’ occurs between the possessor and the possessed and connects the two NPs. The sentence in (42) can be uttered when there is one specific bottle of drink in the context. Sentences with possessor-MW-N phrases (including (42)) have a grammaticality point of 2.25 for the Ningbo Wu speakers and 2.0 for the Shanghai Wu speaker. Note that the level of acceptability of possessor-MW-N phrases is somewhat lower than that of possessor-Cl-N phrases.

Another type of modified Cl/MW-N phrases is the Adjective-Cl/MW-N construction. Li (2013: 280) mentions that classifiers can be employed between an adjective and a noun to form a definite NP in Fuyang Wu.⁶³ The acceptability of this construction and the related Adj-MW-N construction in Northern Wu is examined with this part of the data. The two examples below contain an Adj-Cl-N phrase and a Adj-MW-N phrase respectively.

(43) 黄黄 部 汽车 是 侬个 伐? [Ningbo Wu]

Waonwaon bu chitsho zy nou-ghoq va

Yellow CL car COP you-POSS Q⁶⁴

‘Is the yellow car yours?’

(44) 黄黄 瓶 饮料 是 啥 味道?

Waonwaon bin inlio zy so midau

yellow MW=_{bottle} soft.drink COP what flavour

‘What is the flavour of the yellow drink (in the bottle)?’

In (43), the subject NP which means ‘the yellow car’ contains three words, namely the adjective *waonwaon* ‘yellow’, the classifier *bu* and the noun *chitsho* ‘car’. The classifier

⁶³ In Fuyang Wu, the modification marker /kə/ occurs obligatorily between the adjective and the classifier in this construction (Li 2013: 284). In Ningbo Wu, no modification markers can be inserted in the Adj-Cl-N construction.

⁶⁴ The possessive marker following the pronoun ‘you’ can be deleted, and a Ningbo Wu speaker prefers to omit it.

connects the adjective to the head noun and makes the phrase definite and singular.⁶⁵ Both (43) and the corresponding sentence in Shanghai Wu are considered perfectly acceptable by all the participants. The grammaticality points for Adj-Cl-N phrases are 3.0 in both dialects of Northern Wu. In (44), it is the measure word *bin* ‘bottle’ that connects the adjective ‘yellow’ to the head noun ‘soft drink’, and this phrase refers to one specific bottled drink that is salient in the context. Regarding the acceptability of (44), one native speaker of Ningbo Wu found it “perfect”, and the other selected “good”. The average grammaticality points for Adj-MW-N phrases are 2.0 in Ningbo Wu and 3.0 in Shanghai Wu. Regarding the interpretation of (44), the adjective ‘yellow’ can refer to the colour of the bottle or the drink according to the participants, although the latter reading (i.e. ‘yellow drink’) is preferred if the speaker and the interlocutor know the colour of the liquid (e.g. when it is visible).

It is worth noting that the Adj-Cl/MW-N construction is compatible with a relatively large set of adjectives in Northern Wu, such as *dou* ‘big’, *yu* ‘round’ and *waon* ‘yellow’. As a result, this construction is more frequent than its Cantonese counterpart, in which only two adjectives (viz. *daai* ‘big’ and *sai* ‘small’) can be used (Matthews & Yip 2011: 112). However, there are also adjectives (e.g. *ko* ‘fake’) that cannot occur in this construction in Northern Wu. Exactly which ones are compatible with this construction remains an open question and will need to be determined in a future study on Northern Wu.

Another possible modifier in the modified Cl/MW-N phrases is a relative clause. Li (2013: 284) and Qian (1997: 99) mentioned the existence of modified Cl-N phrases with relative clauses in Fuyang Wu and Shanghai Wu respectively. Its counterpart with measure words (i.e. RelCP-MW-N phrases) has not been mentioned in previous works on Northern Wu. To fill this lacuna, I collected some data and examined the acceptability of RelCP-Cl-N phrases and RelCP-MW-N phrases in Ningbo Wu and Shanghai Wu. These types of phrases are illustrated in the examples below.

⁶⁵ Another common way of forming NPs with adjectives in Northern Wu is by using a modification marker (AKA linking particle) such as *ghoq* in Ningbo Wu and *geq* (or *gheq*) in Shanghai Wu (Qian 1997; Zhu 2006). The modified NPs of this type are more ambiguous in Ningbo Wu, because they can be definite, indefinite and generic.

(45) 我 斩 肉骨头 把刀 寻弗着 嘞。 [Ningbo Wu]

Ngo tsae nioq-kuaqdeu po tau zhin-vaq-dzaq le

1st.SING cut meat-bone CL knife find-NEG-RES SFP

'The knife that I use to chop bones is missing.'

(46) 我 泡茶瓶 热水 来 阿里 啦?

Ngo phau dzo bin nyiq-syu le aqli la?

1st.SING steep tea MW=_{bottle} hot-water be.LOC where SFP

'Where is the bottle of hot water that I'm going to use to make tea?'

In (45), the relative clause (which is underlined) is followed by the classifier *po* and the head noun *tau* 'knife'. The choice of classifiers only depends on the head noun, and the whole NP must be definite and singular in meaning.⁶⁶ Both native speakers of Ningbo Wu find (45) perfect. The grammaticality points for the RelCP-Cl-N phrases (including (45)) are 2.5 in Ningbo Wu and 2.75 in Shanghai Wu.⁶⁷

In (46), the relative clause (underlined) is followed by the measure word *bin* 'bottle' and the head noun *nyiqsyu* 'hot water'. According to the participants who accept (46), it refers to one specific bottle of hot water. There is some inter-speaker variation regarding the acceptability of this sentence: one speaker of Ningbo Wu found it unacceptable, while the other considered it good. The average grammaticality points for the RelCP-MW-N phrases (including (46)) are 1.25 in Ningbo Wu and 2.0 in Shanghai Wu.

There are some minor findings related to the modified Cl/MW-N phrases in Ningbo Wu. First, the modifier can also be demonstratives, and Dem-Cl-N phrases and Dem-MW-N

⁶⁶ Another way to make NPs with relative clauses in Ningbo Wu is to put the modification marker *ghoq* in the same position as the classifier. Those NPs with *ghoq* are ambiguous: they can be definite or indefinite, singular or plural.

⁶⁷ One of the Ningbo Wu speakers seems to prefer relative clauses that are instruments (such as the one in (45)) to those that are objects in this construction. It is unclear whether this is merely an idiolectal preference and whether anything significant lies behind it.

phrases are grammatical in Ningbo Wu.⁶⁸ Second, the classifier can be omitted in the Dem-Cl-NP construction if the head noun denotes human beings. For example, the Dem-Cl-N phrase *kiq-goq-nin* ‘that person’ and the Dem-N phrase *kiq-nin* ‘that person’ are both felicitous in Ningbo Wu. The variants with classifiers are sometimes preferable, esp. with longer NPs. With other types of nouns (e.g. animals, objects), deletion of the classifier generally makes the phrase poor in acceptability, and often unacceptable. Third, inserting the numeral *iq* ‘one’ between the demonstrative and the classifier (or measure word) in a Dem-Cl/MW-N phrase makes it poor in acceptability (cf. Mandarin Dem-*yi*-Cl/MW-N phrases, which are grammatical). Both speakers of Ningbo Wu consider it odd and redundant, even though its meaning is still interpretable.⁶⁹ Lastly, when the acceptability of Adj-MW-N and RelCP-MW-N phrases is low, those phrases are transformed into Dem-MW-Adj-N phrases or RelCP-Dem-MW-N phrases respectively with minimal change in meaning.

The major results of this part of the survey are summarized in Table 4. For both dialects of Northern Wu, the grammaticality point of each construction is indicated.

Table 4: The acceptability of five constructions with classifiers or MWs in Northern Wu

Dialect	Category	Pre-V bare Cl/MW-N	Post-V Cl/MW-N	Possessor- Cl/MW-N	Adjective- Cl/MW-N	Relative C- Cl/MW-N
Ningbo Wu	Classifier	3.0	3.0	2.75	3.0	2.5
	MW	1.4	3.0	2.25	2.0	1.25
Shanghai Wu	Classifier	3.0	3.0	3.0	3.0	2.75
	MW	1.0	3.0	2.0	3.0	2.0

The pre-verbal Cl-N phrases and the modified Cl-N phrases in Northern Wu are definite (Qian 1997; Li 2013). As we see in the table above, they are almost always more

⁶⁸ Two subtypes of measure words are not used with demonstratives, viz. standard measures and temporary MWs.

⁶⁹ According to Li (2013: 280), Dem-*iq*-Cl-N phrases are unacceptable in Fuyang Wu.

acceptable than pre-verbal MW-N phrases and the modified MW-N phrases.⁷⁰ On the other hand, the post-verbal Cl/MW-N phrases are indefinite (Li & Bisang 2012), and there is no difference in acceptability between these two categories.

The remaining part of the survey is about the use of classifiers and measure words in two more complex constructions in Northern Wu, namely the Num-Adjective-Cl/MW-N construction and the Cl/MW-N-*and*-Cl/MW-N construction with a coordinating conjunction (i.e. *taq* in Ningbo Wu and *theq* or *kau* in Shanghai Wu).

Both Num-Adj-Cl-N phrases and Num-Adj-MW-N phrases are found in Mandarin, although the adjective has to be a dimensional adjective and some count classifiers work less well in this construction (Li 2013: 55-61). The Cantonese data presented in the previous section suggest that the Num-Adj-Cl-N phrases are somewhat less acceptable than Num-Adj-MW-N phrases in Cantonese. I collected some data to investigate the acceptability of these types of phrases in Northern Wu, and two examples are given below.

- (47) 小林 天亮饭 吃了 一大只 淡包。 [Ningbo Wu]

Shiolin thinianvae chioq-leq iq-dou-tsaq daepau
Shiolin breakfast eat-ASP one-big-CL steamed.bun
'Xiao Lin ate a big steamed bun for breakfast.'

- (48) 小林 天亮饭 吃了 一大碗 汤饭。

Shiolin thinianvae chioq-leq iq-dou-un thaonvae
Shiolin breakfast eat-ASP one-big-MW_{=bowl} congee
'Xiao Lin ate a big bowl of congee for breakfast.'

In (47), the Num-Adj-Cl-N phrase 'a big steamed bun' is the object in the sentence and contains the size adjective *dou* 'big'. Its meaning is essentially the same as that of the corresponding Num-Cl-Adj-N phrase. Regarding its acceptability, one Ningbo-Wu-speaking

⁷⁰ Here the only exception is the acceptability of Adjective-Cl/MW-N phrases in Shanghai Wu. In this case, the sample size is extremely small, and a difference between classifiers and measure words in this construction may emerge when more data becomes available.

participant considered it “good”, and the other speaker chose “poor”. The grammaticality point of Num-Adj-Cl-N phrases is 1.5 in Ningbo Wu and 1.0 in Shanghai Wu. These speakers all prefer the corresponding Num-Cl-Adj-N phrase, where the adjective *dou* ‘big’ is adjacent to the head noun and follows the classifier *tsaq*.

In (48), the Num-Adj-MW-N phrase ‘a big bowl of congee’ is also the object of the sentence. In this case, the size adjective *dou* ‘big’ modifies the container measure word ‘bowl’ rather than the head noun ‘congee’. Both speakers of Ningbo Wu found it perfectly acceptable. The grammaticality point of Num-Adj-MW-N phrases is 3.0 in Ningbo Wu and 3.0 in Shanghai Wu. In short, Num-Adj-MW-N phrases are more acceptable than Num-Adj-Cl-N phrases in Northern Wu. Note that the Mandarin counterparts of (47) and (48) are both grammatical. In other words, Northern Wu classifiers and measure words and their Cantonese counterparts behave similarly in Num-Adj-Cl/MW-N phrases, and they behave differently from the classifiers and measure words in Mandarin.

As for the conjoined Cl/MW-N phrases, they can be subdivided into four subtypes, namely pre-verbal Cl-N-*and*-Cl-N phrases, pre-verbal MW-N-*and*-MW-N phrases, post-verbal Cl-N-*and*-Cl-N phrases and post-verbal MW-N-*and*-MW-N phrases. The use of the pre-verbal ones will be described first, followed by the post-verbal ones. The acceptability of Cl-N-*and*-Cl-N phrases and that of MW-N-*and*-MW-N phrases in the pre-verbal position turn out to be different. The use of these phrases is illustrated in the examples below.

(49) 只 狗 搭 只 猫 来盖 打相打。 [Ningbo Wu]

Tsaq keu taq tsaq mae leke tanshiantan

CL dog and CL cat PROG.DIST fight.against.each.other

‘The dog and the cat are fighting against each other.’

(50) 块⁷¹ 蛋糕 搭 瓶 饮料 侬 忒 甜 嘞。

Kuo daekau taq bin inlio dze thou die le

MW_{=piece} cake and MW_{=bottle} beverage both too sweet SFP

⁷¹ This morpheme was written like this in the questionnaire. The etymologically correct character for this morpheme is probably 瓜.

‘Both the piece of cake and the bottle of drink are too sweet.’

In (49), the Cl-N-*and*-Cl-N phrase is the subject of the sentence and is followed by the verb. Both Cl-N phrases that are conjoined by *taq* ‘and’ must be interpreted as definite NPs (e.g. ‘the dog’ rather than ‘a dog’). According to the Ningbo Wu speakers, this sentence is perfectly acceptable and can be uttered quite naturally in speech. The grammaticality points for pre-verbal Cl-N-*and*-Cl-N phrases are 3.0 in Ningbo Wu and 3.0 in Shanghai Wu. In (50), the pre-verbal MW-N-*and*-MW-N phrase is the subject, and the two conjoined MW-N phrases are likewise definite (e.g. ‘the piece of cake’ rather than ‘a piece of cake’). Like the simplex MW-N phrases in the pre-verbal position, their level of acceptability is not very high. The grammaticality points for pre-verbal MW-N-*and*-MW-N phrases are 1.5 in Ningbo Wu and 1.0 in Shanghai Wu.

When occurring in the post-verbal position, the conjoined Cl/MW-N phrases are not used in the same way. Two examples with these phrases are given below.

(51) 我 昨末 买了 本 书 搭 支 铅笔。 [Ningbo Wu]

Ngo zoqmaq ma-leq pen syu taq tsy khaepiq

1st.SING yesterday buy-ASP CL book and CL pencil

‘I bought a book and a pencil yesterday.’

(52) 阿拉 买了 袋 橘子 搭 箱 苹果。

Aqlaq ma-leq de ciuiqtsy taq shian binkou

1st.PL buy-ASP MW=_{bag} tangerine and MW=_{box} apple

‘We bought a bag of tangerines and a box of apples.’

In example (51), the Cl-N-*and*-Cl-N phrase ‘a book and a pencil’ (i.e. the underlined part of the sentence) is the direct object and follows the verb ‘bought’. In the post-verbal position, the two Cl-N phrases that are conjoined with *taq* ‘and’ are generally interpreted as indefinite NPs. One of the Ningbo Wu speakers considers the definite interpretation of these phrases to be less acceptable but still possible (in certain contexts), while the other

participants believe that the post-verbal Cl-N-*and*-Cl-N phrases can only be indefinite. In other words, the indefinite reading is at least the main interpretation of these phrases, and it is often the only reading. The level of acceptability of post-verbal Cl-N-*and*-Cl-N phrases is high in both dialects of Northern Wu. The grammaticality points for this construction are 2.75 in Ningbo Wu and 3.0 in Shanghai Wu.

In (52), the MW-N-*and*-MW-N phrase ‘a bag of tangerines and a box of apples’ is the object in the sentence and follows the verb. The two MW-N phrases connected by the conjunction *taq* ‘and’ are also semantically indefinite. In this case, the definite reading does not seem to be possible for any of participants. Such phrases are also highly acceptable in Northern Wu. The grammaticality points of the post-verbal MW-N-*and*-MW-N phrases are 2.5 in Ningbo Wu and 3.0 in Shanghai Wu.

To summarize the findings from this part of the data, there are three points that are worth noting. First, Num-Adj-Cl-N phrases (with size adjectives) are less acceptable than Num-Adj-MW-N phrases in both dialects of Northern Wu. Second, in the pre-verbal position, Cl-N-*and*-Cl-N phrases are semantically definite and highly acceptable, while MW-N-*and*-MW-N phrases are much less acceptable. Lastly, in the post-verbal position, Cl-N-*and*-Cl-N phrases and MW-N-*and*-MW-N phrases are about equally acceptable, and they are generally indefinite in meaning. The results of the surveys that are presented in this section and the previous one will be discussed in section 3.4.

3.4 Discussion of the results

Several patterns emerged from the Cantonese and Northern Wu data in the two previous sections. In this section, I make two claims based these data. First, I argue that classifiers and measure words belong to two separate categories. Second, I demonstrate that the sortal classifiers in Cantonese and Northern Wu differ from their Mandarin counterparts in terms of timing of merger and syntactic position. The interpretations of Cl/MW-N phrases are discussed in the last part of this section.

The most conspicuous characteristic of Cantonese and Northern Wu classifiers is that they are found in a variety of nominal constructions. In Cantonese, they are used in

pre-verbal Cl-N phrases, in possessor-Cl-N phrases, in Deg-Adj-Cl-N phrases and in Dem-Cl-N phrases. In these four types of NPs, the classifiers are consistently more acceptable than measure words. Only in indefinite NPs (i.e. Num-Cl/MW-N phrases and post-verbal Cl/MW-N phrases)⁷² are classifiers and measure words equally acceptable in Cantonese. Likewise, in Northern Wu, classifiers can occur in pre-verbal Cl-N phrases, possessor-Cl-N phrases, Adj-Cl-N phrases, and RelCP-Cl-N phrases, while the measure words cannot (or are less acceptable in these contexts). On the other hand, the measure words are more acceptable than classifiers in one type of NPs, namely in Num-Adj-Cl/MW-N phrases (with a size adjective), and this is true of Cantonese and Northern Wu. In other words, the acceptability of classifiers and measure words are frequently different from each other, and practically speaking they are used in different constructions. In view of these differences, classifiers and measure words should be two separate categories in Cantonese and Northern Wu.⁷³

These distributional differences between classifiers and measure words are also related to the question of how they are merged into the structure of NPs. I propose that the sortal classifier in a Cantonese or Northern Wu NP merges directly with the head noun, unlike their Mandarin counterparts which merge with numerals first. In other words, the classifiers in Cantonese and those in Northern Wu can be termed “classifiers-for-nouns”, while the classifiers in Mandarin can be termed “classifiers-for-numerals” (Little et al. 2022: 10). Classifiers-for-nouns are expected to occur in contexts without numerals, and they may not be obligatory depending on the noun that they co-occur with (ibid. 20). As described in sections 3.2 and 3.3, the classifiers in Cantonese and Northern Wu can be used in quite a few constructions without numerals. It is worth noting that Mandarin classifiers are also used in two types of NPs without numerals, namely Dem-Cl-N phrases and (post-

⁷² The post-verbal Cl/MW-N phrases can be definite or indefinite in Cantonese, and there are some contexts in which only the indefinite reading is available (Matthews & Yip 2011: 106).

⁷³ There are other (non-syntactic) differences between the classifiers and MWs in these languages. For instance, in Adj-Cl-N phrases in Northern Wu, the adjective modifies the noun, while the adjective in an Adj-MW-N phrase can also modify the MW. The classifiers are always monosyllabic and monomorphemic, while measure words can be longer and morphological complex (Qian 1997). Moreover, loanwords and neologisms are found in the MWs but not among the classifiers in Northern Wu and Cantonese (Qian 1997: 93-95; Mak & Taam 2011: 333-339).

verbal) bare Cl-N phrases. However, the numeral *yi* ‘one’ can be inserted in both types of NPs, and the interpretations of the NPs with *yi* ‘one’ and those without it (e.g. Dem-*yi*-Cl-N and Dem-Cl-N) are very similar if not identical. It is therefore reasonable to argue that the Dem-Cl-N phrases and bare Cl-N phrases in Mandarin contain an unpronounced pro-form of the numeral *yi* ‘one’, namely *YI_{pro}* (Zhang 2019). With this special pro-form of *yi* ‘one’, one can argue (on the basis of other evidence) that classifiers always merge with numerals in Mandarin NPs. The situation is rather different in Cantonese and Northern Wu. In most of the constructions that contain a classifier but no numerals in these two languages, the numeral ‘one’ cannot be inserted (Li 2013: 279-280). For example, the Dem-*iq*-Cl-N phrases in Northern Wu are less acceptable than Dem-Cl-N phrases, and the same is true of the Dem-*jat*-Cl-N phrases in Cantonese. Although the numeral ‘one’ may sometimes be inserted into bare Cl-N phrases, it would change their meaning radically (from definite to indefinite) in Cantonese and Northern Wu (Cheng & Sybesma 1999: 528). I propose that there is no pro-form of ‘one’ in Cantonese or Northern Wu that is comparable to *YI_{pro}* in Mandarin, and that the classifiers in Northern Wu and Cantonese are indeed used in several constructions without any numerals. According to the criteria proposed by Little et al. (2022), this would constitute evidence that Northern Wu and Cantonese classifiers are classifiers-for-nouns, which merge with the head noun first to the exclusion of numerals.

