2025

The Grammar of Approximating Number Pairs (ANPs) in Mandarin Chinese

Xuehuai He

The Grammar of Approximating Number Pairs (ANPs) in Mandarin Chinese

Xuehuai He

Pomona College
March 17, 2024

Submitted to the Department of Linguistics and Cognitive Science at Pomona College
in Partial Fulfillment of the Degree of Bachelor of Arts

Contents

1 Introduction

Numerals, despite being mathematically well-defined, display different semantic behaviors in natural languages to convey more complex and nuanced meanings (Bylinina and Nouwen, 2020). Hence, much less is known about how numerals and quantitative phrases are conditioned in natural speech: whether there are rules from cultural systems, school-taught mathematical number systems, or inherent cognitive reality on our perceptions of quantities and estimation (Eriksson et al., 2010). The much more precisely defined numbers in mathematical number systems also have to adapt to the frequent occurrences of uncertainties in real discourses, leading to many modifications so as to express approximate quantity and measure. One of such means is precision regulators, such as *approximately, more or less,* etc. In addition, Krifka (2005) observed that even unmodified numbers themselves, such as *thirty minutes*, in fact allow some approximate interpretations. People also use pairs of numerals, namely Approximating Number Pairs (ANPs), for a more precise description of the extent of uncertainty involved (Eriksson et al., 2010).

Approximating number pairs (ANPs) are a kind of range-approximation-like expression which takes a pair of numbers and expresses a quantity close to the pair. Examples in English include:

(1) There were **twenty or thirty** people in the hallway.

Despite taking the form of a disjunction, ANPs do not express a literal disjunctive meaning of $20 \vee 30$ in (1). Rather, it expresses a rough range between 20 and 30 (Solt, 2018). We could further observe that the roughness of the range is a necessary part of its meaning: we note that (1) is not equivalent to *There were between twenty and thirty people in the hallway.*

In addition, there are constraints on the structure and choice of numerals: while expressions like 'six or seven books' and 'three or four thousand participants' are commonly used, 'twenty- or thirty-six cats' and 'thirty or fifty biscuits' are either anomalous or denote a precise disjunction rather than being approximative.

The phenomenon of ANPs is observed cross-linguistically (Solt, 2018), but the constructions have only been thoroughly described in English, French, German and Dutch as in Pollmann and Jansen (1996); experimental studies have been conducted in Swedish and English (Eriksson et al., 2010), and formal semantics analyses have been done on English ANPs based on the theory of granularity (Solt, 2018). Few studies have been conducted on similar or equivalent concepts in languages that are less genealogically and structurally approximate to the aforementioned languages. This paper extends the current analyses and explorations to Mandarin Chinese, a classifier language with a different number and measure system from the aforementioned languages; more on that in section 2.2.

The structure of the paper is as follows. I first give an overview of relevant literature on numerals, theories of approximation and characterizations of ANPs in other languages. Then, I dive into the grammar of the number-generation system in section 2.2.1 and nominal classifiers in Mandarin Chinese in section 2.2.2, noting their relevance to ANPs. Next, in section 2.3, I look at some other approximative expressions in Mandarin. Regarding these observations, I propose that the approximation in ANPs only scope over simple numerals in the last summand of each numeral construction, which I would elaborate in section 3. Furthermore, I am also giving a formalization of the denotation of ANPs in section 4.2 after identifying that ANPs are quantifier-like in section 4.1. Finally, I explore some future directions of relevant research.

2 Background & Motivation

2.1 Approximating number pairs

2.1.1 Constraints of ANP constructions

Functioning as the approximate range around two specified numbers, ANPs also pose constraints on which choices of numbers are allowed in the constructions of the pairs. Not every two random numbers could combine to form an ANP (Solt, 2018), and the two numbers generally cannot commute, making ANPs *ordered* pairs. Based on the corpora of the four languages Pollmann and Jansen (1996) analyzed, the restrictions on number choices are summarized as follows by Eriksson et al. (2010):

(2) Rules for well-formed ANPs:
   1. the two numbers must be in ascending order;
   2. the gap between them must be a divisor of both values;
   3. the gap must be a so-called favored number, being of the form {1/2/2.5/5}*10;
   4. the gap must be at least 5% of the second value.