Another type of evidence for Northern Wu classifiers being classifiers-for-nouns is that some nouns in Ningbo Wu do not require a classifier when used with a demonstrative. For instance, the classifier *ghoq* is optional in *kiq-ghoq-nin* ‘that person’, while the classifier *tsy* (or *tsyu* for some speakers of Ningbo Wu) in *kiq-tsy-piq* ‘that pen’ is obligatory.⁷⁴ As the optionality of the classifier depends on the noun, it should be considered evidence that these two words merge and form a constituent (Little et al. 2022).

Her (2017: 38-39) proposes a constituency test using coordination and argues that the Num-Cl sequence is a constituent in Mandarin. According to my data, the Cl-N sequence passes this test in both Cantonese and Northern Wu, because the Cl-N-and-Cl-N phrases are grammatical in pre-verbal and post-verbal positions. In other words, the Cl-N part of an NP

⁷⁴ In Shanghai Wu, classifiers are always obligatory in Dem-Cl-N phrases (Qian 1997: 110), hence the lack of evidence from that dialect. Cantonese classifiers are obligatory in this context (Matthews & Yip 2011: 107).

is a constituent in Cantonese and Northern Wu according to this criterion. Num-Cl-*and*-Num-Cl phrases are somewhat less acceptable than Cl-N-*and*-Cl-N phrases in these two languages, which also lends some support to the classifiers-for-nouns account.

There are also some arguments against the classifiers-for-numerals analysis of Cantonese and Northern Wu classifiers. First, classifiers are rarely used when counting in Cantonese and Northern Wu and are never used when talking about the number. In other words, whether the classifier occurs or not depends on the noun and not on the numeral. Second, the numerals *yī* ‘one’ and *èr* ‘two’ in Mandarin have special forms (*yí* or *yì* for ‘one’ and *liang* for ‘two’) that are mainly used with classifiers, and I consider it evidence for the classifiers-for-numerals analysis of Mandarin classifiers (see section 2.3). In Northern Wu, the pattern is completely different. The numeral ‘one’ is always *iq*, and the variant *lian* of the numeral ‘two’ (corresponding to Mandarin *liang*) is used in several positions other than Num-Cl/MW-N phrases, including in fractions and percentages and when counting⁷⁵. In short, Cantonese and Northern Wu classifiers are not classifiers-for-numerals.

There are also some notable semantic patterns in the data. First, the pre-verbal Cl-N phrases and MW-N phrases (when they are not completely unacceptable) in Northern Wu are always definite in meaning, while the same phrases are consistently indefinite in the post-verbal position. Definiteness is known to correlate with information structure. The most identifiable information (i.e. accessible and shared information) is most likely to be the topic, while the unidentifiable information tends to be the informational focus rather than the topic (Li & Bisang 2012: 340-341). In Chinese, the topic is pre-verbal, and the informational focus is on the sentence-final (hence post-verbal) position (Li 2013: 246). Since definiteness correlates with identifiability, it is unsurprising that definite NPs tend to be pre-verbal and that indefinite NPs tend to be post-verbal, which is true for many SVO languages (Li & Bisang 2012: 340). This semantic distinction between pre-verbal Cl/MW-N phrases and post-verbal Cl/MW-N phrases is particularly clear in Northern Wu because it is a “prototypical topic-prominent language” (Li 2013: 246). When the object of a sentence

⁷⁵ In terms of the use of these variants of ‘two’, Shanghai Wu and Ningbo Wu are almost identical. There is another variant of ‘two’, namely *gher*, which is borrowed from Mandarin and is rarely used (Qian 1997: 81).

is definite, object fronting with or without an overt marker⁷⁶ is common in Northern Wu, thus creating OSV and SOV word orders and eliminating post-verbal definite Cl/MW-N phrases in this language (Li 2013: 246-247). The paucal forms of these bare Cl/MW-N phrases, namely *lian*-Cl/MW-N, are also definite when used pre-verbally and indefinite when post-verbal. The modified Cl-N phrases (and the modified MW-N phrases, which are less acceptable) are generally definite in meaning and can occur post-verbally.

In Cantonese, the pre-verbal Cl/MW-N phrases are also obligatorily definite, but the post-verbal ones can be definite or indefinite (Matthews & Yip 2011; Li & Bisang 2012). It is worth noting that Cantonese is a relatively typical SVO language in which topic-fronting is less common, and that the disposal construction (with the marker *jeung*)⁷⁷ is rarely used in spoken Cantonese (Li 2013: 246; Li & Bisang 2012: 339-340). Even though post-verbal Cl/MW-N phrases can be definite or indefinite in Cantonese, the last part of my data seems to show that they are still preferably indefinite. In Cl-N-*and*-Cl-N phrases and MW-N-*and*-MW-N phrases, the participants can all interpret the Cl-N or MW-N phrases as indefinite, but only some of them accepted the definite interpretation. Curiously, in these complex NPs with the coordinator conjunction *tung* ‘and’, it is more acceptable to interpret the first Cl/MW-N phrase as an indefinite NP and the second one as a definite NP than the reverse (see examples (32) and (33) at the end of section 3.2).⁷⁸ This phenomenon seems to support the theory that the initial conjunct is structurally higher than the final conjunct (i.e. specifier vs complement) (Zhang 2009: 192). A fuller theoretical explanation will need to be sought once this pattern is confirmed by a larger set of data.

Another intriguing finding is related to the use of classifiers and measure words with adjectives. In Ningbo Wu, the Adj-Cl-N phrases, which are definite in meaning are preferable to the (definite) Adj-MW-N phrases,⁷⁹ whilst the Num-Adj-MW-N phrases, which

⁷⁶ In Ningbo Wu, the two main markers are *taq* in the disposal construction and *paq* in the passive construction.

⁷⁷ This morpheme is transliterated as such in Li & Bisang (2012). It is 將 in Chinese and *zoeng* in Jyutping.

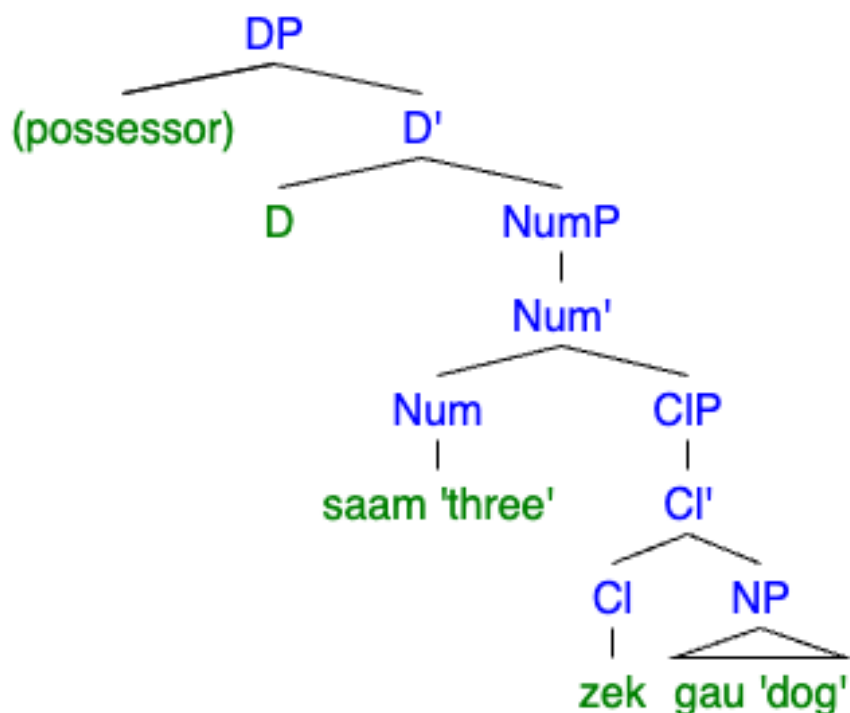
⁷⁸ For example, in (32), three participants accepted the meaning ‘I bought a pen and the magazine’, while only one participant accepted the meaning ‘I bought the pen and a magazine’.

⁷⁹ Recall that no such difference is found in my Shanghai Wu data (the size of which is very small).

are indefinite, are more acceptable than the (indefinite) Num-Adj-Cl-N phrases. Likewise, in Cantonese, the definite Deg-Adj-Cl-N phrases are more acceptable than Deg-Adj-MW-N phrases, while the indefinite Num-Adj-MW-N phrases are more acceptable than Num-Adj-Cl-N phrases. It appears that Northern Wu and Cantonese classifiers are associated with definiteness to some extent, at least occurring in these constructions with adjectives. Note that there is nothing comparable to (Deg)-Adj-Cl-N phrases in Mandarin, and that Num-Adj-Cl-N phrases and Num-Adj-MW-N phrases are both acceptable in Mandarin (Li 2013).

Based on my interpretations and analyses of the Cantonese and Northern Wu data, I propose the following tree structures for these two languages.

Tree 2: Structure of Num-CL-N phrases in Cantonese

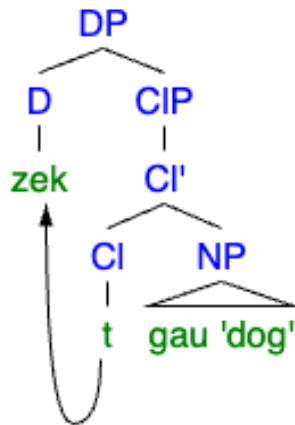


(Based on Cheng & Sybesma 1999: 527 (35))

Tree 2 represents the structure of the Cantonese NP *saam-zek-gau* 'three dogs'. In Cantonese, the classifier merges with the head noun to form a Classifier Phrase, which then merges with the numeral. I propose that the same structure is found in Num-CL-N phrases

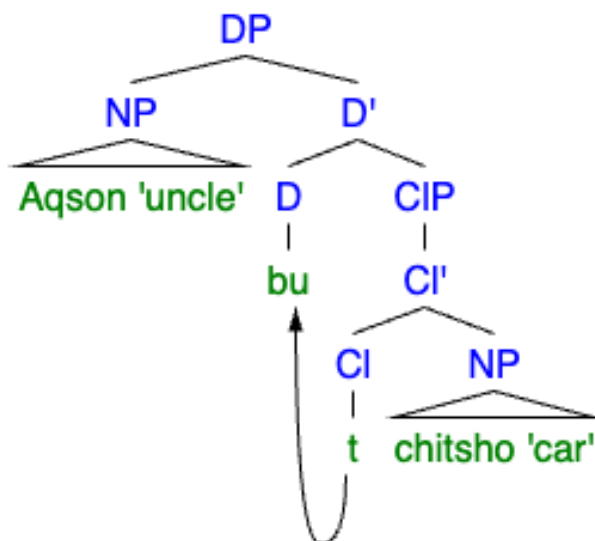
in Northern Wu, which will be omitted here. The next tree shows the structure of preverbal bare CL-N phrases in Cantonese, which are semantically definite.

Tree 3: Structure of pre-verbal bare CL-N phrases in Cantonese



Tree 3 is for the Cantonese NP *zek-gau* 'the dog'. In this tree, the classifier merges with the head noun in the bare CL-N, and it then moves to D⁰, where it gets its definite meaning. Pre-verbal bare CL-N phrases in Northern Wu have the same structure as this one. Tree 4 shows the structure of a modified CL-N phrase in Ningbo Wu.

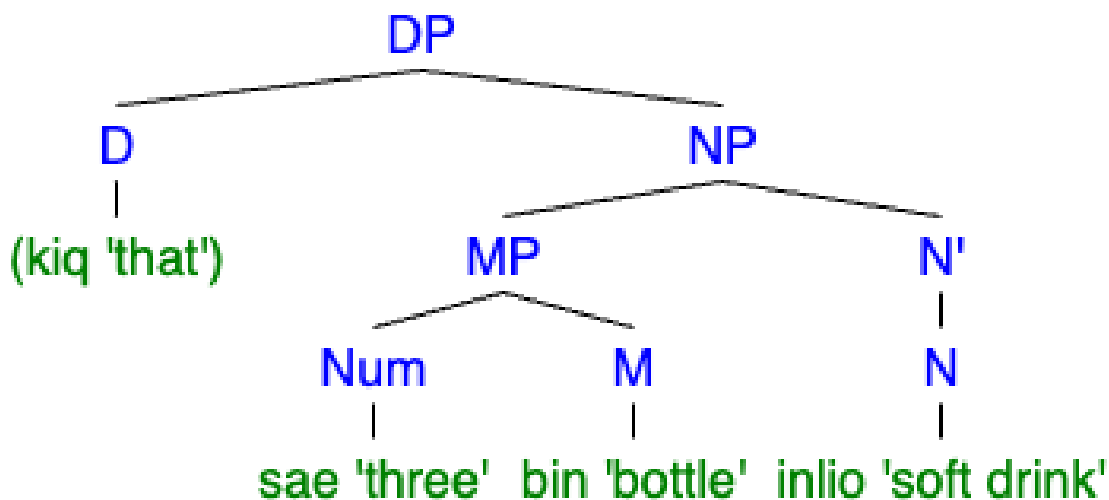
Tree 4: Structure of possessor-CL-N phrases in Northern Wu



(Based on Li 2013: 259 (41))

Tree 4 represents the structure of the Ningbo Wu NP *aqson-bu-chitsho* ‘uncle’s car’. The classifier *bu* moves from Cl^0 to D^0 , where its meaning becomes definite. The modified CL-N phrases (e.g. Possessor/Adj-CL-N phrases) in Cantonese and Northern Wu all have structures comparable to this one, with Cl^0 -to- D^0 movement. Tree 5 indicates the structure of Num-MW-N phrases in Northern Wu.

Tree 5: Structure of Num-MW-N phrases in Northern Wu



(Based on Bale et al. 2019: 2 (1))

Tree 5 shows the structure of the NP *sae-bin-inlio* ‘three bottles of soft drink’ in Ningbo Wu. Unlike the classifiers, the measure word *bin* ‘bottle’ first merges with the numeral *sae* ‘three’ rather than the head noun. Note that the Num-MW-N phrases in Northern Wu, Cantonese and Mandarin have virtually the same structure.

In the next section, I summarize my arguments on the category and syntax of the classifiers and measure words in Mandarin, Cantonese and Northern Wu.

3.5 Interim conclusion

In the last two chapters, I addressed my research questions and proposed two major claims for each of the three Sinitic languages. In Chapter 2, I argued that classifiers and measure words belong to the same grammatical category (which could be termed “unit words”) in Mandarin, and that they always merge with the numeral, which can be overt or null, before the [Num-Cl/MW] constituent merges with the head noun or other morphemes. For Cantonese and Northern Wu, I argued that classifiers and measure words are two separate categories in these languages. I proposed the [Num-[Cl-N]] structure for Cantonese and Northern Wu quantified NPs with classifiers, and I assumed that the quantified NPs with measure words have the [[Num-MW]-N] structure in these two languages.

In Chapter 2, I showed that Mandarin classifiers and measure words share the same distribution and the same position in five types of NPs, and that they exhibit the same behaviour in sentences with N-ellipsis, N-extraction or coordination. These facts indicate that classifiers and measure words belong to one and the same grammatical category in Mandarin. Although there are some differences between Mandarin classifiers and measure words, those differences are phonological, morphological or semantic rather than syntactic. Furthermore, classifiers do not differ consistently from measure words even in those regards: Mandarin classifiers and measure words behave alike in Tone-3 sandhi and have the same disambiguating effect in bare Cl/MW-N phrases. The last major argument of that chapter was about how classifiers and measure words merge into Mandarin NPs. I argued that they first merge with the numeral to form a [Num-UW] constituent.⁸⁰ This claim was supported by a few types of evidence. First, three syntactic tests (based on extraction, ellipsis and coordination) indicated that [Num-UW] is a constituent. Second, there were reasons to believe that Dem-UW-N phrases and post-verbal bare UW-N phrases contain a special unpronounced form of the numeral *yi* ‘one’, which makes the left-branching account compatible with these types of NPs in Mandarin. Third, some [Num-UW] sequences (esp.

⁸⁰ Since classifiers and measure words belong to the same category, I use the abbreviation “UW” (unit words) when referring to this category.

with measure words, such as *san-da* ‘three dozen’) are semantically very similar to complex numerals, and they should be structurally similar as well. Lastly, two lower numerals (i.e. *yī* ‘one’ and *èr* ‘two’) in Mandarin have special variants (i.e. *yí/yì* ‘one’ and *liǎng* ‘two’) that are mostly employed with unit words. In short, Mandarin classifiers and measure words belong to the same category (i.e. unit words), and they merge with numerals first.

In Chapter 3, I described the use of classifiers and measure words in Cantonese and Northern Wu. I argued that classifiers and measure words are two separate categories in these two languages. This was supported by a few types of evidence. First, I showed that the distribution of Cantonese and Northern Wu classifiers is different from that of the measure words (as well as Mandarin classifiers). In both Cantonese and Northern Wu, the sortal classifiers are used in pre-verbal bare Cl-N phrases and in a few types of modified Cl-N phrases (e.g. possessor-Cl-N phrases), in which measure words are consistently less acceptable.⁸¹ On the other hand, Num-Adj-Cl-N phrases are less acceptable than Num-Adj-MW-N phrases in both languages. Second, the classifiers often mark definiteness in Cantonese and Northern Wu, while the measure words almost always occur in indefinite NPs. Lastly, there are some phonological, morphological and lexical differences between classifiers and measure words in Cantonese and Northern Wu. My second argument was that the classifier merge with the head noun first to form a [Cl-N] constituent in Cantonese and Northern Wu NPs. This works better than the [Num-Cl] analysis since the classifiers occur in a variety of constructions, esp. when no null form of the numeral ‘one’ can be postulated in Cantonese or Northern Wu. This claim was also supported by the fact that the [Cl-N] sequences pass the coordination test in Cantonese and Northern Wu. Cantonese and Northern Wu measure words do not seem to be very different from their Mandarin counterparts in syntactic behaviour, and I assumed that the [Num-MW] analysis of Mandarin measure words applies to them as well.

In the next chapter, I describe the diachronic changes that occurred in the classifier system of Chinese in the past three millennia, and I argue that its evolution is best analyzed as an ongoing grammaticalization process, which explains many of the differences among the three modern Chinese languages.

⁸¹ This difference in acceptability appears to be more obvious in Northern Wu than in Cantonese.

4. The evolution of Chinese classifiers: an account based on grammaticalization

In the previous chapters, we have seen that sortal classifiers are frequently used and often obligatory in three modern Chinese languages. However, historically, Chinese has not always been a classifier language. Old Chinese has prototypes of sortal classifiers, but they are uncommon and are comparable to English unit counters. A fully fledged classifier system emerged in Early Middle Chinese, with some further changes in Late Middle Chinese (C. Zhang 2012; Yang-Drocourt 2004). This classifier system was later inherited by Mandarin, Cantonese and Northern Wu, with varying degrees of alteration.

Although the development and changes in the classifier system were gradual, here I divide the history of Chinese into four eras to make the description clear and concise. The earliest written records of Chinese date back to the 13th century BCE, and there is no doubt that the Chinese texts written by the 3rd century BCE are in Old Chinese (Jiang 2009; Yang-Drocourt 2004).⁸² The second era is that of Early Middle Chinese, which I assume to last from the 3rd century CE to the 9th century CE, and I focus mainly on the first few centuries (i.e. the 3rd to the 6th century) of this era. This is followed by the era of Late Middle Chinese from the 10th century to the 13th century CE. Lastly, I describe the changes in the classifier system from Late Middle Chinese to modern Mandarin, Cantonese and Northern Wu. The history of Chinese classifiers will be summarized in Section 4.1, which consists of four subsections, each describing the changes in one era.

After a description of the diachronic changes, I demonstrate that these four eras correspond to four stages in the grammaticalization of Chinese classifiers. I argue that the classifiers in Cantonese and Northern Wu are grammaticalized to a higher degree than their Mandarin counterparts, and that this difference accounts for many of the syntactic and other differences among the three modern languages. This account and the relevant arguments will be presented in Section 4.2.

In Section 4.3, I compare the history of Chinese classifiers and that of measure words in that language. I determine how and why the classifiers diverged from measure

⁸² According to J.-P. Li (2017), the Old Chinese era lasts until the end of Han Dynasty (i.e. early 3rd century).

words in Cantonese and Northern Wu. Some lexical and morphological changes will also be described in this section, and I argue that classifiers and measure words evolve for different reasons.

Section 4.4 is the conclusion of the chapter, where I summarize my findings related to the history of Chinese classifiers and measure words.

4.1 A brief history of Chinese classifiers

As mentioned previously, the history of Chinese classifiers described in this section is organized into four eras, namely Old Chinese, Early Middle Chinese, Late Middle Chinese, and the modern Chinese languages. The Old Chinese era starts from the 13th century BCE, when the first extant Chinese inscriptions were produced during the Shang Dynasty (16th - 11th century BCE), and it ends with the Qin Dynasty (221-206 BCE). The Old Chinese texts were initially written in a few different scripts on various objects (e.g. oracle bones, bronze, bamboo slips and silk fabric), some of which were only discovered in the 20th century (J.-P. Li 2017: 36-37).¹ In this chapter, all the examples in Old Chinese or Middle Chinese will be shown in Traditional Chinese characters and Pinyin Romanization.²

4.1.1 Old Chinese

In Old Chinese, classifiers are optional and infrequent, and they must co-occur with numerals (Yang-Drocourt 2004). It is unclear whether they are different enough from their lexical sources to be considered a separate grammatical category (Jiang 2009: 141). I shall therefore refer to these words in Old Chinese as “proto-classifiers”. An important subtype of Old Chinese proto-classifiers consists of repeater classifiers (AKA echo classifiers), which

¹ Some of the ancient Chinese scripts, esp. the Oracle-bone script and the Bronze script, are very different from Modern Chinese (both the traditional and the simplified version).

² Note that the transliteration of the examples is based on their pronunciations in Mandarin. Reconstructed Old Chinese or Middle Chinese pronunciations will only be mentioned when necessary.

are identical to the nouns they classify. There were other classifiers emerging during the Old Chinese (OC) era, although the quantified NPs in Old Chinese are most often Num-N or N-Num phrases (Yang-Drocourt 2004: 40, 83; J.-P. Li 2017). Old Chinese quantified NPs without classifiers are shown in the examples below.

(1) 二 子 乘 舟... [Old Chinese]

er zi cheng zhou

two person ride boat

‘Two people are sailing in a boat...’

(ASTC-OC: *The Book of Odes – Odes of Bei* [8th-6th century BCE])

(2) 貞 又 牢 三 羌 十。

zhen you lao san qiang shi

submit.to.the.oracle offer sacrificial.ox three Qiang ten

‘We submit to the oracle: we shall offer three oxen and ten Qiang people in sacrifice.’

(HJ-341 [13th-11th century BCE], as cited in Yang-Drocourt 2004: 42)³

In (1), the subject NP *er-zi* ‘two people’ is a Num-N phrase without any classifiers. In many Old Chinese texts (e.g. oracle-bone inscriptions, *The Book of Odes* and *Mencius*), this is by far the most common type of quantified NPs (Yang-Drocourt 2004: 40, 83). The reverse order of the numeral and the head noun is seen in (2), where the object NPs are N-Num phrases. This type of quantified NPs, again without any classifiers, is reasonably common in OC. They account for 25.8% of the quantified NPs in oracle-bone inscriptions and 29% of the occurrences in bronze inscriptions (Yang-Drocourt 2004: 42, 83). In total, Num-N and N-Num phrases often make up about 70% to 90% of the occurrences of quantified NPs in Old Chinese texts (Yang-Drocourt 2004: 40, 83). The remaining types of NPs are more

³ Abbreviation: HJ = *Jiaguwen Heji* (Extensive collection of the inscriptions on bones or tortoise shells). For further bibliographical information, see Yang-Drocourt 2004: 210.

complex, including N-Num-MW/Cl phrases and Num-MW/Cl-N phrases.⁴ Since measure words are not particularly rare in Old Chinese (J.-P. Li 2017: 179), the proto-classifiers only account for a small fraction of the occurrences of the quantified NPs in OC. The use of the proto-classifiers is illustrated in the examples below.

- (3) 孚 牛 三百五十五 牛，羊 廿八 羊。 [Old Chinese]
 fu niu san-bai-wu-shi-wu niu yang nian-ba yang
 capture ox three-hundred-five-ten-five ox sheep twenty-eight sheep
 ‘They captured three hundred and fifty-five oxen and twenty-eight sheep.’
 (Xiao-Yu Ding [10th century BCE], as cited in J.-P. Li 2017: 173)

- (4) 有 人 於此，力 不 能 勝 一 匹 雛。
 you ren yu ci, li bu neng sheng yi pi chu
 there.be person at here strength NEG can bear one CL chick
 ‘Here is a person whose strength does not allow him to carry a chick.’
 (ASTC-OC: *Mencius* [late 4th century – early 3rd century BCE])

In (3), the head noun *niu* ‘ox’ is repeated later in the NP, and the second copy of it is a proto-classifier of the repeater type. The same happens with the noun *yang* ‘sheep’, which is followed by a numeral and the repeater classifier *yang*. There are at least eight repeater classifiers in Old Chinese, all of which have become obsolete by the 3rd century BCE (J. P. Li 2017: 172-176). In (4), the object NP ‘a chick’ is a Num-Cl-N phrase containing the classifier *pi*. It is one of the first instances of Num-Cl-N phrases in Chinese, although such NPs are less frequent than Num-N phrases and N-Num-Cl phrases (Yang-Drocourt 2004: 83). Lexically, the non-repeater proto-classifiers in Old Chinese are also very different from the classifiers in Mandarin: some of the OC classifiers have disappeared, and the ones that survived often became compatible with different sets of nouns (J. P. Li 2017).

⁴ Num-MW/Cl-N phrases are virtually absent in the early OC inscriptions and remained infrequent in late Old Chinese (Yang-Drocourt 2004).

Most of the proto-classifiers are used with a small number of nouns in Old Chinese, and the generic classifiers only began to emerge during the last two centuries of the Old Chinese era (J. P. Li 2017; Yang-Drocourt 2004). The use of the generic classifier *gè* in Late Old Chinese is shown in example (5).

- (5) 負 服 矢 五十 箇。 [Old Chinese]
 fu fu shi wushi ge
 carry.on.the.back quiver arrow fifty CL
 ‘(A soldier from Wei) carries fifty arrows in his quiver.’
 (Xunzi [3rd century BCE], as cited in Yang-Drocourt 2004: 88)

In (5), the sentence-final NP ‘fifty arrows’ contains the classifier *gè*.⁵ This classifier is a grammaticalized variant of the adjective 介 *jiè* ‘single, sole’, and it was already a generic classifier when it became a classifier (J.-P. Li 2017: 58-59). The grammaticalization process was ongoing during the 4th century BCE, when the word *gè* was sometimes a classifier and sometimes an adjective meaning ‘single, alone’.⁶ This classifier is compatible with a variety of nouns, including those denoting humans, animals, bamboos, arrows or deerskin (Yang-Drocourt 2004: 88; C. Zhang 2012: 67). This is the etymon of the generic classifier *gè* in Mandarin. Another major generic classifier is 枚 *méi*, which first occurred in the late 3rd century BCE (at the very end of the Old Chinese era) when it was grammaticalized from the noun *méi* ‘tree trunk’ and became a classifier (J.-P. Li 2017: 50, 56). The classifier 隻 *zhī* (i.e. the etymon of the generic classifier *tsaq* in Northern Wu) appeared somewhat later during the 2nd or the 1st century BCE (J.-P. Li 2017: 172).

In addition to the individual classifiers, Old Chinese also has a few types of measure words, including group measure words (e.g. *shù* ‘bundle’), container MWs (e.g. *hú* ‘flagon, pot’ and standard measures (e.g. *jīn* ‘catty’) (J.-P. Li 2017: 188-189, 245-246, 303). Both Num-MW-N phrases and N-Num-MW phrases are found in Old Chinese, although the latter

⁵ The characters 箇 and 个 are homophonous and interchangeable (Yang-Drocourt 2004: 89).

⁶ Note that both meanings of this word are found in *Zuozhuan*, which was written in the 4th century BCE (Yang-Drocourt 2004: 86-87; J.-P. Li 2017: 26).

is more common, especially in the early OC texts (Yang-Drocourt 2004: 83; J.-P. Li 2017). Two Old Chinese examples with measure words are given below.

(6) 或 取 一 秉 杆 焉。 [Old Chinese]

huo qu yi bing gan yan
someone take one MW_{=handful} hay PARTICLE

‘Someone took a handful of hay.’

[*Zuozhuan* [4th century BCE], as cited in Yang-Drocourt 2004: 80)

(7) 陳 饋 八 簋。

chen kui ba gui
arrange food eight MW_{=round.food.vessel}

‘I arranged eight pots of food.’

[*The Book of Odes – Xiaoya* [9th-6th century BCE], as cited in J.-P. Li 2017: 260)

In (6), the object NP ‘a handful of hay’ is a Num-MW-N phrase with the measure word *bing* ‘handful’. In (7), the object NP ‘eight pots of food’ is a N-Num-MW phrase with the container measure word *gui* ‘bronze pot (for grain)’. N-Num-MW phrases, such as the one in (7), are found in the earliest texts in Old Chinese, whilst the Num-MW-N phrase emerged several centuries later (Yang-Drocourt 2004: 47-48, 74). Lexically, the measure words in Old Chinese and those in Mandarin are vastly different because the containers and standard measures have changed greatly and many archaic MWs (including group MWs) were replaced by new ones throughout the history of Chinese (J.-P. Li 2017).

To sum up, classifiers in Old Chinese have a few characteristics that make them different from their modern counterparts. First, the individual classifiers are optional and uncommon in Old Chinese.⁷ Second, they are more often found in N-Num-Cl phrases than in

⁷ The proto-classifiers in OC are comparable to the unit counters in English (e.g. *head* in *a head of cabbage*) in this regard (Li 2013: 34-35). Note that the label “NU” (abbreviation for *nom d’unité* ‘unit nouns’) is used for Old Chinese proto-classifiers in Yang-Drocourt (2004), which is different from the CL label for the classifiers

Num-Cl-N phrases (Yang-Drocourt 2004). Third, none of the classifiers was frequent since generic classifiers did not start to occur until the last two centuries of the Old Chinese era. Lastly, several repeater classifiers are used in Early Old Chinese (J.-P. Li 2017). In short, the prototypes of classifiers were evolving in embryonic form during the Old Chinese era.

4.1.2 Early Middle Chinese

While Old Chinese is not a classifier language, there is good reason to believe that a fully-fledged classifier system emerged by the Early Middle Chinese era (i.e. 3rd century to 9th century CE) (Jiang 2009: 153; Yang-Drocourt 2004: 119). Some written records of the vernacular of Early Middle Chinese have been excavated from ancient garrisons and burial sites, and those texts indicate that classifiers had become virtually obligatory in Middle Chinese quantified NPs by the 3rd or the 4th century CE (Yang-Drocourt 2004: 115-119).⁸ The birth of this classifier system comes with a few other changes. First, many new classifiers appeared, including some shape classifiers (Jiang 2009: 152). Second, some of the existing classifiers (such as the shape classifier *tiáo* and the generic classifier *méi*) become compatible with other nouns (Yang-Drocourt 2004: 127-132). Third, abstract nouns start to be quantified with classifiers (ibid. 134). Fourth, the Num-Cl-N word order becomes more common than before and may have become the unmarked order during this era, although the N-Num-Cl order remains frequent (Yang-Drocourt 2004: 135-140; Jiang 2009: 153-154). The use of Early Middle Chinese classifiers is illustrated below.

(8) 支道林 常 養 數 匹 馬。 [Early Middle Chinese]
Zhidaolin chang yang shu pi ma

in Middle Chinese and Mandarin. In other words, OC classifiers were still in the same grammatical category as their lexical sources.

⁸ In the literary texts from that era, quantified NPs with classifiers are still less frequent than those without classifiers (Yang-Drocourt 2004: 121). This conservative feature is attributable to the influence of Old Chinese.

Zhidaolin usually keep several CL horse
'Zhi Daolin usually keeps several horses.'
(ASTC-MC: *Shishuo xinyu* [5th century])

(9) 復 有 石碑 四十八 枚。
fu you shibei sishi-ba mei
again there.be stela forty-eight CL
'There were again forty-eight stelae.'
(*Luoyang qielan ji* [6th century]; as cited in Yang-Drocourt 2004: 130)

In (8), the object NP 'several horses' is a Num-Cl-N phrase with the specific classifier *pi*,⁹ which is mostly used with just one noun, namely *ma* 'horse' (Zheng 2016: 209; Yang-Drocourt 2004: 165; ASTC-MC). In (9), the NP 'forty-eight stelae' is a N-Num-Cl phrase with the generic classifier *mei*. This classifier can be used with over a hundred nouns, which are semantically diverse (e.g. *yu* 'fish', *li* 'pear', *jian* 'sword', *qian* 'coin') (Zheng 2016: 207-208). Sometimes both the general classifier *mei* and a specific classifier can be used with a noun (e.g. *yu* 'fish' and *zhu* 'bead') in Early Middle Chinese (C. Zhang 2012: 203-4). The N-Num-Cl and Num-Cl-N constituent orders coexisted in Chinese texts throughout the Middle Chinese era and were usually interchangeable (Yang-Drocourt 2004: 154-157). One noteworthy exception is that only N-Num-Cl phrases can be used when enumerating the items on a list (Yang-Drocourt 2004: 151), which is one of the few contexts in which N-Num-Cl phrases can occur in Mandarin and Northern Wu.

The measure words also underwent some changes from Old Chinese to Early Middle Chinese. First, while Num-MW-N phrases were uncommon in Old Chinese, they had become somewhat more frequent than N-Num-MW phrases by the 5th century CE (Yang-Drocourt 2004: 135). Second, there had been some lexically changes since the Old Chinese era. For instance, the group measure words *dui* 'row, group' was first used in the 6th century (Zheng 2016: 221-222). Some Old Chinese measure words had become obsolete by the 3rd century CE (i.e. the beginning of the Early Middle Chinese era) (J.-P. Li 2017: 200-201, 265). Lastly,

⁹ In Middle Chinese, this morpheme is also used as a measure word for textiles (Zheng 2016: 123-125).

indefinite MWs (e.g. *dian* ‘a little’) and temporary MWs (e.g. *shen* ‘body’) appeared during the Early Middle Chinese era (Wang 2018: 331-334, 647-655).

In short, classifiers in Early Middle Chinese have a few notable features. They are obligatory at least in the colloquial version of Early Middle Chinese. In terms of word order, they can occur in both N-Num-Cl phrases and Num-Cl-N phrases, and the latter type is becoming increasingly common. Lexically, the set of classifiers in Early Middle Chinese is considerably larger than that of Old Chinese, as shape classifiers and generic classifiers¹⁰ become frequent (Jiang 2009: 151-153). More than thirty Early Middle Chinese classifiers survived into Mandarin, often with some changes in their usage, although many classifiers became obsolete (Zheng 2016; C. Zhang 2012: 230-231).

4.1.3 Late Middle Chinese

The Late Middle Chinese era from the 10th century to the 13th century witnesses a few additional changes in the classifier system, after which it is already comparable in many ways to the classifier system of Mandarin. First, quantified NPs are now predominantly Num-Cl-N in terms of constituent order, and the N-Num-Cl word order becomes “marginal” (Yang-Drocourt 2004: 171-173).¹¹ Second, bare Cl-N phrases and bare MW-N phrases with no overt numerals began to be used in the post-verbal position in the 10th century (or slightly earlier) (Yang-Drocourt 2004: 175-177). Third, classifiers (esp. *gè*) start to occur in Dem-Cl-N phrases (ibid. 178-181). Lastly, *gè* is the only generic classifier in Late Middle Chinese as the use of *méi* declined rapidly towards the end of the Early Middle Chinese era (Wang 2018: 508, 519; C. Zhang 2012: 252-253). Example (10) shows how Late Middle Chinese classifiers are used in quantified NPs.

¹⁰ The generic classifier *méi* has been discussed above. The other major generic classifier is *gè*, which became increasingly common from the 7th century onwards and was the primary generic classifier during the last part of the Early Middle Chinese era (Wang 2018). The “quasi-generic” classifier *zhī* is also quite versatile and is used with several types of nouns (e.g. animals, weapons, body parts and jewellery) (C. Zhang 2012: 268).

¹¹ Exceptionally, the N-Num-Cl word order sometimes occurs when the noun denotes humans or books, especially if an appositive reading is available (Yang-Drocourt 2004: 172).

- (10) 師 入 園 中， 見 一 株 菜。 [Late Middle Chinese]
 shi ru yuan zhong jian yi zhu cai
 teacher enter garden inside see one CL vegetable
 ‘The (Zen) Master went into the garden and saw a vegetable.’
 (*Zutangji* [10th century], as cited in Wang 2018: 68)

In (10), the object NP ‘a vegetable’ is a Num-Cl-N phrase with the classifier *zhū* for plants. In addition to a relatively large set of shape classifiers, Late Middle Chinese has several classifiers whose usage is based on animacy (e.g. *tóu* for animals, *yuán* and *míng* for humans), and *zhū* is one of them (Wang 2018: 67-68, 124, 139; C. Zhang 2012: 262). In the texts that record colloquial language of this era, over 70% of the quantified NPs contain a numeral, a classifier and the head noun in that order (Yang-Drocourt 2004: 173). In other words, Num-Cl-N is the main and unmarked form of quantified NPs in Late Middle Chinese.

Another notable change in the use of classifiers is that they can now occur in two types of NPs without numerals. First, they are used in post-verbal bare Cl-N phrases which are indefinite in meaning (Yang-Drocourt 2004: 176). An example in Late Middle Chinese with a bare Cl-N phrase is given below.

- (11) 門 前 有 個 老 人。
 men qian you ge lao ren
 door in.front.of have CL old person
 ‘In front of the door, there is an elderly person.’
 (*Shunzhibian* [10th century], as cited in Yang-Drocourt 2004: 176)

In (11), the post-verbal NP ‘an old person’ is a bare Cl-N phrase and does not contain any numerals. In this case, the generic classifier *gè* is “a veritable grammatical morpheme void of meaning” which is approaching indefinite articles in function (Yang-Drocourt 2004: 176). It is worth noting that Old Chinese has another type of bare Cl-N phrases which are fundamentally different from the post-verbal bare Cl-N phrases in Late Middle Chinese. The (emphatic) bare Cl-N phrases in Old Chinese usually co-occur with negation and can be

translated as “not a single N”, and they can be used as subjects in the pre-verbal position (Wang 2018: 778-781). On the other hand, the post-verbal bare CI-N phrases in Late Middle Chinese are not emphatic and not related to negation in any way. In Mandarin, post-verbal bare CI-N phrases remain productive and are used in essentially the same way as in Late Middle Chinese, while the emphatic CI-N phrases that are found in Old Chinese only survive in certain idiomatic expressions.¹² In Late Middle Chinese, other classifiers and measure words are also used in the post-verbal bare CI/MW-N phrases, although *gè* is by far the most frequent one in this construction (Yang-Drocourt 2004: 176; C. Zhang 2012: 222), presumably because it tends to be semantically vacuous.

The second type of NPs without numerals in which Late Middle Chinese classifiers can occur is Dem-Cl-N phrases, and the classifiers began to occur in that construction in the 10th century (Yang-Drocourt 2004: 178).¹³ This use of classifiers in Late Middle Chinese is illustrated in example (12).

- (12) 此 個 真珠 若 採 得...
 ci ge zhenzhu ruo cai de
 this CL pearl if collect obtain
 ‘If I can dive and get this pearl...’
 (Zutangji [10th century], as cited in Wang 2018: 525)

In (12), the NP ‘this pearl’ contains the generic classifier *gè*, which occurs between the demonstrative *ci* ‘this’ and the head noun *zhenzhu* ‘pearl’. This construction is already more common than the corresponding “classical” construction without classifiers (Dem-N) in Late Middle Chinese (Yang-Drocourt 2004: 178). Although several other individual classifiers also occur in this construction, most of the occurrences of such NPs contain the generic classifier *gè* (Yang-Drocourt 2004: 178-180; C. Zhang 2012: 222). Among the Dem-

¹² For example, the idiom 寸草不生 (*cun-cao-bu-sheng* in Pinyin) contains the MW-N phrase *cun-cao* (literally ‘inch-grass’) and the negation marker *bu*, and it means ‘(so barren that) not a single leaf of grass can grow’.

¹³ According to Wang (2018: 760), Dem-Cl-N phrases first occurred in the 5th or 6th century (during the Early MC era), and it became much more common in Late Middle Chinese.

ge-N phrases, *zhe-ge*-N phrases (with the demonstrative *zhe* ‘this’) are more common than *ci-ge*-N phrases and *na-ge*-N phrases (with the demonstrative *na* ‘that’) combined (Wang 2018: 770). It seems that the rise of the *zhe-ge*-N phrases largely triggered this change in Late Middle Chinese. Note that Dem-N phrases were not eliminated after this change, as they were still used in the late 12th century, often for no clear reason (Yang-Drocourt 2004: 180-181). Dem-MW-N phrases are found in Late Middle Chinese texts, but they seem to be relatively infrequent (Wang 2018: 480).