The aforementioned rules applied on >90% tokens in the corpora; more recent corpus analyses done by Eriksson et al. (2010); Solt (2018) using the Swedish PAROLE corpus at Språkbanken and/or Corpus of Contemporary American English (COCA; Davies 2008- as cited) also largely corroborated the above descriptions. However, we note that these corpus analyses are all carried out in number-marking languages (Scontras, 2013) that do not systematically make use of classifiers. Since classifiers accompany nouns and frequently compose with numerals, it is far from clear whether and how this composition affects the formation of ANPs. Hence, one of the aims of this paper is to explore the possible violations or corroboration of the rules in (2) due to interactions of numerals with classifiers in Mandarin Chinese; more on that in section 2.2.2.

2.1.2 Semantics of ANPs

In addition to the construction, the current semantic analysis of ANPs characterized mainly by Solt (2018) also makes a few language-specific assumptions. Firstly, the analysis assumes that the function of the "or" between the two numbers in the ANP like the one in (1) could be viewed as an ordinary disjunction. As Solt (2018) pointed out in the paper itself, this assumption is questionable due to ANPs not being interchangeable with sentential disjunctions (e.g. *There were twenty people there or there are thirty people there*). In addition, the "or" is simply not observed in many other languages, as shown in the German and Mandarin Chinese examples in (3).

(3) a. German:
fünfzig, sechzig Meter
fifty sixty meters

fifty or sixty meters

b. Chinese:
五六十米
wu liu shi mi
five six 10 meters
fifty or sixty meters

Secondly, Solt (2018) approaches the roughness of the range through considering the set of alternatives of a numeral based on granularity, which characterizes the coarseness and fineness of uncertainties of the numeral (e.g. we can say that 30 has a larger granularity of 10 compared with 33, to which we could set the granularity as 1). The choice of granularity is quantified by the *gran* unit, which are chosen to be powers of tens, halves and doubles of tens, or by cultural conventions. The Ruler Model (Solt, 2018, p. 10) was proposed, where the set of alternatives of a numeral (or a pair, in which case the alternatives is the union of the two sets of alternatives individually) ALT_{gran} is then the set of integer multiples S_{gran} of *gran*. Then, for a number to be considered approximate enough to the truth (and hence the proposition judged semantically true), it needs to be closer to the truth than any other alternatives in the set S_{gran}. A concrete example illustrating this model is in (4).

(4) Consider the proposition: There are **fifty or sixty** people.
Let $gran = 10$. Hence, $S_{gran} = \{\dots, 30, 40, 50, 60, 70, \dots\}$.
We observe that the alternatives for this ANP excluding themselves would be

$$ALT_{gran}(\text{fifty or sixty}) \backslash \{50, 60\} = \{\dots, 30, 40, 70, \dots\}$$

Say now there are 47 people. We observe that $|47-50| \leq |47-\alpha|$ for any $\alpha \in ALT_{gran} \backslash \{50, 60\}$. Hence, it is closer to one of the numerals in the ANP than any alternatives based on the granularity of 10. It can be concluded that the proposition is true if the fact is that there are 47 people.

Then, consider the truth that there are 69 people. We note that $|69 - 70| = 1 < |69 - 60|$ and $|69 - 70| < |69 - 50|$. This means that there exists an alternative 70 that is closer to 69 than either of 50 or 60. Therefore, the proposition would be judged false according to the Ruler Model.

Offering us with a systematic way to derive the acceptable ranges of quantities approximated by ANPs, this analysis is, however, almost purely a mathematical one detached from the grammar of the language. In contrast, the construction of ANPs are highly rooted in the grammar and many alternative expressions containing the two mathematically equivalent numbers would not produce the same meaning, especially in Mandarin as discussed more in section 3. This leads us to question whether the derivation of meanings of ANPs should be completely disjoint from the grammar. We also observe that the analysis does not give a proper justification of how the unit *gran* is chosen other than mathematical divisibility concerns and social conventions. While these are highly plausible sources of influence on the approximation ranges, the resulting analysis of the choice of *gran* would be idiosyncratic to each specific number and not universally generalizable. Moreover, the rule for determining whether a quantity would be considered true if described by a certain ANP involves a quite complex calculation of absolute values of differences to all alternatives. Whether there could be a grammar-motivated and simpler explanation for the semantics of ANPs is still to be investigated.

Finally, Bylinina and Nouwen (2020) also pointed out that there are other semantic types numerals could bear: as modifiers to give the modified predicates the property of being that number, or as entities presenting the mathematical concept of the specific number. We shall also look at whether these use cases of numerals could be replaced with ANPs in Mandarin in section 4.1.

Before diving deep into the complex constructions, semantics and pragmatics of ANPs, we shall look at some characteristics of Mandarin Chinese to obtain clues on how numerals and ANPs might function in this language.