By the Late Middle Chinese era, classifiers have also been used with interrogative words (e.g. *ana* ‘which’ and *ji* ‘how many’), with the quantifying determiner *mei* ‘every’, and in some separable words (Wang 2018: 762-763, 775-778; C. Zhang 2012: 207). In short, Late Middle Chinese classifiers are used in a few contexts without any numerals.

Lexically, the classifiers in Late Middle Chinese are overall comparable to Mandarin classifiers. The generic classifier *gè* remains the main generic classifier in Mandarin, and the “quasi-generic” classifiers *jiàn* and *zhī* are still frequent in Mandarin (C. Zhang 2012: 240-242, 246, 268; Wang 2018: 41-44; Chen 2024). However, it is not uncommon for a Late Middle Chinese classifier to undergo changes and become compatible with a different set of nouns in Mandarin. For example, the classifier *jià* was used with nouns denoting buildings in Late Middle Chinese, while its reflex in Mandarin is used with other nouns including *tizi* ‘ladder’ and *gangqin* ‘piano’ (C. Zhang 2012: 245; Chen 2024: 178).

To sum up, the classifiers in Late Middle Chinese have three notable features. First, they are typically used in Num-Cl-N phrases. Second, they can occur in a few types of NPs that do not contain any numerals, notably post-verbal bare Cl-N phrases and Dem-Cl-N phrases. Lastly, the most frequent classifier is *gè* (which is the only generic classifier in Late Middle Chinese), followed by *zhī*, *jiàn*, and the most common shape classifiers such as *tiáo* and *piàn* (C. Zhang 2012: 233-271; Wang 2018).

4.1.4 Modern Chinese languages: Mandarin, Northern Wu and Cantonese

The last stage in the evolution of Chinese classifiers to be discussed here is the modern era. The most noteworthy changes from Late Middle Chinese to the three modern

Chinese languages will be summarized in this subsection, and I start with the changes that occurred in Mandarin.

Relatively few changes have happened in the use of classifiers in Mandarin since the end of the Late Middle Chinese era (in the 13th century). One of the few new constructions in Mandarin is the Num-Adj-Cl-N construction in which some classifiers can occur after a numeral and a dimensional adjective (Li 2013: 56-61). In addition, the classifier *gè* can be used after certain prepositions (viz. *lián* ‘even’ and *wèi* ‘for’) in the P-Cl-N construction, and such NPs are indefinite and singular like the post-verbal bare Cl-N phrases in Mandarin, but other classifiers are normally incompatible with that construction (C. Zhang 2012: 214-215). The non-generic classifiers have become more common in the bare Cl-N phrases and Dem-Cl-N phrases, and the classifier is generally obligatory in the Dem-Cl-N phrases in Standard Mandarin, except when it is used with a noun denoting human beings (Yang-Drocourt 2004: 175-179; Aikhenvald 2000: 208).¹⁴

The classifier system and its syntax in Northern Wu have undergone more radical changes since the 13th century. The most remarkable change is the creation of a variety of constructions with semantically definite classifiers. Among these types of NPs in Northern Wu, only Adj-Cl-N phrases are also found in Late Middle Chinese, while all the others (viz. pre-verbal Cl-N phrases, possessor-Cl-N phrases and RelCP-Cl-N phrases) are most likely created by relatively recent changes. However, the Adj-Cl-N phrases in Late Middle Chinese usually contain the adjective *hao* ‘good’ and the classifier *ge*, and they occur in exclamatory sentences (C. Zhang 2012: 192-193). Since the Adj-Cl-N construction in Northern Wu and that in Late Middle Chinese are lexically and semantically very different, the former could be an entirely new construction as well. As for the reason why Northern Wu classifiers can occur in those four types of NPs, a plausible explanation is that they followed the same path of change as the generic classifier *gè*. In a southern dialect of Late Middle Chinese,¹⁵ *gè* had been used with adjectives as a modification marker by the 13th century (Wu 2005: 12). In Northern Wu, the corresponding morpheme *geq* is used with adjectives, possessor NPs and

¹⁴ The situation in colloquial Beijing Mandarin appears to be quite different, as Dem-N phrases without classifiers tend to be allowed with other nouns.

¹⁵ According to Wu (2005), that dialect is the ancestor of Wenzhou Wu (a variety of Southern Wu).

relative clauses. When the other classifiers became grammaticalized later in the history of Northern Wu, they did not become modification markers (possibly in order to avoid having too many modification markers in the language) and instead became definiteness markers, which may also serve as linking particles (e.g. in Adj-Cl-N phrases) with a definite meaning. This explains why Northern Wu has pre-verbal bare Cl-N phrases which are definite as well as post-verbal bare Cl-N phrases which are indefinite in meaning. The latter is a descendant of the bare Cl-N phrases in Late Middle Chinese (Yang-Drocourt 2004), while the former is a relatively new construction that was created after the classifiers became associated with definiteness in Northern Wu.

Two changes happened in the Dem-Cl-N phrases in Northern Wu. First, the classifier has become obligatory in Shanghai Wu (Qian 1997: 110), and the only exception in Ningbo Wu is that nouns referring to humans do not require a classifier. In other words, classifiers are used more systematically in Northern Wu than in Late Middle Chinese (Yang-Drocourt 2004: 180-181). Second, the results of my survey indicate that Dem-Num-Cl-N phrases are ungrammatical in Northern Wu if the numeral is *iq* ‘one’, and Dem-Cl-N phrases must be used instead. Dem-*one*-Cl-N phrases are found in Middle Chinese texts written in the 9th and 10th centuries (Wang 2018: 773-774), and this construction is lost in Northern Wu.

Lexically, the most noteworthy change in Northern Wu is the use of *tsaq* as the main generic classifier. Its etymon is the quasi-generic classifier 隻 *zhī* in Middle Chinese,¹⁶ which is compatible with various nouns including *ge* ‘pigeon’, *lū* ‘shoe’ and *chuan* ‘boat’ (C. Zhang 2012: 268; Wang 2018: 40-44). In Northern Wu, the classifier *tsaq* is employed with nouns for virtually all animals, body parts, fruits, places, and many tools and containers (e.g. *khaedon* ‘bucket’) (Zhu 2006: 72). In addition, it is compatible with some abstract nouns (e.g. *zenwu* ‘task’ in Shanghai Wu and *loemaon* ‘dream’ in Ningbo Wu) and can occur with some nouns denoting humans when they are used pejoratively (Zhu 2006: 72; Luo 2005: 364).¹⁷ Nouns for recently invented objects often co-occur with this classifier (e.g. *tsaq-fizoe* ‘the spaceship’ and *tsaq-di’in* ‘the movie/film’) (Luo 2005: 361; Zhu 2006: 72).

¹⁶ The original MC pronunciation of this morpheme is something like /tɕiek/ (Chen 2003: 226, 395), from which Mandarin *zhī*, Northern Wu *tsaq* and Cantonese *zek* are derived.

¹⁷ Note that the cognates of this classifier are very frequent in Xiang and Gan as well (Luo 2005: 360-361).

The other generic classifier is *goq* (or *geq* in Shanghai Wu), which continues the generic classifier *gè* in Late Middle Chinese and is used with abstract nouns and most nouns for humans (Zhu 2006: 72-73). The classifier *méi* (i.e. the primary generic classifier in Early Middle Chinese) has become virtually obsolete in colloquial Ningbo Wu.

Cantonese classifiers underwent parallel changes that only differ slightly from those in Northern Wu. First, Cantonese classifiers started to be used in pre-verbal Cl-N phrases, possessor-Cl-N phrases, Adj-Cl-N phrases and RelCP-*go*-Cl-N phrases, all of which have a definite interpretation (Matthews & Yip 2011: 104-129, 328-330). While the post-verbal bare Cl-N construction is inherited from Late Middle Chinese, its meaning has become ambiguous with both definite and indefinite interpretations (Yang-Drocourt 2004: 176; Matthews & Yip 2011: 106). Second, the classifier in Dem-Cl-N phrases became obligatory in Cantonese (Matthews & Yip 2011: 107). Third, Dem-*jat*-Cl-N phrases with the numeral *jat* 'one' has become degraded and less acceptable than Dem-Cl-N phrases. Lastly, some sortal classifiers have been created in Cantonese, including 棟 *dung* for buildings, 圍 *wai* for tables and the plural classifier 啲 *di* (Matthews & Yip 2011: 115, 123; C. Zhang 2012; Wang 2018). The classifier *zek* (i.e. the reflex of the Middle Chinese quasi-generic classifier 隻 *zhī*) has become compatible with a somewhat larger set of nouns, as it can occur with *gau* 'dog' and some nouns for humans when used pejoratively (Matthews & Yip 2011: 122; Luo 2005: 362-364; Wang 2018: 41-44). However, *go* (which continues the generic classifier in Late MC) is still the main generic classifier in Cantonese (Matthews & Yip 2011: 120). The lexical changes that Cantonese and Northern Wu underwent seem to be more different than the syntactic and semantic changes.

In conclusion, the use of classifiers has undergone many far-reaching changes over the past three millennia. Chinese became a classifier language at the beginning of the Early Middle Chinese era. Num-Cl-N phrases have been the main type of quantified NPs since the Late Middle Chinese era, and Dem-Cl-N phrases and post-verbal bare Cl-N phrases started to occur during that period. More recently, in Cantonese and Northern Wu, the classifiers acquired the additional function of definiteness markers and began to be used in some new contexts. The most notable features of the classifier systems in Old Chinese (OC), Middle Chinese (MC), and the three modern Sinitic languages are summarized in Table (5).

Table 5: Major changes in the history of Chinese classifiers¹⁸

Feature Language	OC	Early MC	Late MC	Mandarin	Northern Wu	Cantonese
Num-N or N-Num	Dominant	Rare	Rare	No	No	No
N-Num-Cl	Infrequent	Frequent	Rare	Rare	Rare	Rare
Num-Cl-N	Rare	Frequent	Dominant	Dominant	Dominant	Dominant
Dem-Cl-N (with Cl.)	No	No	Frequent / dominant	Dominant	Dominant	Obligatory
Post-V bare Cl-N	No	No	Yes	Yes	Yes	Yes
Pre-V bare Cl-N (def.)	No	No	No	No	Yes	Yes
Modified Cl-N (def.)	No	No	No	No	Yes	Yes
Repeater classifiers	Yes	No	No	No	No	No
Generic classifiers	None	枚 <i>méi</i> & 個 <i>gè</i> ¹⁹	個 <i>gè</i>	個 <i>gè</i>	隻 <i>tsaq</i> & 個 <i>goq</i>	個 <i>go</i> (& 隻 <i>zek</i>)

(Yang-Drocourt 2004; C. Zhang 2012; Zheng 2016; Wang 2018; Matthews & Yip 2011)

With these changes in mind, we shall consider, in the next section, the question of why the classifiers appeared in Chinese and evolved in this way.

¹⁸ I only include the features of the colloquial language of the time in this table. Vestigial constructions in literary texts or idiomatic expressions are excluded. Here the features of Northern Wu are based on those of Ningbo Wu.

¹⁹ When there are two generic classifiers, the first one is more common than the second one.

4.2 Changes in the classifier system as grammaticalization

The changes mentioned in Section 4.1 may seem diverse and unsystematic at first glance. In this section, I argue that there exists an underlying driving force behind these changes, and that most of them are instances of grammaticalization. Grammaticalization is the term for a type of changes from more lexical and independent words or morphemes into more grammatical and dependent ones (Aikhenvald 2000: 374). It consists of several subtypes of changes such as cliticization, phonetic erosion, semantic generalization and pragmatic enrichment (Hopper & Traugott 2003; Bisang 2011). I will demonstrate that the morphosyntactic, semantic and phonetic subtypes of grammaticalization have all occurred in the history of Chinese classifiers.

First, when some nouns (and other lexical items) became classifiers in Old Chinese or Middle Chinese, they lost their syntactic freedom and could no longer be used freely and independently in most contexts. Etymologically, virtually all Chinese classifiers are derived from nouns, verbs or adjectives at some point in the history. For example, the classifier 个 *gè* is derived from the adjective 介 *jiè* ‘alone, sole’,²⁰ and the classifier 枚 *méi* used to be a noun with the lexical meaning of ‘tree trunk’ or ‘small piece of wood’ (J.-P. Li 2017: 58-60; Chen 2024: 258). Likewise, the shape classifier 張 *zhāng* originated as a variant of the corresponding verb, which means ‘to draw a bow, to spread out’ in Old Chinese (J.-P. Li 2017: 84-86; Chen 2024: 429).²¹ Crucially, the nominal, verbal or adjectival etyma of these classifiers could occur more freely in more positions. For instance, nouns can be used as topics, subjects or objects, whilst classifiers (which are mostly derived from nouns) cannot be used independent, as they generally co-occur with numerals in quantified NPs in Old Chinese and Early Middle Chinese (Yang-Drocourt 2004).

Second, classifiers have become morphologically fossilized and incompatible with most (if not all) affixes. In Mandarin, two of the most common nominal suffixes (viz. *-zi* and

²⁰ It merged with the homophonous classifier 箇 *gè* for bamboos around the 3rd century CE (J.-P. Li 2017: 60-62).

²¹ More classifiers are derived from nouns than from verbs or adjectives in Mandarin (Li 2013: 24; Chen 2024).

-*tou*) are never used with any classifiers, and the other common suffix *-r* is compatible with a minority (i.e. less than 20%) of the sortal classifiers in the language (Chao 1968: 589-593). In Northern Wu, the classifiers are always morphologically simplex, and none of the affixes for nouns (e.g. *-tsy*, *-deu*, *-n*, and *aq-*) are compatible with them (Zhu 2006: 57-58). This is a typical phenomenon observed during grammaticalization and is comparable to the loss of person and number inflection in some auxiliaries and modals in English (e.g. **He wills...*) (Campbell 2013: 281; Hopper & Traugott 2003: 55; Bisang 2011). In short, Chinese classifiers have undergone morphological fossilization.

Third, the classifiers in Chinese underwent pragmatic enrichment and semantic generalization simultaneously, which are correlated processes (Hopper & Traugott 2003: 94). In Early Middle Chinese, classifiers only individuate and classify nouns.²² In Late Middle Chinese, as the classifiers started to occur in post-verbal bare Cl-N phrases, they became associated with the indefinite meaning (Yang-Drocourt 2004). In Northern Wu and Cantonese, the classifiers acquire an additional definite meaning as they begin to be used in pre-verbal bare Cl-N phrases and modified Cl-N phrases which are definite. Consequently, when a classifier is used in a post-verbal bare Cl-N phrase in Cantonese, it marks the noun category, singularity, and definiteness or indefiniteness at the same time, which is more complex than the function of Early Middle Chinese classifiers (Matthews & Yip 2011: 106; Yang-Drocourt 2004). Another type of semantic generalization is related to lexical changes. The most notable change of this type is the loss of the repeater classifiers towards the end of the Old Chinese era. For example, the classifier for *yang* ‘sheep’ can be the repeater classifier *yang* in Old Chinese, while in Mandarin the classifier *zhī* is used for *yang* ‘sheep’ and many other animals (J.-P. Li 2017: 172-175). Similar changes happened to some shape classifiers as well. For instance, the classifier *tiáo* has always been used with nouns that denote long and slender objects since the Old Chinese era; in Mandarin it is also compatible with certain nouns for places and humans as well as some abstract nouns (e.g. *faqui* ‘law’) (J.-P. Li 2017: 68-70; Chen 2024: 361-363; C. Zhang 2012: 260-261).

²² I assume that the generic or kind reading of bare nouns is the most basic interpretation in Middle Chinese, since there is good reason to believe that the bare nouns in Mandarin “denote kinds as default” (Li 2013: 88).

Lastly, Chinese classifiers are phonetically reduced to varying degrees. In principle, the etyma of the classifiers that were still nouns, verbs or adjectives in Old Chinese should be able to bear stress just like other nouns, verbs or adjectives. By contrast, classifiers cannot be stressed in Mandarin except when they are contrastively focused, which rarely happens to sortal classifiers (Li 2013: 32-33). In Northern Wu, a Num-Cl-N phrase with a stressed classifier sounds odd, while the numeral and the head noun can both be stressed. In addition, one of the generic classifiers, namely 个 *geq* in Shanghai Wu (corresponding to *gè* in Mandarin) has undergone vowel reduction from /u/ to /ə/ (cf. the pronunciation *ku* in 个人 *kunin* ‘individual’). Moreover, this classifier is often pronounced /həʔ/ instead of /gəʔ/ in connected speech, which is an instance of lenition (Zhu 2006: 72). The same type of lenition also happens to the classifier *goq* in Ningbo Wu in many contexts. It seems plausible that the generic classifier *gè* had become stressless by the 12th century, when it was first used with the interrogative pronoun 甚麼 *shenme* ‘what’ for seemingly euphonic reasons, in view of the fact that it does not have any syntactic or semantic roles to play in that position (Yang-Drocourt 2004: 177).²³

Since Chinese classifiers have undergone all these changes which can be subsumed under grammaticalization, it seems reasonable to argue that grammaticalization played a major role in the evolution of the classifiers in Chinese. It is then helpful to determine the stages of the evolution of Chinese classifiers in terms of grammaticalization. At the first stage, the classifiers are grammaticalized to a limited extent and are not very different from their lexical sources; this stage corresponds to the Old Chinese era. At the second stage, the classifiers lose their syntactic freedom completely and become dependent on the numeral. This corresponds to the Early Middle Chinese era. At the third stage, the classifiers start to occur in some other types of NPs and their meaning becomes more complex. The generic classifier is probably becoming stressless and somewhat weakened. This stage corresponds

²³ Note that the effects of grammaticalization is also observable in some lexical changes. For example, the generic classifier *méi* is incompatible with abstract nouns in Early Middle Chinese (Zheng 2016: 207-208). Since the Late Middle Chinese era, generic classifiers (including Mandarin *gè* and Northern Wu *tsaq*) are compatible with abstract nouns even though they were initially only used with concrete nouns (C. Zhang 2012: 240-242, 268).

to the Late Middle Chinese era, and the level of grammaticalization of Mandarin classifiers has not really passed this stage yet. At the fourth stage, the classifiers are fully fledged functional morphemes that are pragmatically versatile and semantically complex. They are syntactically and morphologically distinct from words of other categories including nouns. They can no longer bear stress, and some of the most common classifiers may undergo phonetic reduction (e.g. having a reduced vowel or a weakened consonant). The classifier systems of Northern Wu and Cantonese are at this stage.

To sum up, most of the changes in the history of Chinese classifiers are related to the grammaticalization of those morphemes. This process turned the prototypes of classifiers with lexical meanings in Old Chinese into functional morphemes in the modern Chinese languages. In the next section, I compare the evolution of Chinese classifiers and that of measure words, and I discuss the synchronic effects of the diachronic differences.

4.3 A diachronic comparison of Chinese classifiers and measure words

When the diachronic changes in Chinese classifiers were described in section 4.1, the changes in measure words were also briefly mentioned. Since classifiers and measure words share the individuating function, it is unsurprising that they are syntactically similar and have undergone parallel changes in much of their history. In this section, I summarize the similarities and differences between Chinese classifiers and measure words in each of the historical eras, and I argue that they can be structurally similar or different depending on the extent to which they are grammaticalized.

During the Old Chinese era, both classifiers and measure words gradually increased in number (J.-P. Li 2017). Yang-Drocourt (2004: 56) proposes that the N-Num-MW phrases existed in Chinese before there were prototypes of classifiers, and that this usage extended first to N-Num-N' phrases (with echo-classifiers) and then to N-Num-Cl phrases with other

proto-classifiers.²⁴ This theory explains the fact that only a handful of non-repeater count classifiers appear in the oracle-bone inscriptions (13th-11th century BCE) while measure words seem to be somewhat more common (Wu 2005: 5; J.-P. Li 2017: 232-233, 280). In the early texts, measure words are often used with a small number of nouns (e.g. 丙 *bīng* ‘a pair (of horses)’, 卣 *yóu* ‘a big oval jug (of fragrant alcohol)’ (J.-P. Li 2017: 204, 238-239). It is worth noting that two “generic” container measure words (i.e. 盛 *chéng* and 器 *qì* (both ‘container’) appeared towards the end of the Old Chinese era (i.e. 4th century to early 2nd century BCE), when the generic classifiers 枚 *méi* and 個 *gè* were also emerging (J.-P. Li 2017: 49-63, 236-238). I propose that Old Chinese already has a grammatical category for the measure words, and that they are only grammaticalized to a limited extent.

There are few changes in the use of measure words in Early Middle Chinese except for the increasing frequency of Num-MW-N phrases (which are rare in Old Chinese) (Yang-Drocourt 2004: 83, 135). Many lexical changes have happened in the measure words by the Early Middle Chinese era, including the loss of the generic container measure words (J.-P. Li 2017: 237-238; Zheng 2016). The first loanwords in the measure words (e.g. 由旬 *youxun* from Sanskrit *yojana*, a unit of length) entered Chinese during this era (Wang 2018: 565). The meanings of the container and group measure words were usually close to that of their lexical sources, which are mostly nouns or verbs (Zheng 2016; Wang 2018).

In Late Middle Chinese, Num-MW-N phrases become more frequent than N-Num-MW phrases as Num-Cl-N becomes the predominant form of quantified NPs (H.-Q. Zhang 2014: 18-19; ASTC-EMC: *Zutangji*; Yang-Drocourt 2004: 173). Like the Dem-Cl-N phrases which appeared in Chinese at the beginning of this era, Dem-MW-N phrases occurred, albeit infrequently, in Late Middle Chinese texts (Wang 2018: 769-770; Yang-Drocourt 2004: 178). Post-verbal bare MW-N phrases appeared slightly earlier (in the 9th century), also around the same time as the first occurrences of post-verbal bare Cl-N phrases (Wang

²⁴ Note that many Tibeto-Burman languages (e.g. Burmese, Drung), which are distantly related to Chinese, also have echo-classifiers (Jiang 2009: 79-85; Greenberg 2013 [1974]: 309). It is possible that Old Chinese and those Tibeto-Burman languages both inherited the N-Num-N’ construction from Proto-Sino-Tibetan. If that is the case, then N-Num-N’ phrases must have existed for many centuries before the Old Chinese era and were disappearing.

2018: 787-788; Yang-Drocourt 2004: 176). In short, quantified NPs with measure words and those with classifiers underwent very similar changes simultaneously in Late Middle Chinese. This suggests that the classifiers and measure words belong to one and the same grammatical category in Late Middle Chinese.

After the Late Middle Chinese era, syntactic differences between measure words and classifiers emerged in Cantonese and Northern Wu. Most notably, Cantonese and Northern Wu classifiers can be used in definite NPs without demonstratives (e.g. in pre-verbal bare Cl-N phrases), while such NPs with measure words are less acceptable and extremely rare if not non-existent. In other words, classifiers have the function of definiteness markers in some contexts, while measure words do not. Furthermore, Cantonese and Northern Wu have acquired the Num-Adj-MW-N construction which does not seem to exist in Middle Chinese (possibly from Mandarin), but the similar Num-Adj-Cl-N construction is not really used in these two modern languages (C. Zhang 2012; Wang 2018).²⁵ This difference is likely attributable to the tendency for measure words to retain their lexical meanings and to be closer to the corresponding nouns. For example, *dou-un* 'big bowl' occurs quite naturally in Northern Wu as an Adj-N sequence and as a complete NP, which can be reanalyzed as an Adj-MW sequence in another context. On the other hand, Adj-Cl sequences such as *dou-tsaq* 'big-classifier' do not occur elsewhere in Northern Wu, and reanalysis cannot happen. In view of these phenomena, I propose that the classifiers have become fully functional in Northern Wu and Cantonese, and that the measure words are still semi-lexical.

Typologically, the [Num-MW] constituency seems to be universal. The relationship between numerals and measure words is analogous to that between multipliers and bases in complex numerals both semantically and in terms of word order, as there is a cross-linguistic correlation between Num-MW order and multiplier-base order (Her 2017A). In other words, a measure word is always expected to merge with the numeral in a quantified NP. This should also be true of the classifiers that are only slightly grammaticalized and exhibit the same syntactic behaviour as measure words. I propose that, at this point, the measure word and the classifier belong to the category of Unit words, and that the Unit Phrase is an adjunct in the Numeral Phrase (e.g. in Old Chinese) or the complement of that

²⁵ Num-Adj-Cl/MW-N phrases had appeared in Early Mandarin by the late 16th century (Ye 2008: 108-109).

phrase (e.g. in Mandarin). As unit words (e.g. *dozen*) and bases (e.g. *hundred*) are similar both functionally and semantically, the Num-Unit order correlates with the multiplier-base order in the world's languages (Her 2017A). However, when the classifiers become highly grammaticalized (e.g. in Cantonese and Northern Wu), the NPs acquire the additional Classifier Phrase structure, and the classifiers merge with the noun first within the CLP, and Num-MW-N phrases (with the [Num-MW] constituent including MW as Unit⁰) and Num-Cl-N phrases (with the [Cl-N]_{CLP} constituent) become structurally different. Note that the surface word order Num-Cl-N remains unaffected by this change.

To sum up, the evolution of classifiers and that of measure words share three similarities and are different in two ways. They occur in the same constructions in Middle Chinese and Mandarin and were affected by the same change in constituency order (i.e. N-Num-Cl/MW to Num-Cl/MW-N) in Middle Chinese. They are pragmatically and lexically enriched throughout the history and still share very similar syntactic behaviour in Mandarin. As for the differences, measure words were employed more systematically than the prototypes of classifiers in the earliest texts in Old Chinese. In addition, the classifiers and measure words in Cantonese and Northern Wu appear in different types of NPs, and their functions and meanings are different to a considerable extent.

4.4 Conclusion

In this chapter, I described the most notable diachronic changes related to Chinese classifiers and argued that most of those changes are triggered by the grammaticalization of classifiers. I also compared the use of classifiers and that of measure words at different stages in the history. The most important finding is that Chinese classifiers were once fully lexical nouns, verbs or adjectives (i.e. the prototypes of classifiers in Old Chinese) but gradually lost their lexical meanings and acquired grammatical functions. Similar changes from nouns to classifiers have been observed in Hmong, Palikur and Korean (Aikhenvald 2000: 355-356; Y.-H. Huang 2005: 484-485), and processes creating classifiers from verbs are found in some Mayan languages (Aikhenvald 2000: 364). The syntactic differences between classifiers and measure words in Northern Wu and Cantonese arose relatively late

in the history. The differences are attributable to grammaticalization as well: the classifiers have been grammaticalized to a very high level, which the measure words cannot attain as they must (at least partly) retain their lexical meanings.²⁶

On the other hand, the classifiers and measure words in Mandarin are syntactically much more similar. This is related to the fact that Mandarin classifiers are less versatile and fossilized to a lesser extent, and these features suggest that Mandarin classifiers are still semi-lexical. Since that is a level of grammaticalization that measure words can reach, the classifiers and measure words in Mandarin are both semi-lexical and belong to one grammatical category.

²⁶ For example, when the noun from which a container or group measure word originates falls out of use in Chinese (e.g. 緋 *gun* 'woven ribbon', 簍 *dan* 'small bamboo basket'), the measure word generally becomes obsolete as well (J.-P. Li 2017: 213, 261).

5. Theoretical implications and related issues

In this chapter, I briefly explore a number of theoretical issues related to classifiers, such as the mass-count distinction, grammatical number in Chinese and other classifier languages, and the optionality of Chinese classifiers. I argue that the Chinese languages only have mass nouns and do not have a mass-count distinction. Number is marked with special classifiers rather than affixes in these languages, and numeral classifiers and plural morphology on nouns can co-occur in some languages. Lastly, I propose that classifiers tend to be optional with complex numerals as the multiplicand may be reanalyzed as a dummy unit word, and that the optionality of classifiers in Chinese is also affected by a few other factors.

In addition, I discuss the main implications of my findings and theory about Chinese classifiers for Chinese syntax, linguistic typology and diachronic syntax. These implications include a rethinking of syntactic differences among the Sinitic languages, a new explanation for the constituent order in NPs with classifiers and a prediction about the likely features of classifiers that behave differently from measure words.

Section 5.1 is about the issues related to Chinese classifiers. The implications of this study are discussed in Section 5.2. Section 5.3 is the conclusion of this chapter.

5.1 Relevant issues

This section consists of three subsections, each about a topic related to classifiers. In subsection 5.1.1, I discuss the (non)existence of the mass-count distinction in the Chinese languages and in other languages. In subsection 5.1.2, I examine whether and how plurality is marked morphologically in Chinese. Subsection 5.1.3 is about the optionality of Chinese classifiers, and I shall describe a few factors including one that seems cross-linguistically common. I start with the mass-count distinction in subsection 5.1.1.

5.1.1 The mass-count distinction in Chinese and beyond

In previous studies, the existence of a classifier system is often linked to the absence of a mass-count distinction in classifier languages. Chierchia (1998: 354) lists both “no PL” (i.e. no plural morphology) and “generalized classifier system” as properties of argumental languages such as Mandarin Chinese. Cheng & Sybesma (1999: 514-517) argue that there is a mass-count distinction in Chinese nouns²⁷ and support this claim with their argument that the count classifiers and mass classifiers in Chinese are syntactically different. According to Borer (2005: 93-94), all nouns are universally mass (by default) and require a stuff divider to “interact with the ‘count’ system”; the stuff divider can be a classifier or plural inflection, which are essentially the same thing as they are in complementary distribution (i.e. used in different languages or in different contexts in the same language).

My view differs from all these three theories. I suggest that classifiers and number inflection have partly different functions, and that the costly number-marking systems with both are naturally uncommon but still attested. Relatively recent studies indicate that the simple complementary distribution of classifiers and plural inflection does not hold, and that there are in fact four types of languages instead of two. First, there are “mass/count languages” such as English where some nouns are compatible with plural markers while other nouns are not. Second, there are “count” languages such as Yudja [Tupian] in which all nouns are compatible with numerals, and there are no classifiers and no plural marking. Third, there are “mass” languages including Mandarin where plural inflection is absent and nouns must co-occur with classifiers (or measure words) to be modified by a numeral (Li 2013: 71-72). The fourth type of languages includes Ejagham [Niger-Congo], which has plural prefixes as well as numeral classifiers, and they can co-occur in the same NP (Bisang

²⁷ Note that I am only discussing the mass-count distinction as a grammatical distinction here. There is little doubt that Chinese speakers have the capability to perceive the ontological difference between discrete objects (e.g. trees) and homogeneous entities that are not spatially delimited (e.g. water) (Li 2013: 81-82).

2012: 24-26). In short, it is typologically possible for languages to have neither classifiers nor number marking or both classifiers and number inflection.²⁸

Furthermore, classifiers have a classifying function that number morphology lacks. For example, different nouns select different classifiers in Mandarin, while the plural suffix -s in English is compatible with most nouns. Although there are mass/count languages (e.g. Classical Greek) with declension classes that affect how nouns are pluralized, it seems that such inflectional classes tend to be selected based on morphophonological criteria, unlike the selection of classifiers which is “predominantly semantic” (Aikhenvald 2000: 98). It is also worth noting that relatively few languages have both numeral classifiers and genders or noun classes (Aikhenvald 2000: 202).

A better generalization is that the basic functions of classifiers are a combination of the function of number morphology (i.e. individuating or “stuff-dividing”) and that of noun classes (i.e. classifying). The nominal systems with both classifiers and plural marking (and those with both classifiers and noun classes) are possible but costly, hence their rareness. The account proposed in Chierchia (1998) is overly general and simplistic, because the use of bare arguments, the existence of a classifier system and the absence of plural marking are not dependent on one another. Cheng & Sybesma’s (1999: 515-6) arguments about the syntactic differences between the count classifiers and mass classifiers in Mandarin turned out to be mostly spurious (Li 2013: 49-67).²⁹ Consequently, the mass-count distinction that they propose for Mandarin nouns is no longer tenable.

Regarding Borer’s (2005) claim about nouns being universally mass and requiring stuff dividers to be quantified, it seems largely compatible with the patterns in the Chinese languages. However, this theory may not be able to account for the reason why the nouns in Yudja do not co-occur with stuff dividers and why Ejagham nouns need two types of stuff dividers (viz. plural marking and numeral classifiers) (Li 2013; Bisang 2012). Furthermore, the mass nouns in French with obligatory dividers (e.g. *du vin* ‘(some) wine’) may also pose

²⁸ In some languages with neither classifiers nor plural morphology, it is possible to distinguish count nouns from mass nouns because only the former can combine directly with numerals, and this is the case in Dëne Suliné [Na-Dené] (Li 2013: 71).

²⁹ I have discussed their arguments and Li’s (2013) counterarguments in the last part of section 2.1.

a problem, as mass nouns are expected to be unmarked, although the significance of this problem depends on whether the French-type languages (with obligatorily non-bare mass nouns) are typologically common or exceptional (Chierchia 1998: 355).

I propose that all nouns in the three Sinitic languages are mass, and that it is largely independent of the difference between classifiers and measure words. Note that nouns can only be grammatically (i.e. morpho-syntactically) mass, and that the ontological properties of their referents (e.g. discrete, dimensional) are only indirectly related to the grammatical mass-count distinction (Li 2013: 82). Although individual classifiers tend to co-occur with some nouns that are labelled “count” and measure words with others (labelled “mass”), this correlation must not be exaggerated. First, nearly all nouns are compatible with one or two (if not more) measure words. For example, Num-MW-N phrases such as *yi-che-xuesheng* ‘a busload of students’ occur naturally in Mandarin, even though the noun *xuesheng* ‘student’ refers to a discrete individual and is normally translated with count nouns in mass/count languages such as English and French. Second, many nouns that are typically mass can select some special unit words from a group of “individuating [classifiers]” (Zhang 2013: 164). Some individuating classifiers are shown in the examples below.

(1) 两 剂 汤药 [Mandarin]

liang ji tang-yao

two Cl_{=dose} liquid-medicine

‘Two doses of liquid medicine’

(Based on Zhang 2013: 164, example (338-a))

(2) 两 滴 红酒

liang di hong-jiu

two Cl_{=drop} red-wine

‘Two drops of red wine’

In (1), the individuating classifiers for *tang-yao* ‘liquid medicine’ is *ji* ‘dose’, which is normally used as the minimal unit of this noun. In (2), the individuating classifiers for *hong-*

(Qian 1997: 88), and this use of *tsaq* does not seem to have a stylistic effect. Likewise, the classifier *zek* in Cantonese is a frequent sortal classifier for animals and round objects, as well as a kind classifier meaning ‘variety’ (Matthews & Yip 2011: 122). In short, although Chinese nouns can be ontologically count-like (i.e. referring to discrete objects) or mass-like, they co-occur with similar types of unit words which may even be lexically identical.

To sum up, all nouns in the three Sinitic languages under review are mass, since they do not take number marking and cannot merge directly with numerals.³¹ Mass nouns must undergo individuation before merging with numerals and other quantifiers, normally with the help of classifiers or measure words in the Chinese languages. Classifiers have the additional function of classifying nouns and co-occur with number marking in certain languages. The link between the mass-count distinction and the classifier-MW distinction in Chinese is dubious and is at least weaker than previously proposed (contra Chao 1968: 508; Cheng & Sybesma 1999). The mass-count distinction does seem to be a genuine syntactic distinction in some other languages, and some languages might only have underlyingly count nouns.³²

In the next subsection, I describe a few ways to mark grammatical number in three Sinitic languages and differentiate plural marking from collective marking.

5.1.2 Grammatical number in Chinese

Contrary to the belief that plurality is not marked in Chinese (Chierchia 1998), there are in fact a few different types of plural marking, including suffixes, special classifier forms and reduplication. I shall start with the plural suffixes.

³¹ The only possible exception is the nouns for humans in Mandarin, which can take optional plural morphology (if it is semantically definite) and compatible with some numerals. This is discussed in more detail in subsection 5.1.2.

³² An examiner points out that only some Chinese nouns can co-occur with size and shape adjectives. I have no doubt about its empirical validity, but that difference among nouns just reflects the ontological differences between discrete and homogeneous objects. Here I focus on the grammatical mass-count distinction between nouns rather than the physical differences between the referents of nouns.

In Mandarin, some nouns that are semantically plural can sometimes be followed by the suffix *-men*. Unlike the plural suffix *-s* in English, this suffix in Mandarin is could only co-occur with nouns denoting humans and pronouns, making the nouns definite, and it never occurs in Num-Cl-N phrases (Y.-H. A. Li 1999: 77-78; Cheng & Sybesma 1999: 536-537). It is portrayed as a collective suffix in Cheng & Sybesma (1999) due to its limited distribution and compatibility with personal pronouns and proper names. However, according to Y.-H. A. Li (1999), *-men* is a plural suffix that is generated at Num⁰ and raises to D⁰, where it is attached to a pronoun or a noun raising from N⁰ if there is no intervening Cl⁰.

I propose that the suffix *-men* is a plural marker, and that it has certain functions of a collective suffix only for some speakers. When co-occurring with common nouns, it simply makes the noun plural with no collective force. For example, the plural NP *xuesheng-men* (i.e. *student-PL*) must refer to be a specific group of students (and only students) instead of a student or some students together with the people associated with them (Y.-H. A. Li 1999: 94). This analysis also explains the fact that this suffix is not obligatory in definite nouns denoting humans: its optionality is attributable to the ambiguity of the bare nouns, which can be definite and plural in Mandarin (Cheng & Sybesma 1999: 528). For the speakers who accept the use of this suffix with proper names, it can also serve as a collective suffix. For instance, the NP *Xiao Qiang-men* means ‘Xiao Qiang’s group’, including others that are considered similar to or associated with Xiao-Qiang (Cheng & Sybesma 1999: 536). Since relatively few speakers of Mandarin accept this use of *-men* (Y.-H. A. Li 1999: 94), and it is more frequently used as a plural marker with common nouns and pronouns, I consider it to be a plural suffix rather than a collective suffix. Y.-H. A. Li’s (1999) account appears to be largely correct, although classifiers may be incompatible with the plural suffix *-men* for a different reason. Semantically, the Num-Cl-N phrases and bare Cl-N phrases in Mandarin are indefinite and specific in terms of quantity, whilst NPs with *-men* are definite and vague in terms of quantity, and it is plausible that they are semantically incompatible.³³

³³ Another possibility is that the suffix *-men* is in fact incompatible with numerals. Since the numerals generally co-occur with classifiers (or measure words) in Mandarin, this suffix is also incompatible with classifiers. If that is the case, then the suffix *-men* is comparable to the plural markers in Hungarian and Turkish, which are not compatible with numerals either (Borer 2005: 116).

Northern Wu has a suffix that is *prima facie* very similar to the plural suffix *-men* in Mandarin. This suffix is pronounced *-laq* in Ningbo Wu and *-la* in Shanghai Wu, and it is a collective marker for nouns (Zhu 2006: 59; Qian 1997: 103). This suffix is used obligatorily in the plural forms of personal pronouns in Northern Wu, where it is just a plural marker. However, with proper names and common nouns, it is usually used as a collective marker, although the plural interpretation is sometimes possible as well. Like the suffix *-men* in Mandarin, the *-la(q)* suffix is definite in meaning and incompatible with classifiers, and they are only used with nouns denoting humans. The use of this collective suffix in Shanghai Wu is shown in (5), which is compared to a very similar sentence in Mandarin in (6).

- (5) 搿 趟 聚会， 娘舅拉 是 勿 会 去 个。 [Shanghai Wu]
 geq thaon zhiuwe nianjieu-la zy veq we chi geq
 this CL gathering maternal.uncle-LA COP NEG FUT go SFP
 'The uncle and his family will not go to this gathering.'
 (Qian 1997: 103)

- (6) 这 次 聚会， 舅舅们 是 不 会 去 的。 [Mandarin]
 zhe ci juhui jiujiu-men shi bu hui qu de
 this CL gathering maternal.uncle-MEN COP NEG FUT go SFP
 'The uncles will not go to this gathering.'

In (5), the NP with the suffix *-la* refers to a specific (maternal) uncle and the group of people associated with him, which is assumed to be his family in this case. Only one person of that group is an uncle of the speaker, while all the other ones are not. The speaker does not need to have two or more maternal uncles for (5) to be interpretable. In short, the *-la* suffix should be analysed as a collective marker rather than a plural marker.

The Mandarin sentence corresponding to (5) is shown in (6), and the suffix *-men* is attached to the noun *jiujiu* 'maternal uncle'. This NP refers to a set of uncles of the speaker and no one else, and it presupposes that the speaker has at least two maternal uncles. The suffix *-men* should, therefore, be considered a plural marker instead of a collective suffix.

With other common nouns that are not kinship terms, the *-la(q)* suffix in Northern Wu marks both plurality and collectivity, making the NP ambiguous in some cases (Qian 1997). It also occurs freely with proper names where only the collective reading is allowed (Zhu 2006: 59). In view of the additional definite meaning of this collective suffix (like the plural suffix *-men* in Mandarin), I propose that it merges with nouns at D⁰, which explains why demonstratives are normally incompatible with this suffix. It seems that the Northern Wu suffix *-la(q)* is more comparable to the Burmese suffix *-do* (with a phonetic variant *-to*) than Mandarin *-men* (cf. Y.-H. A. Li 1999: 92-93). Both Northern Wu *-la(q)* and Burmese *-do* are used to form plural personal pronouns, and they express the “associative” or collective meaning with common nouns (Jenny & Hnin Tun 2016: 129-130). On the other hand, they differ in two major ways. First, *-do* and classifiers are compatible in Burmese N-*do*-Num-Cl phrases, whilst Num-Cl-N-*laq* phrases are ungrammatical in Northern Wu. Second, *-do* can be used with nouns for non-human referents such as books and dogs in Burmese,³⁴ while the suffix *-la(q)* is not compatible with those nouns (Jenny & Hnin Tun 2016: 130-131).