2.2 Features of Mandarin Chinese

This section aims to highlight a few features of Mandarin Chinese that differ significantly from the previously mentioned Indo-European languages. Specifically, I am going to point out a few properties that might interfere with the usages of ANPs.

2.2.1 Number system

Mandarin Chinese numerals employ the decimal system of counting (He and Zhang, 2021), with the ten unique simple numerals displayed in (5); the numerical bases are {十，百，千，万，亿，兆}, translating to numerals {10, 10^2, 10^3, 10^4, 10^8, 10^{12}}.

(5) Ten simple cardinal numbers in Chinese
零，一，二，三，四，五，六，七，八，九
ling yi er san si wu liu qi ba jiu
0, 1, 2, 3, 4, 5, 6, 7, 8, 9

The simple numerals, like the ones in English, are unsystematic and have no rules for generation (Ng and Rao, 2010). However, unlike English using more irregular number names (e.g. *eleven* for 11) or attach the morphemes of numerical bases to form larger numbers (e.g. *forty* formed by attaching the morpheme '-ty' for 10), Chinese numerals are generated systematically by placing a simple numeral before the numerical base to express the product from multiplication as in (6a), or gluing the digit after the base to express the sum from addition as in (6b). The two processes could be applied multiple times or composed to generate most digits in a regular fashion, as shown in (6c). This kind of construction of numerals is called *base-final*, where the base serving as the multiplicand attaches after the simple numeral multipliers (Her, 2017).

(6) a. 五十
wu shi
5 10
fifty ($50 = 5 \times 10$)

b. 十五
shi wu
10 5
fifteen ($15 = 10 + 5$)

c. 四千三百五十八
si qian san bai wu shi ba
4 10^3 3 10^2 5 10 8
four thousand three hundred and fifty-eight ($4358 = 4 \times 10^3 + 3 \times 10^2 + 5 \times 10 + 8$)

We also observe that the smallest base in the entire numeral is optional whenever there are no missing intermediate powers of 10s, as seen in (7). That is, whenever the number ends in the form $10^n + a \times 10^{n-1}$ for any $a \in [1, 9]$ and $n, n-1 \in \{2, 3, 4, 8, 12\}$, then the 10^{n-1} character could be omitted as it would be implied. Otherwise, all bases are mandatory, and the missing bases are usually marked by the zero digit 'ling'.

(7) a. Smallest base is optional when no intermediate powers of 10s are missing:
五千八（百），三十五万六（千），四百七（十）
wu qian ba (bai), san shi wu wan liu (qian), si bai qi (shi)
5 10^3 8 (10^2), 3 10 5 10^4 6 (10^3), 4 10^2 7 (10)
5800, 356000, 470

b. Smallest bases not optional when there are missing powers of 10s:
两万**零**五 *(百)，七千**零**二 *(十)
liang wan **ling** wu *(bai), qi qian **ling** er *(shi)
2 10^4 **zero** 5 *(10^2), 7 10^3 **zero** 2 *(10)
20500, 7020

One thing we should note here is that despite using the same mathematical notations of + and × as defined for integers, the implied addition and multiplication when generating Mandarin numerals are not exactly the same as that in mathematics. First, we observe that such operations are not commutative in Mandarin, as the *multiplier* × *multiplicand* sequence could not be reversed. Specifically, only bases given in the set $\{10, 10^2, 10^3, 10^4, 10^8, 10^{12}\}$ are licensed to occupy the *multiplicand* position, whereas simple numerals $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ can only occupy the *multiplier* position, as illustrated in (8).

(8) a. 四十
si shi
four 10
forty ($4 \times 10 = 40$)

b. * 四七
*si qi
four seven
Intended: $4 \times 7 = 28$

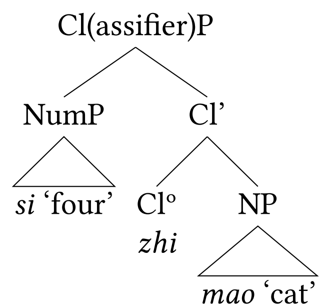
2.2.2 Nouns and classifiers

Mandarin Chinese is usually considered a classifier language (Her et al., 2022), which means that it requires nominal classifiers in the presence of numerals in order to count the referents of nouns (Scontras, 2013). There are still some ongoing inconsistencies in the classification of different kinds of classifiers (Li, 2011; Gerner, 2014; Zhang, 2011; Her et al., 2022; Jiang et al., 2022). This paper is going to proceed being less concerned about the different kinds of classifiers, but more about how they combine with numerals and nouns.

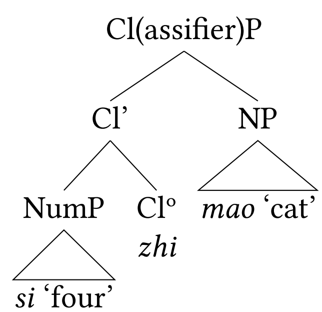
In addition, the syntactic structure of classifier phrases are still under debate. Specifically, regarding how the [Num CL NP] sequence combines, there is strong evidence corroborating both the [Num [CL NP]] hypothesis (Scontras, 2013; Jiang et al., 2022) and the [[Num CL] NP] hypothesis (Her, 2017; Jiang et al., 2022). The two structures are shown in (9) using a simple example.

(9) Possible syntactic structures of *si zhi mao* 'four cats':

a. The [Num [CL NP]] hypothesis:



b. The [[Num CL] NP] hypothesis:



The most recent comprehensive analysis by Jiang et al. (2022) argues that both structures above are needed to describe all the data observed, but both also have certain limitations and issues. The [Num [CL NP]] structure as illustrated in (9a) is more semantically favorable. As Chierchia (1998) argued, the nouns in Mandarin are number-neutral and denote the *kinds* of referents, which are incompatible with numerals, since a 'kind' cannot be counted. This differs from the count nouns in number-marking languages like English, where the noun refers to each individual object defined with measures. For instance, while the unmarked noun 'cat' would default to a concrete individual cat in English and other number-marking languages, it would be referring to an abstract "kind of thing that is named 'cat' " in Mandarin. Chierchia (1998) then proposes that numeral classifiers are functions converting kinds to atomic predicates and hence are of a semantic type $\langle e, \langle e, t \rangle \rangle$. We observe that the fact that classifiers perform a type-converting operation on the proceeding NP argues in favor of the [Num [CL NP]] hypothesis.

However, various historical word order and cross-linguistic observations have argued otherwise (Her, 2017). In addition, when classifiers internally encode a quantity, they still combine multiplicatively with the numerals in the same formulation as characterized in section 2.2.1, which made the [Num CL] constituency in (9b) favorable. I will discuss the composition of numerals and classifiers more in section 2.2.3. Furthermore, He and Tan (2019) also made observations that certain approximative expressions in Mandarin support the [[Num CL] NP] structure.

Building on top of the coexistence of both hypotheses, this paper makes use of both the type-converting and the number-denoting properties of nominal classifiers. That is, I assume that its type-converting property acts on the NP and ensures that the combination of numerals and nouns are grammatical, while simultaneously its number-denoting property would still act on the numeral multiplicatively to denote the correct quantity. While the choice of syntactic structure between (9a) and (9b) is still under debate, the coexistence of the type-converting and the number-denoting properties in classifiers is to a great extent built on sound ground.

2.2.3 A comparison between numbers and classifiers

Despite the inconsistencies in the subcategorization of nominal classifiers in Mandarin, there are several main subcategories that could accommodate all current theories (Jiang et al., 2022). For instance, it is generally accepted that the subcategory of measure classifier describes a standard unit of measurement, such as 一**斤**土豆 'a **kilo** of potatoes'; group classifiers collect individuals into aggregates, such as 两**束**花 'two **bunches** of flowers' (Jiang et al., 2022).

From these examples, we could note that the classifiers internally encode a quantity; in English, 'Two **kilos** of potatoes' is certainly not the same as 'two potatoes', but instead means the number 'two' multiplied by the quantity implicitly encoded in the classifier of 'kilo'. This multiplicative process also applies to Mandarin classifiers as in (10).

(10) Multiplication continues with container/measurement/group classifiers:
二十打鸡蛋
er-shi da jidan
two-10 dozen egg
twenty dozens of eggs ($2 \times 10 \times 12 = 240$ eggs!)

Hence, we see that classifiers could also be a multiplicand in continuation of the numbers preceding them. We could further extend this to also include individual classifiers, whose function is roughly to individuate nominal concepts and to provide counting units[1]. In this case, the implicitly encoded number would simply be 1 unit for each individual.

Despite both functioning as multiplicands, nominal classifiers and number bases are distinct in that classifiers have the ability to convert the proceeding nouns from *kinds* to atomic predicates to

[1]Another theory argues that individual classifiers provide a count/mass distinction for nouns, the existence of which is still currently debated. For more details about the two analyses, see Jiang et al. (2022, Section 24.2.4)

be counted, as discussed in Section 2.2.2. In contrast, bases do not have such function and cannot directly precede nouns in modern Mandarin without a classifier. This difference is illustrated in (11).