Cantonese does not have a true plural or collective suffix, and the only morpheme that is roughly comparable to Mandarin *-men* is the suffix *-dei* which is the plural suffix for the personal pronouns in Cantonese (Matthews & Yip 2011: 45). It is only compatible with one noun (i.e. *jan* ‘person’), and the meaning of *jan-dei* is not exactly compositional as it means ‘other people’ or ‘someone’ (which can be obvious from the context) (ibid. 92-93). The plural suffix *-mun* in formal Cantonese is borrowed from Mandarin *-men* and does not seem to occur in colloquial Cantonese (Matthews & Yip 2011: 93).

Plurality is in fact more often expressed using classifiers rather than with affixes in Chinese, and this is especially true for Northern Wu and Cantonese. That function can be fulfilled by a special plural classifier form or by classifier reduplication in each of the three Sinitic languages (Matthews & Yip 2011: 39). I start by describing the plural classifiers.

In Mandarin, plurality is sometimes encoded using the special classifier *xiē*, which co-occurs with the demonstratives and the numeral *yī* ‘one’, but not with other numerals (Aikhenvald 2000: 208). There is controversy as to whether it is a plural marker. According

³⁴ According to Bernot et al. (1997: 25-26), the collective suffix *-to/-do* is only compatible with nouns for animate beings. There might be variation in the use of this suffix among Burmese speakers.

to Chao (1968: 598), it is a partitive measure word which means “some, an amount of, a number of”. Aikhenvald (2000: 208) suggests that it is a plural classifier and mentions that it is compatible with the plural suffix *-men*. According to Wang (2018: 654-655), its etymon in Late Middle Chinese is an indefinite classifier that means ‘a small amount of’. Its meaning has been reasonably stable since the 12th century, as it denotes a small indefinite amount in Mandarin as well (Chen 2024: 399-400; Wang 2018). Zhang (2013: 110) points out that it does not create a plural-singular contrast and is used “with mass nouns such as *shui* ‘water’ and *mianfen* ‘flour’”. In other words, if this classifier has a plurality-marking function, it is not nearly as consistent as the suffix *-s* in English. The use of the classifier *xie* is illustrated in the example below.

- (7) 这件衬衫 还是 这些 衬衫? [Mandarin]
 zhe jian chenshan haishi zhe xie chenshan
 this CL shirt or this CL:PL shirt
 ‘This shirt or these shirts?’

In (7), the classifier *xie* makes the noun *chenshan* ‘shirt’ plural in meaning, and it does differ semantically from the usual sortal classifier for shirts (i.e. *jian*). Such sentences occur naturally in Mandarin when talking about the number of objects that are needed or wanted. In view of these sentences and its compatibility with the plural suffix *-men* (which is exceptional among classifiers), I propose that *xie* is a plural classifier for ontologically count nouns, even though it is not used very systematically. It appears that this classifier is evolving into a pure plural marker, as it can co-occur with the generic individual classifier *gè* (e.g. in *zhe-xie-ge-xuesheng* ‘these students’), although such NPs are still infrequent and *xie* is incompatible with other sortal classifiers. When it co-occurs with “mass” nouns (for homogeneous objects, e.g. *shui* ‘water’), *xie* is arguably an indefinite classifier.

The Northern Wu equivalent of Mandarin *xie* is the classifier *ngae* (Qian 1997: 95), and the “collective classifier” *di* in Cantonese also corresponds to that word (Matthews & Yip 2011: 115). The Northern Wu classifier *ngae* is very similar to Mandarin *xie*: it is used with demonstratives and with the numeral *iq* ‘one’ but not with other numerals, and it is

compatible with nouns denoting discrete things as well as those for homogeneous objects. Matthews & Yip (2011: 115) and the Cantonese data I collected indicate that these features are shared by the Cantonese classifier *di* as well. It seems reasonable to assume that both Northern Wu *ngae* and Cantonese *di* are plural classifiers. Like their Mandarin counterpart, they can also occur in bare Cl-N phrases. However, the bare Cl-N phrases with the plural classifier *xie* can only occur in the post-verbal position and is always semantical indefinite in Mandarin, while the plural classifiers in Northern Wu (*ngae*) and in Cantonese (*di*) can occur pre-verbally and can be definite or indefinite (Matthews & Yip 2011: 115). In this aspect, Northern Wu³⁵ *ngae* and Cantonese *di* behave like special classifiers rather than special measure words.

Another way of marking plurality in these Sinitic languages is by reduplicating the classifier.³⁶ The sortal classifier that is selected by a noun can be reduplicated in a Num-Cl-N phrase when the numeral is one, and the reduplication marks abundant plurality, that is ‘many’ rather than ‘a few’ (Zhang 2013: 115-117). This use of classifiers is shown in (8).

(8) 河里 漂着 一 朵朵 莲花。 [Mandarin]

He-li piao-zhe yi duo-duo lianhua

river-in float-DUR one CL-RED lotus

‘There are many lotuses floating on the river.’

(Zhang 2013: 116, example (230-a))

Example (8) illustrates how the classifier *duo* for flowers is reduplicated to mark the plurality of the noun *lianhua* ‘lotus’ which would have been singular if this classifier were not reduplicated. In Mandarin, the meaning of classifier reduplication differs from that of

³⁵ The plural classifier *ngae* appears to be somewhat less frequent in Shanghai Wu than in Ningbo Wu, as it can be replaced by the “indefinite classifiers” *ti* and *tiq* (Qian 1997: 95). In Ningbo Wu, *ngae* is used predominantly in all the contexts in which *xie* is used in Mandarin, except for *xie-ge* (i.e. the sequence of two classifiers mentioned in the previous paragraph), which does not have an equivalent in Ningbo Wu.

³⁶ Reduplication of nouns has been observed in a variety of languages, and it is often a way of marking plurality (e.g. in Washo and Malay) or distributivity (e.g. in Yoruba) (Xu 2012: 44; Yuko 2017).

the plural classifier *xiē* in that the former refers to a relatively large number of things, while the latter often refers to a small number of objects. For instance, one would use the NP *yi-xie-lianhua* ‘some lotuses’ rather than the NP with classifier reduplication in (8) to refer to two or three lotus flowers (Zhang 2013: 116). Unlike NPs with the plural suffix *-men*, NPs with classifier reduplication can be indefinite, and the referents of the nouns can be non-human things. This plural marking strategy is largely productive, applying to classifiers and measure words, although it must co-occur with the quantifier *dou* ‘all’ or *yi* ‘one’³⁷ which license classifier reduplication (Zhang 2013: 117-118, 120). Longer measure words (i.e. more than one syllable long, e.g. *kache* ‘truckload’) cannot be reduplicated.

In Northern Wu, classifier reduplication is also a strategy of marking a special type of plurality. Like in Mandarin, the reduplicated classifier must be semantically selected by the noun, and this type of plurality marking is less acceptable with a very small number of things (e.g. two or three objects). The sortal classifiers and measure words can generally be reduplicated in Northern Wu, except for disyllabic measure words and standard measures. However, the reduplicated classifiers in Northern Wu have a stronger distributive meaning than those in Mandarin, and they often co-occur with quantifiers (e.g. *dze* ‘all’) but are not licensed by them (Ruan 2009: 94; Qian 1997: 97). It is worth mentioning that Ningbo Wu classifiers appear to be stressed when they are reduplicated: the classifier for most nouns denoting humans is pronounced *ghoq* /ɦoʔ/ in most contexts (e.g. Num-Cl-N phrases, Dem-Cl-N phrases), but it must occur in the stressed form *kou* /kɔʊ/ when reduplicated (e.g. in the Cl-Cl-N phrase *kou-kou-banyeu* ‘every friend’). Noun reduplication is relatively rare in Northern Wu, but it is not restricted to monosyllabic nouns, and it also adds an abundant and distributive meaning to the noun (Zhu 2006: 58-59). In short, classifier reduplication is used for marking abundant plurality with a distributive meaning in Northern Wu.

Classifier reduplication is also found in Cantonese, although it does not seem to be a type of plural marking. In Cantonese, the meaning of reduplicated classifiers is consistently distributive (i.e. ‘every’) and they do not seem to create a singular-plural contrast (such as singular Num-Cl-N phrases vs plural Num-Cl-Cl-N phrases in Mandarin) (Matthews & Yip 2011). This use of Cantonese classifiers appears to be licensed by the quantifier *dou* ‘all’.

³⁷ According to Zhang (2013: 121), this *yi* is an existential quantifier rather than the numeral ‘one’.

Noun ellipsis is frequent in the Cl-Cl-N phrases, thus creating bare Cl-Cl phrases (e.g. *go-go-jan* ‘everyone’) in Cantonese (Matthews & Yip 2011: 39, 113, 305-306).

To sum up, plurality can be marked with suffixes or classifiers in Chinese. The suffix *-men* in Mandarin renders some nouns plural and definite. The suffix *-la* or *-laq* in Northern Wu can sometimes mark the plurality of certain nouns, but it is mainly used as a collective marker. Plurality is more frequently marked by the plural classifiers, which are found in all these Sinitic languages, namely *xie* in Mandarin, *ngae* in Northern Wu and *di* in Cantonese. In addition, classifiers can be reduplicated to mark plurality in Mandarin and Northern Wu, although the reduplicated classifiers in Northern Wu usually have a distributive meaning as well. Noun reduplication can occasionally be used as a plurality-marking strategy in these Chinese languages (Zhu 2006; Matthews & Yip 2011).

In the next subsection, I describe when Chinese classifiers are optional and propose some explanations for their optionality where possible.

5.1.3 Optionality and omission of Chinese classifiers

In the previous chapters, I depicted classifiers as obligatory morphemes that always occur in the Num-Cl-N phrases in Mandarin, Northern Wu and Cantonese. Although this generalization is mostly valid, it has two notable exceptions. First, Mandarin classifiers are not strictly obligatory in a Num-Cl-N phrase with a complex numeral and a noun for humans. Second, some idiomatic expressions and historical terms in the Chinese languages contain Num-N sequences without classifiers. I shall start with the first exception.

In Mandarin, individual classifiers can be exceptionally omitted in a Num-Cl-N phrase when the numeral is a complex numeral ending in a typical base (e.g. *bai* ‘hundred’, *qian* ‘thousand’) and the noun refers to human beings (e.g. *ren* ‘person’, *shibing* ‘soldier’). A quantified NP must meet both criteria for the classifier to be optional; otherwise, omission of the classifier would be ungrammatical. One such phrase with an optional classifier in Mandarin is shown in the following example.

(9) 岛上 驻扎着 三千 (名) 美军 士兵。 [Mandarin]

Dao-shang zhuzha-zhe san-qian (ming) mei-jun shibing
 island-on garrison-DUR three-thousand CL American-troops soldier
 'Three thousand American soldiers garrison the island.'

In (9), the classifier *ming* for the noun *shibing* 'soldier' is optional because it follows the numeral *san-qian* 'three thousand' and classifies the noun 'American soldier'. However, if the numeral were *san* 'three' instead of *san-qian* 'three thousand', then the classifier *ming* would have been obligatory. Likewise, the classifier *ge* in the NP *sanqian-ge-pingguo* 'three thousand apples' is also obligatory, since the head noun does not refer to humans.

Greenberg (2013 [1974]: 307-308) points out that "it is particularly common for classifiers not to occur with higher units of the numerical system and their multiples". This typological tendency implies that numerals can sometimes fill the position of the classifier in many classifier languages, thus rendering the classifier optional. It is plausible that these higher units can fill the position of Unit⁰ in certain contexts, if they are interpreted as units instead of part of the numeral. Another explanation is that the optionality of classifiers is dependent on phonological weight. It appears that the phonological unit the consists of the numeral and the classifier must be disyllabic or longer; since a complex numeral is already disyllabic (if not longer), the classifier can be optional. A problem with this account is that the complex numeral *san-qian-ling-yi* 'three thousand and one' is clearly less acceptable than *san-qian* 'three thousand', if the classifier is omitted in (9).

This type of optional classifiers is mainly found in formal and written Mandarin, and omission of classifiers in this type of NPs is relatively uncommon in casual speech. Overall, the classifier is more frequently optional as the base or unit gets numerically greater. For example, the classifier *ge* in the NP *sanshi-ge-ren* 'thirty people' (with the base *shi* 'ten') is more often retained than omitted even in a reasonably formal register. On the other hand, the same classifier *ge* in *bashi-yi-ge-ren* 'eight billion people' (with the base *yi* 'a hundred million') is almost always omitted, even in the most informal form of Mandarin. Optional classifiers of this kind are rare in colloquial Northern Wu, and this seems to be true of colloquial Cantonese as well, but they can obviously occur in some NPs borrowed from written Chinese, which is based on Mandarin (Matthews & Yip 2011).

The second type of contexts in which the classifiers are omitted includes idiomatic expressions and specialized terminology (esp. historical terms). In these contexts, it is in general infelicitous to insert a classifier in the phrase. An idiomatic expression in Mandarin with omitted classifiers is shown in the example below.

- (10) 九 牛 二 虎 之 力 [Mandarin]
Jiu niu er hu zhi li
Nine bull two tiger MOD strength
'a Herculean/tremendous effort'

The literal meaning of the idiom in (10) is 'the (physical) strength of nine bulls and two tigers', and it refers to a tremendous effort in a figurative way. Note that *jiu-niu* 'nine bulls' and *er-hu* 'two tigers' are Num-N phrases with no classifiers. Inserting the classifier *tou* for *niu* 'bull' or the classifier *zhi* for *hu* 'tiger' would make the idiom uninterpretable. The absence of classifiers in (10) and similar idiomatic expressions is arguably attributable to the influence of Old Chinese. Some idioms were created during the Old Chinese era, and others were created when Classical Chinese (which is modelled on Old Chinese) was the prestigious language in the region. Since classifiers are uncommon in Old Chinese (Yang-Drocourt 2004), they tend not to occur in the idiomatic expressions that were constructed based on Old Chinese grammar. The influence of Old Chinese is observable in the choice of lexical items as well: Old Chinese *hu* 'tiger' is used instead of Mandarin *laohu* 'tiger', and the Old Chinese modification marker *zhi* is used here while *de* is normally used in Mandarin.

Likewise, quantified NPs with no classifiers also occur in the context of specialized terminology, including many terms for Chinese history and world history. A historical term of this type is shown in example (11).

- (11) 五代十国 [Mandarin]
wu-dai-shi-guo
five-dynasty-ten-kingdom
'Five Dynasties and Ten Kingdoms' (an era in Chinese history: 907-979 CE)

The historical term in (11) does not contain any classifiers, and the generic classifier *ge*, which is normally used with these two nouns, cannot be inserted into the term. This and other specialized terms belong to the category of fixed expressions which also includes the idioms discussed above. Consequently, the reason why they contain no classifiers should be similar: the grammar of Classical Chinese also seems to influence how those terms were coined. Very similar specialized terms also exist in Northern Wu and Cantonese, and the absence of classifiers in this context is also observable in these two languages.

Chinese classifiers can also be optional in some other types of NPs (i.e. other than the Num-Cl-N phrases). In a Dem-Cl-N phrase in Mandarin, the classifier is optional when the head noun denotes human beings (Aikhenvald 2000: 208), although it is usually the generic classifier *ge* that gets omitted in those Dem-Cl-N phrases, and the specific classifiers for humans (e.g. the honorific classifier *wei*) is rarely omitted. The same phenomenon also occurs in Ningbo Wu, and the generic classifier *goq* can be omitted in that type of phrase as well. In both Standard Mandarin and Ningbo Wu, omission of the classifier tends not occur if the head noun or the core NP is long (i.e. with more than four or five syllables). Cantonese classifiers are obligatory when they co-occur with a demonstrative, but they are optional with some quantifiers, including 好多 *houdo* 'many' and 幾多 *geido* 'how many', especially when the head noun is an abstract noun (Matthews & Yip 2011: 107, 111). In Ningbo Wu, classifiers cannot be used with the quantifier 蠻多 *maetou* 'many', but with 几 *ci* 'how many' classifiers are required. In other words, each of these languages has specific rules about the use of classifiers with a given quantifier (i.e. obligatory, optional or prohibited).

To sum up, the classifier is optional in formal Mandarin NPs with a higher round numeral and a noun denoting human. Classifiers are omitted in many fixed expressions that were inherited from Classical Chinese. Classifiers are optional in some Dem-Cl-N phrases in Mandarin and Ningbo Wu, and they are optional with some quantifiers in Cantonese. Based on the patterns observed in these Chinese languages, there seem to be three major factors that affect the regular use of classifiers. First, classifiers tend to be optional when they are used with higher round numeral. Second, they may be optional when co-occurring with nouns referring to human beings. Third, the use of classifiers can be affected by borrowing,

which may create a stylistic effect in the use (or non-use) of classifiers. Whether and how these and similar factors play a role in other classifier languages will need to be determined by typological studies in the future.

The next section is about the implications of the current study.

5.2 Theoretical and typological implications

The discussion of Chinese classifiers and of other related issues in this thesis has several theoretical implications. First, one group of morphemes with superficially similar meanings and functions may correspond to two or more categories in the grammar and be associated with different syntactic structures. Second, numeral classifiers and number morphology can occur in different contexts in the same language, which can be explained by a tendency toward economy. Third, diachronic accounts can sometimes simultaneously explain the syntactic, morphological, phonological, pragmatic and other characteristics of some words and reveal the correlation between them. Lastly, this thesis invites a thorough rethink of syntactic variation in the Chinese languages and proposes that each variety should be described and analyzed as a separate system.

First, although classifiers and measure words can be used in similar contexts with similar meanings, they are represented differently in the mental grammar. The position of a classifier in the structure of NPs is determined by a parameter, hence the possibility of it merging with a numeral or with the head noun. On the other hand, so long as a numeral and a measure word co-occur in the same NP, one can expect a [Numeral-MW] constituent and consider it universal, and this is governed by a principle in the mental grammar. The cross-linguistic variation between classifiers-for-numerals and classifiers-for-nouns is thus expected (Little et al. 2022), which does not apply to measure words, and one should not be able to find measure-words-for-nouns in any of the world's languages.

Second, number morphology and numeral classifiers can co-occur in the same NPs in several languages (Bisang 2012: 24-25), and the cross-linguistic infrequency of NPs with both is best accounted for with economy rather than with a syntactic constraint. As heads of different projections (Y.-H. Li 1999), there is no compelling syntactic reason for number

and classifiers not to co-occur. However, since numeral classifiers are used with numerals, the exact number has been explicitly stated, and the number marking would be functionally redundant. In some languages with number morphology and numeral classifiers, their co-existence is not considered uneconomical, because they are used in different contexts. This is true of Mandarin: nouns with the plural suffix *-men* are definite whilst Num-Cl-N phrases are generally indefinite. Similar patterns should exist in other classifier languages, which may turn out to be more common than Ejagham-type languages in which plural markers and classifiers co-occur in the same phrases (Bisang 2012).

Third, some seemingly unrelated synchronic patterns may originate from the same diachronic change in the language and are thus correlated. For instance, Northern Wu classifiers are syntactic dependent on other morphemes in NPs (i.e. nouns in most cases), and they are phonetically weak, pragmatically versatile and incompatible with affixes. All these features are attributable to the grammaticalization of classifiers in the history of this language, and this combination of features is far from random. Comparable features are expected to be found in many other classifier languages, as grammaticalization is a cross-linguistically common change.

Lastly, there has been a popular belief that different varieties (“dialects”) of Chinese are only lexically and phonetically different. The assumption that they all have a common “Chinese grammar” is even accepted by some linguists (Matthews & Yip 2011: 1-2). For example, Chao (1968: 13) claims that “there is practically one universal Chinese grammar”. The current study disproves that belief. The syntactic behaviour of Mandarin classifiers is so different from that of their counterparts in Cantonese and Northern Wu that they should be considered categorially and structurally different. For instance, preverbal Cl-N phrases and possessor-Cl-N phrases are perfectly acceptable in Cantonese and Northern Wu, while they are ungrammatical in Mandarin. The syntactic differences among the Sinitic languages would be even greater if the patterns in Old Chinese and Middle Chinese were considered (contra Chao 1968: 13). In other words, like phonologists and lexicographers, syntacticians working on any of the Chinese languages should be aware of the existence of the “dialectal” differences and avoid vague terms such as “Chinese grammar”. This will make descriptions of varieties of Chinese more accurate, and the related discussion more pertinent.

In addition, this thesis has a few typological implications. First, it offers a novel explanation for the cross-linguistically attested constituent orders in quantified NPs with classifiers. Second, it demonstrates how the syntax of classifiers may interact with the lexical and phonological properties of numerals or classifiers. Third, this thesis makes predictions about the probable characteristics of the classifiers that are syntactically different from measure words in other classifier languages. Lastly, the syntax of classifiers and measure words, or at least some aspects of it, should be analyzed as emergent rather than strictly fixed, which could explain certain unexpected synchronic patterns.