(11) a. * 三十书
*san-shi shu
three-10 book
Intended: 30 books

b. 三打书
san da shu
three dozen book
3 dozens of books ($3 \times 12 = 36$ books)

To summarize, Table 1 draws the three way comparison between the functions of simple digits, bases and nominal classifiers in Mandarin. These similarities and distinctions become useful when we consider the scope of ANPs as well as other approximative constructions in Mandarin.

Table 1: Features of simple digits, bases and nominal classifiers in Mandarin

	simple digits	bases	nominal classifiers
Multiplier	√	×	×
Multiplicand	×	√	√
Type conversion	×	×	√

2.3 Other approximative expressions in Mandarin

Although ANPs have not been characterized in Mandarin Chinese, various studies have attempted to characterize other forms of approximative numerals in Mandarin. One of them is the 上 *shang* (literally ‘above’) expressions, which are roughly equivalent to English and French approximatives like *thousands, millions, centaines, milliers* and so on. In languages like English and French, these approximating numbers combine with ‘of/de’ and only appear in the morphological plural form (Kayne, 2005). As a non-number-marking language, Chinese does not have a plural form for the numeral bases, but rather attaches the character ‘上’ *shang* before each base. Similarly, there is also an expression ‘多’ *duo* (literally ‘much’), which could be added after each ‘round number’ to mean the number plus some value less than itself, as in (12b) (Luo, 2018).

(12) a. 上千
shang qian
SHANG 10^3
Thousands (e.g. $5000 \in [\![shang\ qian]\!]$)

b. 七百多
qi bai duo
seven 10^2 DUO

More than 700 (e.g. $747 \in [\![qi\ bai\ duo]\!]$)

We observe that these particles still obey the 'base-final' number system of Mandarin: the *shang* particle attaches before any bases to multiply the bases by a generic number, whereas the *duo* particle is additive and hence comes after the 'round numbers', which are interpreted as bases using the terminology of this paper. Luo (2018) also noted that the particle *duo* could also be attached after mass/measure classifiers; this phenomenon could be accommodated by our observation that nominal classifiers, like bases, could also serve as multiplicands, as discussed in section 2.2.3. Furthermore, Luo (2018) observed that constructions involving the particle *duo* has two subtypes with the particle attached before or after the classifier respectively, as illustrated in (13):

(13) a. 十多杯水
shi duo bei shui
10 DUO CL(cup) water
More than ten cups of water (Between 10 and 20 cups of water)

b. 十杯多水
shi bei duo shui
10 CL(cup) DUO water
More than ten cups of water (Between 10 and 11 cups of water)

According to the theories of Luo (2018), the reason the DUO exhibits a 10-interval when added before the classifier but an 1-interval when after is that the *duo* particle (circled above) could only scope over the single unit-morpheme immediately preceding it. Luo called this the **immediacy constraint**. The ill-formed example in (14a) illustrates how this rule applies: as the *duo* scopes over the immediately preceding morpheme 'six', it violates the constraint that the number it acts on must be a 'round number'. This constraint is no longer violated when the *duo* particle moves after the classifier.

(14) a. * 六多杯水
liu duo bei shui
six DUO CL(cup) water
Intended: Between 6 and 12 cups of water

b. 六杯多水
liu bei duo shui
six CL(cup) DUO water
More than six cups of water (Between 6 and 7 cups of water)

Based on the observations above, Luo (2018) proposed the denotation of the approximative expression *duo* to be that in (15).

(15) $[\![duo]\!] = \lambda m.(\lambda n.(\lambda x.(\mu_{CARD}(x) = (n+r) \times m)))$
(where r is a real number between 0 and 1, i.e. $r \in [0,1]$; μ is a map outputting the cardinal quantity of the referent nouns x.)

An example of its application to a concrete number is illustrated in (16).

(16) qi shi
seven 10
seventy
$[\![qi\ shi\ duo]\!] = \lambda m.(\lambda n.(\lambda x.(\mu_{CARD}(x) = (n + r) \times m)))(10)(7)$
$= \lambda x.(\mu_{CARD}(x) = (7 + r) \times 10)$

I would later partially build upon this analysis and formalize the semantics of Mandarin ANPs in a way compatible with that of the *duo* construction, connecting other features of Mandarin Chinese as well as the construction of ANPs.

3 The structure of ANPs in Mandarin

Based on four European languages, Pollmann and Jansen (1996) came up with restrictions on the constructions of ANPs as formulated in (2). While making successful predictions in the select number-marking languages, whether they apply to classifier languages is unknown. Furthermore, we note that the rules in (2) generally govern the mathematical properties of numbers, such as divisibility, rather than their syntactic and semantic roles as they appear in sentences. This becomes especially worthy of note when it comes to Mandarin, in which the construction of numbers is a highly regular and productive process generated by the additive and multiplicative patterns discussed in section 2.2.1, incurring many distributional constraints for the simple numbers and bases.

As a start, we make some observations of under which circumstances well-formed and invalid ANPs would form.

3.1 Scope of approximation

3.1.1 Absence of base reduplication and disjunction

In many ways, the ANPs in Mandarin mirror the ones in languages such as English, following Pollmann and Jansen (1996). For instance, it appears that the difference cannot be a non-divisor or non-'favorite' number in both languages:

(17) a. Anomalous ANP construction in English:
He ate **eleven or fifty** biscuits.

b. Same ANP construction also anomalous in Mandarin:
他吃了**十一（或）五十块**饼干。
#ta chi-le shi-yi (huo) wu-shi kuai binggan
he eat-PERF 10-one (or) five-10 CL biscuits
He ate **eleven or fifty** biscuits.

However, when it comes to bigger numbers with bases attached to numbers as multiplicands, the behaviors start to diverge. If we consider the following pair of ANP constructions in English with the base '100' only reduplicated in (18a) but not in (18b), we would realize that the meaning remains the same and both forms are natural. In other words, the numerical base is optionally reduplicated in ANPs in English.

(18) a. Two hundred or three hundred cats.
b. Two or three hundred cats.

As English does not have a fully multiplicative number-generating system with several additional hard-coded numerals beyond ten (e.g. 'twelve') as well as the ten-meaning base not separated by a word boundary (e.g. 'forty', 'seventeen'), the reduplication of the base might sometimes even be mandatory.

In contrast, the numerical base are strictly not reduplicated when it comes to constructing ANPs in Mandarin. Among the following mirrored examples, only (19a) with the same structure as (18b) (except for the lack of disjunction 'or') can serve the function of ANPs. Reduplicating the base in (19b) would result in ungrammaticality; adding the disjunction '或' (or) in (19c) restores the grammaticality, but subsequently diminishes the intended approximation meaning and results in the meaning 'precisely 200 or 300', which makes the phrase no longer an ANP construction. Similarly, (19d) adds the disjunction between the two simple numerals without reduplicating the base, but the meaning of an ANP is still absent and the construction results in a precise quantity.

(19) a. 两三百只猫
liang-san-bai zhi mao
two-three-10^2 CL cat
Two (hundred) or three hundred cats.
b. * 两百三百只猫
*liang-bai san-bai zhi mao
two-10^2 three-10^2 CL cat
Intended: Two (hundred) or three hundred cats.
c. 两百或三百只猫
liang-bai huo san-bai zhi mao
two-10^2 or three-10^2 CL cat
(Precisely, either) two hundred or three hundred cats.
d. 两或三百只猫
liang huo san-bai zhi mao
two or three-10^2 CL cat
(Precisely, either) two hundred or three hundred cats.

3.1.2 Scope and concatenated pairs

We could make a few hypotheses based on these observations. First, I propose that there is not an underlying disjunction between the two numbers in the approximating pair, since the meaning of the 'or'-inserted concatenated pair in (19d) and that of the 'or'-absent pair in (19a) are contrastive, with only the 'or'-absent concatenated pair having the approximating meaning. Second, reduplicating the numeral base would force us into either syntactic anomaly or the precise disjunction, meaning that we cannot form valid ANPs whenever we try to separate the two simple numerals in the number pair and attempt to attach the base to each one of them (i.e. two-three-10 → two-10 or three-10). In fact, we note that the only valid way to form an ANP is to concatenate two simple numerals from the set $\{1, 2, \ldots, 9\}$[2] and then add a base as needed. Effectively, the two concatenated simple numerals replaces the position of a single numeral (i.e. two-10 → two-three-10).

This leads us to question the constituency of the individual numbers inside each Mandarin ANP construction. We observe that the two simple numerals in the pair are not separable to individually combine with their shared base, and the pair altogether has roughly the same distribution as a single simple numeral. In terms of construction, these observations come down in favor of the hypothesis that the two simple numerals in an ANP is in fact a single, inseparable constituent that functions like a single numeral that could further multiplicatively combine with bases to form the final ANP number, which then could be multiplicatively combined with classifiers. We shall henceforth call the two concatenated simple numerals a **concatenated pair**.