First, previous typological studies suggest that two of the six logically possible constituent orders in Num-Cl-N are unattested in the world's languages, namely Cl-N-Num and Num-N-Cl, while all the other orders are attested (Aikhenvald 2000: 104-105; Her 2017A: 62-63). As the numeral and the classifier are adjacent to each other in all the other four orders (e.g. Num-Cl-N in Mandarin, N-Cl-Num in Bodo [Tibeto-Burman]), it has been argued that cross-linguistically the quantified NP with a numeral, a classifier and a noun has the left-branching structure [[Num-Cl]-N] (Aikhenvald 2000: 105; Her 2017A). My theory concurs with that view in the analysis of the classifiers that are grammaticalized to a limited degree. However, the highly grammaticalized classifiers (e.g. in Cantonese and Northern Wu) will merge with the head noun first. The existence of those languages (with highly grammaticalized classifiers) is not expected to change the word order typology of quantified NPs drastically for two reasons. First, the Num-Cl-N and N-Cl-Num orders are compatible with both the left-branching structure and the right-branching structure, and they are expected to stay the same even when the structure has become right-branching. Second, with the other two attested orders (viz. N-Num-Cl and Cl-Num-N), the classifier may have become very weak and phonetically dependent on the numeral as affixes when it becomes structurally closer to the noun via grammaticalization. In other words, a change in morpheme order would be prohibited for morphophonological reasons. This explains why classifier languages with the N-Num-Cl order are "much more frequent" than those with the Cl-Num-N order: if classifiers are indeed affixes in some of these languages, suffixes are typologically more common than prefixes (Aikhenvald 2000: 105). It is worth noting that the "unattested" Cl-N-Num order is in fact found in some Tai-Kadai languages (e.g. Bouyei) with highly grammaticalized classifiers, even though it is only acceptable with the numeral

‘one’ which could be analyzed as a determiner (Her 2017A: 46-48; Jiang 2009: 42). If my theory is correct, one might find some languages with highly grammaticalized classifiers and an indisputable Cl-N-Num order, which could be rare (if the classifier is phonetically weak and prefix-like) yet possible.

Second, some lexical or phonological characteristics of numerals or classifiers may afford a clue as to how the classifier merges with the other morphemes. For instance, the tone sandhi pattern of trisyllabic Num-Cl-N phrases is the same as that of the [N-N]-Adj¹ phrases and different from that of the Adj-[N-N] phrases in Mandarin. In addition, some numerals may have special forms when they co-occur with classifiers (e.g. Mandarin *liang* and *er* ‘two’). Such patterns should not be omitted from descriptions of classifier languages, as they reveal the status of classifiers at syntax-phonology and syntax-lexicon interfaces. Whether the same patterns occur in classifiers-for-numerals languages and classifiers-for-nouns languages is also worth examining once the typological data is available.

Third, classifiers are expected to behave differently from measure words at two stages in their life cycle. First, in languages with nascent classifier systems, classifiers can be optional and infrequent, whilst measure words tend to be obligatory and more frequent. Second, in languages with highly grammaticalized classifiers, classifiers may occur in many types of NPs, while measure words can be used in fewer (and possibly somewhat different) contexts. In other words, if the classifiers and measure words appear in different syntactic structures in a specific language, and the classifiers are obligatory and frequent, then they are expected to have some of the following characteristics. First, the classifiers should have lost most of their lexical meanings and are no longer used as nouns (or any other lexical words). Second, the classifiers should be phonetically weak, and the most frequent ones may display signs of lenition and vowel reduction. Third, the classifiers will likely be pragmatically versatile, occurring in a relatively large number of contexts. Lastly, they will likely be semantically rich, marking multiple features at once (e.g. number, definiteness, nominalization). These predictions could be verified in future typological studies on other classifier languages.

¹ [N-N] represents a compound noun that consists of two nominal morphemes.

Last but not least, the behaviour of classifiers and measure words described in this thesis lends some support to Hopper's theory of "Emergent Grammar" (Hopper 1998: 155-156). According to that theory, "structure ... is shaped by discourse in an ongoing process", and synchronic grammar is a result of "sedimentation of frequently used forms" (ibid. 156, 158). This explains why Mandarin classifiers and measure words do not always share the same behaviour (e.g. different meanings in Num-Adj-Cl/MW-N phrases) and why Northern Wu and Cantonese classifiers and measure words bear certain unexpected similarities (e.g. ellipsis in Num-Cl/MW-N phrases). In view of the grammaticalization of classifiers in the history of Chinese, the divergence between classifiers and measure words (due to the grammaticalization of the former) can be perceived as an ongoing process, which has barely begun in Mandarin and has not finished in Cantonese and Northern Wu.

Section 5.3 is the conclusion of this chapter.

5.3 Conclusion

In this chapter, I briefly examined three issues related to the classifiers in the three Chinese languages under review and discussed some major implications of the current thesis. I proposed that all nouns in these three languages are mass, and that the mass-count distinction does not correlate directly with the difference between classifiers and measure words. Plurality is most often marked using special classifier forms or reduplicated classifiers, with the plural suffix *-men* in Mandarin being the lone exception. The classifiers can only be optional in formal Mandarin with higher numerals and with nouns for human, and they do not occur in some idioms and specialized terms; otherwise, they are required in the three languages. In terms of implications, functionally similar words can be represented differently in the mental grammar, which can lead to structural differences. The uncommonness of numeral classifiers co-occurring with number morphology is attributable to economy. The absence of certain constituent orders in NPs with numerals and classifiers does not exclude the possibility that such NPs in some languages have the right-branching structure. If some classifiers are grammaticalized to the point that they differ syntactically from measure words, then they will likely have other properties of

highly grammaticalized morphemes, such as phonetic reduction, incompatibility with affixes and pragmatic versatility.

6. Conclusion

In this thesis, I investigated several questions related to the syntactic properties and history of Chinese classifiers. In this chapter, I summarize the major findings and mention some related puzzles that remain to be studied in the future.

In Chapter 2, I argued that Mandarin classifiers and measure words belong to one and the same grammatical category based on their shared distribution and behaviour in sentences with extraction, ellipsis, and coordination. I argued that classifiers and measure words occur in a left-branching structure and that they always merge with a numeral first. The existence of bare Cl/MW-N phrases and Dem-Cl/MW-N phrases in Mandarin can be explained by the presence of a silent numeral in those NPs (Zhang 2019). I linked that left-branching syntactic structure to the tone sandhi pattern in Num-Cl/MW-N phrases and to the lexical variants of the numerals *yi* ‘one’ and *er* ‘two’.

In Chapter 3, I proposed that the classifiers and measure words in Northern Wu and Cantonese belong to two separate grammatical categories. This claim was supported by the different distribution and interpretation of classifiers and measure words: only classifiers can occur in pre-verbal bare Cl-N phrases and modified Cl-N phrases (which are definite in meaning), and only measure words are acceptable in Num-Adj-MW-N phrases (which are indefinite). I argued that classifiers merge with head nouns in Northern Wu and Cantonese, and that Northern Wu and Cantonese measure words first merge with numerals. These claims are based on the use of classifiers and measure words in NPs without numerals and the grammaticality of [Cl-N-and-Cl-N] phrases in these two languages.

In Chapter 4, I examined the history of classifiers and measure words in Chinese. I showed that a fully-fledged classifier system first appeared in Early Middle Chinese, which gradually became more complex and more versatile and evolved into the classifier systems in the modern Chinese languages. I argued that many of the changes affecting the classifiers could be subsumed grammaticalization, which include phonetic reduction, morphological fossilization in addition to the syntactic changes (e.g. occurring in more types of NPs).

In Chapter 5, I discussed the theoretical implications of this study and some related issues. I proposed that all nouns in Mandarin, Northern Wu and Cantonese are underlyingly

mass nouns, and that classifiers individualize and classify nouns. I showed that plurality is usually marked with classifiers in the three Chinese languages, and I argued that numeral classifiers tend not to co-occur with number morphology due to preference for economical systems. Furthermore, I proposed a new explanation for the attested constituent orders in Num-Cl-N phrases and I suggested that the concept of Emergent Grammar could be helpful in explaining some properties of classifiers.

At least three questions related to my theories in this thesis could be investigated by researchers in the future. First, syntacticians may find other classifier languages in which classifiers and measure words occur in different syntactic structures, thereby determining how common those languages are. Second, typologists could determine whether the right-branching structure of [Num-[Cl-N]] phrases must co-occur with highly grammaticalized classifiers. In addition, since there is variation in the acceptability of modified CL/MW-N phrases among native speakers of Cantonese and Northern Wu, researchers could study the variation more thoroughly and see what syntactic, semantic or sociolinguistic factors influence the acceptability of those phrases.

Bibliography

- Academia Sinica. *Academia Sinica Tagged Corpus of Old Chinese*. (Abbr. ASTC-OC).
<http://lingcorpus.iis.sinica.edu.tw/ancient/>
- Academia Sinica. *Academia Sinica Tagged Corpus of Middle Chinese*. (Abbr. ASTC-MC).
<https://lingcorpus.iis.sinica.edu.tw/middle/>
- Academia Sinica. *Academia Sinica Tagged Corpus of Early Mandarin Chinese*. (Abbr. ASTC-EMC). <https://lingcorpus.iis.sinica.edu.tw/early/>
- Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Bale, A. and Coon, J. (2014). Classifiers Are for Numerals, Not for Nouns: Consequences for the Mass/Count Distinction. *Linguistic Inquiry* 45(4): 695-707.
- Bale, A., Coon, J. and López, N. A. (2019). Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa: a journal of general linguistics* 4(1): 77. 1–30.
- Bernot, D., Cramerotti, C. & Yin Myint, M.-Y. (1997). *Dictionnaire Français-Birman* (French-Burmese Dictionary). Paris: Langues et Mondes / L'Asiathèque.
- Bisang, W. (2011). Chapter 9: Grammaticalization and linguistic typology. *The Oxford Handbook of Grammaticalization*; Heine, B. & Narrog, H. (eds.). Oxford: Oxford University Press.
- Bisang, W. (2012). Numeral classifiers with plural marking: A challenge to Greenberg. *Plurality and Classifiers across Languages in China*, (ed.) Xu, Dan. Berlin: De Gruyter Mouton.
- Borer, H. (2005). *Structuring Sense Volume I: In Name Only*. Oxford: Oxford University Press.
- Campbell, L. (2013). *Historical Linguistics: An Introduction* (3rd edition). Cambridge: The MIT Press.
- Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chen, Fu (ed.). (2024). *A Dictionary of Chinese Classifiers with Detailed Explanations and Etymological Annotations* (汉语量词详解与溯源词典). Beijing: Language and Culture Press.
- Chen, G.-Z. (2003). *Yunjing with comprehensive explanations* (韻鏡通釋). Shanghai: Shanghai Lexicographical Publishing House.

- Cheng, L. L.-S. & Sybesma, R. (1999). Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry* 30(4): 509-542.
- Chierchia, G. (1998). Reference to Kinds across Languages. *Natural Language Semantics* 6(4): 339-405.
- Fábregas, A. & Scalise, S. (2012). *Morphology: From Data to Theories*. Edinburgh: Edinburgh University Press.
- Greenberg, J. H. (2013) [1974]. Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type (Chapter 10). *Eight Decades of General Linguistics: The History of CIPL and Its Role in the History of Linguistics*, eds. F. Kiefer and P. van Sterkenberg. Leiden: Brill.
- He, Jie. (2008). *A Study on Modern Chinese Classifiers (Enlarged edition)* (现代汉语量词研究). Beijing: Beijing Language and Culture University Press.
- Her, O.-S. (2017[A]). Structure of numerals and classifiers in Chinese: Historical and typological perspectives and cross-linguistic implications. *Language and Linguistics* 18(1): 26-71.
- Her, O.-S. (2017[B]). Deriving classifier word order typology, or Greenberg's Universal 20A and Universal 20. *Linguistics* 55(2): 265-303.
- Hopper, P. J. (1998). Emergent Grammar. *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure (Volume I)*: 155-175; (eds.) M. Tomasello. New York: Routledge.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd edition). Cambridge: Cambridge University Press.
- Huang, C.-T. James & Li, Y.-H. Audrey, and Li, Yafei. (2009). *The Syntax of Chinese*. Cambridge: Cambridge University Press.
- Huang, C.-R., Wang, H. & Chen, I.-H. (2022). Characters as Basic Lexical Units and Monosyllabicity in Chinese (Chapter 4). *The Cambridge Handbook of Chinese Linguistics* (eds. Chu-Ren Huang, Yen-Hwei Lin, I-Hsuan Chen and Yu-Yin Hsu): 74-96. Cambridge: Cambridge University Press.

- Huang, Y.-H. (2005). Korean classifiers (朝鲜语量词). *Studies on the classifiers in Sino-Tibetan languages* (汉藏语系量词研究), Li, Jin-fang (ed.); pp.475-487. Beijing: China Minzu University Press.
- Jenny, M. and Hnin Tun, S. S. (2016). *Burmese: A Comprehensive Grammar*. Abingdon (UK): Routledge.
- Jiang, Ying. (2009). *A comparative study on the classifiers for nouns in the Sino-Tibetan languages* (汉藏语系语言名量词比较研究). Beijing: The Ethnic Publishing House.
- Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Li, J.-P. (2017). *A Study on the Classifiers in Pre-Qin and Han-Dynasty Chinese* (先秦兩漢量詞研究). Beijing: China Social Sciences Press.
- Li, X.-P. (2013). *Numeral Classifiers in Chinese: The Syntax-Semantics Interface*. Berlin: De Gruyter Mouton.
- Li, X.-P. & Bisang, W. (2012). Classifiers in Sinitic Languages: From individuation to definiteness-marking. *Lingua* 122: 335-355.
- Li, Yen-hui A. (1999). Plurality in a Classifier Language. *Journal of East Asian Linguistics* 8(1): 75-99.
- Little, C. R., Moroney, M. and Royer, J. (2022). Classifiers can be for numerals or nouns: Two strategies for numeral modification. *Glossa: a journal of general linguistics* 7(1): 1-35.
- Luo, Zi-qun. (2005). A Dialectal comparison of the semantic ranges of the individual classifier *zhī* in Chinese (汉语个体量词“只”语义范围的方言比较). *Studies on the classifiers in Sino-Tibetan languages* (汉藏语系量词研究), Li, Jin-fang (ed.); pp.358-372. Beijing: China Minzu University Press.
- Luo, Q.-P., Hsieh, M.-L., and Shi, D.-X. (2017). Pre-classifier adjectival modification in Mandarin Chinese: a measurement-based analysis. *Journal of East Asian Linguistics* 26(1): 1-36.

- Mak, Wan, and Taam, Bou-wan.¹ (2011). *Practical Sorting Dictionary of Guangzhou Cantonese* (實用廣州話分類詞典). Hong Kong: The Commercial Press.
- Matthews, S., and Yip, V. (2011). *Cantonese: A Comprehensive Grammar* (2nd edition). New York: Routledge.
- Myers, J. (2022). Wordhood and Disyllabicity in Chinese (Chapter 3). *The Cambridge Handbook of Chinese Linguistics* (eds. Chu-Ren Huang, Yen-Hwei Lin, I-Hsuan Chen and Yu-Yin Hsu): 47-73. Cambridge: Cambridge University Press.
- Qian, N.-R. (1992). *A study on contemporary Wu language* (当代吴语研究). Shanghai: Shanghai Educational Publishing House.
- Qian, N.-R. (1997). *A Grammar of Shanghai Dialect* (上海话语法). Shanghai: Shanghai People's Press.
- Qian, N.-R. & Xu, B.-H., and Tang, Z.-Z. (2007). *A Comprehensive Dictionary of Shanghainese* (上海话大词典). Shanghai: Shanghai Century Publishing Group and Shanghai Lexicographical Publishing House.
- Ruan, G.-J. (2009). *A study of the grammar of Ningbo dialect* (宁波方言语法研究). Wuhan: Central China Normal University Publishing House.
- Saito, M., Lin, T.-H. J. and Murasugi, K. (2008). N'-Ellipsis and the Structure of Noun Phrases in Chinese and Japanese. *Journal of East Asian Linguistics* 17(3): 247-271.
- Simpson, A. (2005). Classifiers and DP structure in Southeast Asia. *The Oxford Handbook of Comparative Syntax*: 806-838; Guglielmo Cinque and Richard S. Kayne (eds.). Oxford: Oxford University Press.
- Wang, Li. (2023). *Modern Chinese Grammar III: Substitution and Numeration*. Translated by B. Wang & Y.-Y. Ma. Abingdon: Routledge.
- Wang, Shao-Xin. (2018). *A Study of Chinese Classifiers during the Sui-Tang and Five Dynasties Period* (隋唐五代量词研究). Beijing: The Commercial Press.

¹ The authors' names in Chinese are 麥耘 and 譚步雲, and in Pinyin, these would be: Mai, Yun and Tan, Bu-Yun. The Cantonese words in this dictionary are grouped by category.

- Wang, Wei. (1987). The Bouyei language (布依语). *The Languages of the Minority Ethnic Groups in China* (中国少数民族语言): 287-301. The Institute for Minority Languages at the Minzu University of China (ed.). Chengdu: Sichuan Minzu Publishing House.
- Wee, L.-H. (2022). Tonal Processes Conditioned by Morphosyntax (Chapter 15). *The Cambridge Handbook of Chinese Linguistics* (eds. Chu-Ren Huang, Yen-Hwei Lin, I-Hsuan Chen and Yu-Yin Hsu): 313-335. Cambridge: Cambridge University Press.
- Wu, An-qi. (2005). The classifiers for nouns in analytic languages (分析型语言的名量词). *Studies on the classifiers in Sino-Tibetan languages* (汉藏语系量词研究), Li, Jin-fang (ed.); pp.1-13. Beijing: China Minzu University Press.
- Wu, Ying. (2022). *A Study on the Shape Classifiers in Chinese from a Typological Perspective* (类型学视野下汉语形状量词研究). Wuhan: Wuhan University Press.
- Xia, H.-L. (2021). *Classifiers in three Sinitic languages: structure, typology and variation*. [Unpublished manuscript].
- Xu, D. (2012). Reduplication in languages: A case study of languages of China. *Plurality and Classifiers across Languages in China*: 43-63. Dan Xu (ed.). Berlin: De Gruyter Mouton.
- Yang-Drocourt, Zhitang. (2004). *Évolution syntaxique du classificateur en chinois du XIII^e siècle av. J.-C. au XVII^e siècle* (Syntactic evolution of the classifiers in Chinese from the 13th century BCE to the 17th century). Paris: Centre de Recherches Linguistiques sur l'Asie Orientale (École des Hautes Études en Sciences Sociales).
- Ye, Gui-Chen. (2008). *A Study on the Classifiers in Ming-Dynasty Chinese* (明代汉语量词研究). Changsha: Yuelu Publishing House.
- Yuko, Fujimura. (2017). Reduplication in Standard Malay and Japanese. *Journal of Modern Languages* 13(1): 65-92.
- Zhang, Cheng. (2012). *A History of the Evolution of Chinese Classifiers for Nouns from a Typological Perspective* (类型学视野的汉语名量词演变史). Beijing: Peking University Press.

Zhang, Hui-Qiang. (2014). *A comparative study of the classifiers in Bianwen from Dunhuang and in Zaju of the Yuan Dynasty* (敦煌变文、元刊杂剧量词比较研究). Beijing: Guangming Daily Publishing House.

Zhang, Niina Ning. (2009). Explaining the immobility of conjuncts. *Studia Linguistica* 64(2): 190-238.

Zhang, Niina Ning. (2013). *Classifier Structures in Mandarin Chinese*. Berlin: De Gruyter Mouton.

Zhang, Niina Ning. (2019). Complex indefinites and the projection of DP in Mandarin Chinese. *Journal of East Asian Linguistics* 28(2): 179-210.

Zheng, S.-L. (2016). *A Study on the Classifiers for Nouns in Chinese Inscriptions from the Wei-Jin Southern-and-Northern-Dynasties Period* (魏晉南北朝石刻名量詞研究). Shanghai:

Shanghai Century Publishing Group and Shanghai Ancient Books Publishing House.

Zhu, Xiaonong. (2006). *A Grammar of Shanghai Wu*. Munich: LINCOM-Europa.

Appendix: The Cantonese and Northern Wu questionnaires used for data collection

1. Survey on Cantonese classifiers and measure words

I. Sentences with pre-verbal bare CL-N phrases:

個細路 喺度 做乜嘢呀？

Go sailou haidou zou matje aa³

CL child PROG do what SFP

‘What is the child doing?’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?²

間屋好靚呀！

Gaan uk hou leng aa³

CL house very beautiful SFP

‘The house is so beautiful!’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

II. Sentences with post-verbal bare CL-N phrases:

呢件事好複雜，搵個律師問吓先啦。

² A reader mentions that this type of questions related to translation can be hard to interpret. A different type of questions (e.g. about truth values) may need to be included in lieu of these in a follow-up study.