Nonetheless, to make further conclusions on this matter, we would also have to look at whether other structural constraints and the semantics of ANPs also corroborate this hypothesis.

Continuing with the hypothesis that the approximation scopes over the concatenated pair but not the bases, I next delve into which choices of the simple numerals would produce a well-formed ANP.

3.2 Choice and order of numerals in the concatenated pair

As Solt (2018) pointed out for English, not all numbers produce well-formed ANPs. We first see if these are mirrored in Mandarin.

(20) Number choices that are felicitous on approximation reading in English (Solt, 2018).

a. English:
   - (i) There were 5 or 6 people at the public meeting.
   - (ii) ...10 or 12...
   - (iii) ...15 or 20...
   - (iv) ...30 or 40...
   - (v) ...60 or 80...
   - (vi) ...500 or 600...

b. Corresponding Mandarin (character form omitted except for the first sentence):
   - (i) 公共会议里有**五六个**人
     gonggong huiyi li you **wu-liu ge** ren
     public meeting in exist **five-six CL** people

[2] Zero is a trickier situation due to it being phonologically silent in most positions of the number. To my current knowledge, there are no good ways of forming similarly constructed ANPs in Mandarin that begin or end with 0. The equivalent meanings are usually achieved through using other precision-regulating particles like '的样子'(de yangzi) *or so*, '大致/大概'(dazhi/dagai) *about*, '那么'(name) *like*, etc. The following discussions in this paper will not cover these cases.

There were/are 5 or 6 people at the public meeting.

(ii) *shi-shi-er
10-10-two
Intended: ten or twelve

(iii) *shi-wu er-shi
10-five two-10
Intended: fifteen or twenty

(iv) san-si-shi
three-four-10
thirty or forty

(v) *liu-ba-shi
six-eight-10
Intended: sixty or eighty

(vi) wu-liu-bai
five-six-10^2
FIve or six hundred

We observe that, in contrast to English, the only well-formed cases in Mandarin occur when the two numbers in the concatenated pair are consecutive and increasing. That is, the concatenated pairs must be of the form $(n, n+1)$ with n being a simple numeral. The ANPs could not be well-formed with numbers like '...15 or 20...', since they could not be written as a pair of the form $(n, n+1)$ multiplied to an existing base in Mandarin. This corroborates the previous observation that the approximation only scopes over the concatenated pair but not the base.

One might point out that certain common expressions like the one in (21a) does not follow the $(n, n+1)$ rule, as the second number is greater than the first one by 2. However, we note that (21a) is in fact not an ANP: the meaning is roughly to 'a few' in general and not correlated with the magnitudes of the numbers. We observe that such expressions are also no longer productive: shifting both numbers in the pair as in (21b) would not result in a similar meaning. Idiomatic expressions as such are therefore outside of the scope of this analysis.

(21) a. 三五个人
san wu ge ren
three five CL people
a few people (ungrammatical if 'three or five people')

b. * 四六个人
si liu ge ren
four six CL people

Formulating an ANP as containing a concatenated pair followed by bases gives us a few advantages. At a glance, it is effectively equivalent to the rules 1 & 3 proposed by Pollmann and Jansen (1996), which state that (1) the two numbers must be in ascending order; and (3) the gap must be a so-called favored number, being of the form $\{1/2/2.5/5\} \times 10$. However, the isolation of the pair $(n, n+1)$ now respects the constituencies and the internal structures of Mandarin numerals while avoiding the seemingly arbitrary usage of a 'favored number'. In addition, this formulation enables us to extend to other bases; consider:

(22) a. It took 30 or 45 **minutes**. [base-15]
b. #She ate 30 or 45 **biscuits**. [base-10]

As time operates in a base-60 number system with subbases of quarters ($60 \div 4 = 15$), we realise that '30 or 45 **minutes**' is a felicitous expression since it is underlyingly $(2,3) \times 15$, displaying the $(n, n+1)$ pair; however, biscuits are counted in English in the normal base-10 system, which means we could not find a base such that '30 or 45 **biscuits**' displays a $(n, n+1)$ pair underlyingly.