Ni kin si hou fukzaap wan go leotsi man-ha sin laa¹

This CL matter very complex find CL lawyer ask-ASP first SFP

'This matter is very complex, let's ask a lawyer first.'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

你問咗個律師未呀？

Nei man-zo go leotsi mei aa³

You ask-ASP CL lawyer NEG.PERF SFP

'Have you asked the lawyer?'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

III. Sentences with modified CL-N phrases (type 1: possessor-CL-N):

王生架車整好咗未呀？

Wong-saang ga ce zing-hou-zo mei aa³

Wong-sir CL car fix-ASP-ASP NEG.PERF SFP

'Has Mr. Wong's car been repaired?'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

佢把聲都幾好聽㗎。

Keoi ba seng dou kei houteng gaa³

He CL voice also fairly pleasant.to.hear SFP

‘His voice is also pleasant to hear.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

IV. Sentences with modified CL-N phrases (type 2: adjective-CL-N):

咁 大 間 房 幾多 個 人 住 呀?

Gam daai gaan fong geido go jan zyu aa³

so big CL suite how.many CL person live SFP

‘How many people live in such a big suite?’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

V. Sentences with modified CL-N phrases (type 3: demonstrative-(NUM)-CL-N)

我 仲未 讀過 嗰 本 書。

Ngo zungmei duk-gwo go bun syu

1st.SING NEG read-EXP that CL book

‘I haven’t read that book.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

我 仲未 讀過 嗰 一 本 書。

Ngo zungmei duk-gwo go jat bun syu

1st.SING NEG read-EXP that one CL book

'I haven't read that book.'

❖ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

VI. Sentences with pre-verbal bare MW-N phrases:

杯 茶 幾 香 嘅

Bui caa gei hoeng ge²

MW=_{cup} tea quite fragrant SFP

'That cup of tea smells quite good.'

❖ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

包 麵粉 好 貴 嘅 咩

Baau minfan hou gwai ge² me¹

MW=_{bag} flour very expensive SFP Q

'Is that bag of flour so expensive?'

❖ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

班 後生仔 走咗 啦

Baan hausangzai zau-zo laa¹

Group youngster leave-ASP SFP

‘That group of young people has left already.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

VII. Sentences with post-verbal bare MW-N phrases:

買 抽 提子 啦，好 新鮮 㗎

Maai cau taizi laa hou sansin gaa³

Buy MW=_{bunch} grape SFP very fresh SFP

‘(Please) buy a bunch of grapes, they are very fresh.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

食 完 碗 飯 未 呀？

Sik jyun wun faan mei aa¹

Eat finish MW=_{bowl} rice NEG.PERF SFP

‘Have you finished that bowl of rice?’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

VIII. Sentences with modified MW-N phrases (type 1: possessor-MW-N):

你 樽 汽水 係 乜嘢 味 㗎？

Nei zeon heiseoi hai matje mei gaa³

You MW=_{bottle} soft.drink COP what flavour SFP

‘What is the flavour of your drink (in the bottle)?’

❶ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

IX. Sentences with modified MW-N phrases (type 2: adjective-MW-N):

咁 大 舊 石 我 搬 唔 啱 㗎。

Gam daai gau sek ngo bun m juk wo³

So big MW=_{lump} rock 1st.SING move NEG move SFP

‘I cannot move such a big lump of rock.’

❶ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

佢 買咗 咁 大 堆 生果 畀 你 食。

Keoi maai-zo gam daai deoi saanggwo bei nei sik

3rd.SING buy-ASP such big MW=_{pile} fruit give you eat

‘He bought such a big pile of fruit for you to eat.’

❶ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

X. Sentences with modified MW-N phrases (type 3: demonstrative-(NUM)-MW-N)

佢 好 鍾意 呢 種 花 嘅。

Keoi hou zungji ni zung fa ge

3rd.SING very like this type flower SFP

‘She likes this kind of flower very much.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

佢 好 鍾意 呢 一 種 花 嘅。

Keoi hou zungji ni jat zung fa ge

3rd.SING very like this one type flower SFP

‘She likes this kind of flower very much.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

XI. Sentences with bare *di*-N phrases (*di* is the plural classifier and a measure word):

啲 細路 好 得意 呀。

Di sailou hou dakji aa³

CL.PL child very cute SFP

‘The children are very cute.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

啲 血 好 得人驚 呀。

Di hyut hou dakjangeng aa³

MW blood very frightening SFP

‘The blood (there) is very scary.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

古 希臘 人 相信 嗰 啲 神明。

Gu heilaap jan soengseon go di sanming

Ancient Greek person believe.in that CL.PL god

‘Ancient Greeks believed in those gods.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

XII. More complex constructions:

佢 寫 嘅 部 小說 好 有意思。

Keoi se ge bou siusyut hou jaujisi

3rd.SING write LP CL novel very interesting

‘The novel that he wrote is very interesting.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

佢 寫 嗰 部 小說 好 有意思。

Keoi se go bou siusyut hou jaujisi

3rd.SING write that CL novel very interesting

‘The novel that he wrote is very interesting.’

❶ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

我 飲 嘅 樽 汽水 太 甜 啦。

Ngo jam ge zeon heiseoi taai tim laa

1st.SING drink LP MW=_{bottle} soft.drink too sweet SFP

‘That bottle of soft drink that I drank was too sweet.’

❶ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

我 飲 嗰 樽 汽水 太 甜 啦。

Ngo jam go zeon heiseoi taai tim laa

1st.SING drink that MW=_{bottle} soft.drink too sweet SFP

‘That bottle of soft drink that I drank was too sweet.’

❶ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

❷ Is the English translation accurate? Yes / No.

❸ If it is inaccurate, how would you change the translation?

我 買咗 兩 本 雜誌， 佢 買咗 三 本。

Ngo maai-zo loeng bun zaapzi keoi maai-zo saam bun

1st.SING buy-ASP two CL magazine 3rd.SING buy-ASP three CL

'I bought two magazines, and he bought three.'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

我 買咗 兩 本 雜誌， 佢 買咗 三。

Ngo maai-zo loeng bun zaapzi keoi maai-zo saam

1st.SING buy-ASP two CL magazine 3rd.SING buy-ASP three

'I bought two magazines, and he bought three.'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

我 飲咗 兩 罐 啤酒，佢 飲咗 三 罐。

Ngo jam-zo loeng gun bezau keoi jam-zo saam gun

1st.SING drink-ASP two MW=_{can} beer 3rd.SING drink-ASP three MW=_{can}

'I drank two cans of beer, and he drank three.'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

我 飲咗 兩 罐 啤酒，佢 飲咗 三。

Ngo jam-zo loeng gun bezau keoi jam-zo saam

1st.SING drink-ASP two MW=_{can} beer 3rd.SING drink-ASP three

'I drank two cans of beer, and he drank three.'

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

王生 同 李生 分別 想 賣 兩 間 同 三 間 屋。

Wong-saang tung Lei-saang fanbit soeng maai loeng gaan tung saam gaan uk

Wong-sir and Lei-sir respectively want sell two CL and three CL house

‘Mr. Wong and Mr. Lei want to sell two and three houses respectively.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

隻 狗 同 隻 貓 喺度 打交。

Zek gau tung zek maau haidou daagaau

CL dog and CL cat PROG fight

‘The dog and the cat are fighting with each other.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

碗 炒飯 同 碟 牛腩 都 好 好味 呀！

Wun caaafaan tung dip ngaunaam dou hou houmei aa

MW_{=bowl} fried.rice and MW_{=saucer} beef.brisket all very tasty SFP

‘This bowl of fried rice and this dish of beef brisket are both very tasty.’

① How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

聽日 要 我 人 個 兩 個 見。

Tingjat yiu ngo jan go loeng go gin

Tomorrow must I person CL two that see

(*) 'Tomorrow, I need to see those two people.'

Φ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

我 飲咗 一 大 碗 湯 啦。

Ngo jam-zo jat daai wun tong laa

1st.SG drink-PERF one big MW_{-bowl} soup SFP

'I have eaten a big bowl of soup.'

Φ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

我 食咗 一 大 隻 雞髀 啦。

Ngo sik-zo jat daai zek gaibei laa

1st.SG eat-PERF one big CL chicken.thigh SFP

'I have eaten a big chicken thigh.'

Φ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

② Is the English translation accurate? Yes / No.

③ If it is inaccurate, how would you change the translation?

我 買咗 支 筆 同 本 雜誌。

Ngo maai-zo zi bat tung bun zaapzi

1st.SG buy-PERF CL pen and CL magazine

Φ How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

2 If you find it acceptable, select all the meanings that this sentence has:

- A. 'I bought a pen and a magazine.'
- B. 'I bought a pen and the magazine.'
- C. 'I bought the pen and a magazine.'
- D. 'I bought the pen and the magazine.'

佢 飲咗 杯 茶 同 樽 汽水。

Keoi jam-zo bui caa tung zeon heiseoi

3rd.SING drink-ASP MW=_{cup} tea and MW=_{bottle} soda

1 How good does this sentence sound?

Excellent – Good – Poor – Unacceptable

2 If you find it acceptable, select all the meanings that this sentence has:

- A. 'He drank a cup of tea and a bottle of soft drink.'
- B. 'He drank a cup of tea and that bottle of soft drink.'
- C. 'He drank that cup of tea and a bottle of soft drink.'

'He drank that cup of tea and that bottle of soft drink.'

[End of the questionnaire]

To see the online survey and results, please use the link below:

<https://www.surveymonkey.com/r/B9MNX8K>

2. Survey on Ningbo Wu classifiers and measure words

1. Sentences with pre-verbal bare CL-N phrases:

梗 鱼 死掉 嘞

kan ng shi-dio le

CL fish die-ASP SFP

'The fish is dead.'

(Only the first and the last lines will be shown during the test.)

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

2. Sentences with post-verbal bare CL-N phrases:

我 昨末 买了 梗 鱼

Ngo zoqmaq ma-leq kan ng

I yesterday buy-ASP CL fish

'I bought a fish yesterday.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

3. Sentences with possessor-CL-N phrases:

阿姐 部 汽车 是 白 颜色 喔

Acia bu chitsho zy baq ngaeseq ghо

Elder.sister CL car COP white colour SFP

'My elder sister's car is white.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

4. Sentences with adjective-CL-N phrases:

黄黄 部 汽车 是 依个 伐?

Waoawaon bu chitsho zy nou-(ghoq) va

Yellow CL car COP you.POSS Q

'Is the yellow car yours?'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

5. Sentences with Num-adjective-CL-N phrases:

小林 天亮饭 吃了 一大只 淡包。

Shiolin thinianvae chioq-leq iq-dou-tsaq daepau

Shiolin breakfast eat-ASP one-big-CL steamed.bun

'Xiao Lin had a big steamed bun for breakfast.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

6. Sentences with pre-verbal bare MW-N phrases:

杯 茶 冷掉 嘞

pei dzo lan-dio le

MW tea cold-ASP SFP

Intended: 'That cup of tea is cold.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

7. Sentences with post-verbal bare MW-N phrases:

其有事体，吃了 杯 茶 就去嘞

Ji yeu zythi, chioq-leq pei dzo jiu chi le

3rd.SING have business eat-ASP MW tea then go.away SFP

'He had other things to do, so he had a cup of tea and went away.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

8. Sentences with possessor-MW-N phrases:

依 瓶 饮料 是 啥 味道 啦?

Nou bin inlio zy so midau la

You MW soft.drink COP what flavour SFP

Intended: 'What is the flavour of your drink (in the bottle)?'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

9. Sentences with adjective-MW-N phrases:

黄黄 瓶 饮料 是 啥 味道 ?

Waonwaon bin inlio zy so midau

yellow MW soft.drink COP what flavour

Intended: 'What is the flavour of the yellow drink in the bottle?'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

Additional questions: Ambiguity? Which one must be yellow? The bottle or the drink?

10. Sentences with Num-adjective-MW-N phrases:

小林 天亮饭 吃了 一大碗 汤饭。

Shiolin thinianvae chioq-leq iq-dou-wun thaonvae

Shiolin breakfast eat-ASP one-big-MW congee

'Xiao Lin had a big bowl of congee for breakfast.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

11. Plural NPs with the paucal marker 两 *lian*:

两件衬衫还没燥啦。

Lian ji tshensae wae maq sau la

Two CL shirt PARTICLE NEG dry SFP

'The shirts are not dry yet.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

两块蛋糕依吃掉好嘞。

Lian kuei daekau nou chioq-dio hau le

Two piece cake you eat-ASP all.right SFP

'You can eat the two pieces of cake.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

12. Relative clauses:

我斩肉骨头把刀寻弗着嘞。

Ngo tsae nioqkuaqdeu bo tau zhin-vaq-dzaq le

1st.SING cut meat.bone CL knife find-NEG-RES SFP

'The knife that I use to chop bones is missing.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

我泡茶瓶热水来阿里啦？

Ngo phau-dzo bin nyiq-syu le aqli la?

1st.SING steep-tea MW_{-bottle} hot-water be where SFP

'Where is the bottle of hot water that I'm going to use to make tea?'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

13. Complex construction: coordination

只狗搭只猫来盖打相打。

Tsaq mae taq tsaq keu leke tanshiantan

CL cat and CL dog PROG.DIST fight.against.each.other

'The cat and the dog are fighting against each other.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

我昨未买了本书搭支铅笔。

Ngo zoqmaq ma-laq pen syu taq tsy khaepiq

1st.SING yesterday buy-ASP CL book and CL pencil

'I bought a book and a pencil yesterday.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

块蛋糕搭瓶饮料齐忒甜嘞。

Kuei daekau taq bin inlio dze thaq die le

Piece cake and bottle soft.drink both too sweet SFP

'Both the piece of cake and the bottle of drink are too sweet.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

阿拉买了袋橘子搭箱苹果

Aq-laq ma-leq de ciuiqtsy taq shian binkou

We buy-ASP bag orange and box apple

'We bought a bag of orange and a box of apples.'

This sentence is: perfect / acceptable / bizarre / completely unacceptable.

The English translation is accurate / inaccurate.

[End of the questionnaire]

3. Survey on Shanghai Wu classifiers and measure words

1. 吾 昨日 买了 条 鱼。

Ngu zoqniq ma-leq diau ng.

1.SING yesterday buy-ASP CL fish

'I bought a fish yesterday.'

❖ Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

❷ Is the English translation of this sentence accurate? Yes / No.

❸ If the translation is inaccurate, how would you modify it?

❹ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

2. 只 狗 辣 沙发 高头 睡觉。

Tsaq keu laq sofaq kaudeu khuenkau.

CL dog be.LOC sofa on.top.of sleep

'The dog is sleeping on the sofa.'

❖ Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

❷ Is the English translation of this sentence accurate? Yes / No.

❸ If the translation is inaccurate, how would you modify it?

❹ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

3. 爷叔 部 车子 比 搿 部 车子 大。

Yasoq bu tshotsy pi geq bu tshotsy du.

Uncle CL car COMP this CL car big

'Uncle's car is bigger than this one.'

❖ Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

❷ Is the English translation of this sentence accurate? Yes / No.

❸ If the translation is inaccurate, how would you modify it?

❹ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

4. 白 颜色 部 车子 阿 是 依 个?

Baq ngeseq bu tshotsy aq zy non geq

White colour CL car Q COP 2.SING LP

'Is the white car yours?'

❶ Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

❷ Is the English translation of this sentence accurate? Yes / No.

❸ If the translation is inaccurate, how would you modify it?

❹ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

5. 吾 斩 肉骨头 把 刀 勿见脱 了。

Ngu tse nioq-kueqdeu po tau veq-ci-theq leq

1.SING chop meat-bone CL knife NEG-see-ASP SFP

'The knife that I use to chop bones is missing.'

❶ Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

❷ Is the English translation of this sentence accurate? Yes / No.

❸ If the translation is inaccurate, how would you modify it?

❹ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

6. 每 个 学生 伙 买了 本 书 得 支 钢笔。

Me geq ghosqan ze ma-leq pen sy teq tsy kaonpiq

Each CL student QUANT buy-ASP CL book and CL pen

'Each student bought a book and a pen.'

❶ Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

❷ Is the English translation of this sentence accurate? Yes / No.

❸ If the translation is inaccurate, how would you modify it?

❹ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

7. 只 狗 得 只 猫 辣辣 打相打。

Tsaq keu teq tsaq mau laqlaq tansiantan

CL dog and CL cat PROG fight.RECIP

'The dog and the cat are fighting with each other.'

① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

② Is the English translation of this sentence accurate? Yes / No.

③ If the translation is inaccurate, how would you modify it?

④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

8. 两 只 钟 齐 勿 准 个。

Lian tsaq tson ze veq tsen geq

Two CL clock QUANT NEG accurate SFP

A. 'Neither of the two clocks is accurate.'

B. 'None of these clocks is accurate. (There are three clocks, for example.)'

① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

② Is the English translation of this sentence accurate? Only A is correct. / Only B is correct. /

Both A and B are accurate translations. / Neither A nor B is accurate.

③ If the translation is inaccurate, how would you modify it?

④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

9. 杯 茶 有点 冷 脱 了。

Pe zo yeuti lan theq leq

MW_{-cup} tea a.little cold ASP SFP

'This cup of tea is a little cold.'

① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

② Is the English translation of this sentence accurate? Yes / No.

- ③ If the translation is inaccurate, how would you modify it?
 ④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

10. 伊 吃了 杯 茶 就 跑 了。

Yi chiq-leq pe zo zieu bau leq

3rd.SING eat-ASP MW_{-cup} tea then leave ASP

'He had a cup of tea and went away.'

- ① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

- ② Is the English translation of this sentence accurate? Yes / No.

- ③ If the translation is inaccurate, how would you modify it?

- ④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

11. 依 瓶 饮料 是 啥 个 味道 啦?

Non bin inliau zy sa geq midau la

2nd.SING MW_{-bottle} soft.drink COP what LP flavour SFP

'What is the flavour of your drink in the bottle?'

- ① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

- ② Is the English translation of this sentence accurate? Yes / No.

- ③ If the translation is inaccurate, how would you modify it?

- ④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

12. 黄颜色 瓶 饮料 比 罇 瓶 贵。

Waoon-ngeseq bin inliau pi geq bin ciu

Yellow-colour MW_{-bottle} soft.drink COMP this MW_{-bottle} expensive

'The yellow drink in the bottle is more expensive than this one.'

- ① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

- ② Is the English translation of this sentence accurate? Yes / No.

- ③ If the translation is inaccurate, how would you modify it?
- ④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?
- ⑤ Which part(s) of the more expensive drink must be yellow: the bottle / the drink (liquid) / either of them.

13. 吾 用来 练 书法 瓶 墨水 勿见 脱 了。

Ngu yonle li syfaq bin meqsy veq-ci theq leq

1st.SING use practice calligraphy MW_{-bottle} ink NEG-see PARTICLE SFP

'The bottle of ink that I use to practice calligraphy is missing.'

- ① Is this sentence acceptable in Shanghainese?
- Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable
- ② Is the English translation of this sentence accurate? Yes / No.
- ③ If the translation is inaccurate, how would you modify it?
- ④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

14. 块 蛋糕 脱 瓶 饮料 侬 忒 甜 了。

Khue dekau theq bin inliau ze theq di leq

Piece cake and MW_{-bottle} drink both too sweet SFP

'The piece of cake and the bottle of drink are both too sweet.'

- ① Is this sentence acceptable in Shanghainese?
- Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable
- ② Is the English translation of this sentence accurate? Yes / No.
- ③ If the translation is inaccurate, how would you modify it?
- ④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

15. 伊拉 买了 串 葡萄 脱 箱 苹果。

Yila ma-leq tshoe budau theq sian binku

3rd.PL buy-ASP bunch grape and box apple

'They bought a bunch of grapes and a box of apples.'

- ① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

② Is the English translation of this sentence accurate? Yes / No.

③ If the translation is inaccurate, how would you modify it?

④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

16. 介 小 只 包 哪能 塞得进 一 大 包 衣裳 得 一大本 词典
呢?

Ka siau tsaq pau nanen saqteqtsin iq du pau izaon teq iq-du-pen zyti ne

So small CL bag how can.cram one big MW_{-bag} clothes and one-big-CL dict. SFP

‘How can a big pack of clothes and a bulky dictionary fit into such a small bag?’

① Is this sentence acceptable in Shanghainese?

Perfectly acceptable – I incline to accept it – I incline not to accept it – totally unacceptable

② Is the English translation of this sentence accurate? Yes / No.

③ If the translation is inaccurate, how would you modify it?

④ Can you modify the Shanghainese sentence slightly to make it perfectly acceptable?

Do you have any questions and comments on this questionnaire?

(To be answered by the participants)

[End of the questionnaire]