3.3 The 'last summand' hypothesis

So far, I hypothesized that the approximation scopes over the concatenated pair restricted to the form $[\![(n, n+1)]\!]$, which could then multiplicatively combine with any bases if needed. However, if we start to generate ANPs through these two rules, we would still run into a handful of exceptions:

(23) a. si-bai liu-qi-shi wan
four-10^2 six-seven-10 10^4
four million six or seven hundred thousand[3]
b. liang-san-qian wan
two-three-10^3 10^4
twenty or thirty million
c. #liang-san-bai er-shi yi
two-three-10^2 two-10 one
#two or three hundred twenty one

We observe that while the first two examples are valid ANPs, we find (23c) unnatural. Using the base-final multiplicative and additive generation mechanism as discussed in Section 2.2.1, we could rewrite the ANPs in (23) into a numerical form representing the generation mechanism:

(24)
- Number in (23a): $(4 \times 10^2 + [\![(6,7)]\!] \times 10^1) \times 10^4$
- Number in (23b): $[\![(2,3)]\!] \times 10^3 \times 10^4$
- #Number in (23c): $[\![(2,3)]\!] \times 10^2 + 2 \times 10^1 + 1 \times 10^0$

Where the addition and multiplication orders must respect the original word order in Mandarin.

Here, we could observe that the concatenated pairs inhabit different *terms* in the multiplicative and additive structure, where I define terms to be the segments separated by the addition symbol + as canonically defined for polynomials in mathematics.

[3] This might sound unnatural in English, but this is due to English lacking a concise base word for 10^4, which subsequently makes the phrase extremely hard to phonologically realize. However, such problem is not present in Mandarin and the phrase sounds natural.

Each of these terms to be added can be called a *summand*. In (23a), the entire number in the parentheses is in the end multiplied by the base 10^4; inside the parentheses, we note that there are two summands 4×10^2 and $[\![(6,7)]\!] \times 10^1$, where the concatenated pair occupies the second summand position as highlighted in the example. In (23b), there is only one summand $[\![(2,3)]\!] \times 10^3 \times 10^4$ due to there being only multiplicative structures. In (23c), we see that there are three summands present: $[\![(2,3)]\!] \times 10^2$, 2×10^1 and 1, but the concatenated pair $[\![(2,3)]\!]$ appears in the first of all three summands.

Now we could begin to make sense of why (23c) does not form a valid ANP. Since Mandarin numerals are generated in a way such that the summands must be in descending orders of magnitude, placing the concatenated pair on any summand but the last would mean that approximation at a greater range is conveyed while having the knowledge of a more precise number. This would inherently be infelicitous in a discourse. I would discuss the semantic implications of the placement of concatenated pairs in more details in section 4.2.

Therefore, I formulate this constraint of the position of the concatenated pair in numerals as the **last summand hypothesis**: *the concatenated pair must be replacing the multiplier of the final summand in a number.*

One thing to be cautious regarding this constraint is that following the last summand hypothesis does not automatically guarantee a felicitous ANP. If we consider certain anomalous examples like # 一千一百二十四五 '*intended*: one thousand twenty-four or twenty-five', we would notice that this ANP still follows the last summand hypothesis; nonetheless, this ANP is pragmatically infelicitous due to the number being too large yet the range of approximation too small. To see how this is a pragmatic effect and does not undermine the last summand hypothesis, we could replace the ANP by any other means of approximation like # 一千一百二十四左右 '*intended*: around one thousand twenty-four' and see if it is still infelicitous. In this way, we are able to diagnose which infelicitous ANPs are syntactically ill-formed and which are pragmatically unrealistic.

With the constraints of constructions in mind, I could begin to look at the meanings of ANPs in Mandarin Chinese.

4 Semantics of ANPs in Mandarin

ANPs pose us three puzzles in semantics: their semantic types, the way the approximation effect composes with ordinary numerals, and the range of approximation itself. I now take some time to address each puzzle.

4.1 ANPs are quantifier-like numerals

First, we note that the semantic roles of ANPs in sentences are not always equivalent to that of precise numerals. For instance, while the precise numeral in (25) is replaceable by an ANP, those in (26) and (27) are not.

(25) a. 花园里有三十只猫。

Huayuan li you (san-shi) zhi mao.
garden inside have three-10 CL cat.

There are thirty cats in the garden.

b. 花园里有三四十只猫。
Huayuan li you (san-si-shi) zhi mao.
garden inside have three-four-10 CL cat.

There are thirty or forty cats in the garden.

(26) a. 他们是三只很有名的猫。
Tamen shi (san) zhi hen youming-de mao.
they are three CL very famous-DE cat.

They are three very famous cats.

b. # 他们是三四只很有名的猫。
Tamen shi (san-si) zhi hen youming-de mao.
they are three-four CL very famous-DE cat.

Intended: They are three or four very famous cats.

(27) a. 四十五不是质数。
(sishi-wu) bu shi zhi-shu.
forty-five not is prime-number.

Forty-five is not a prime number.

b. # 四十五六不是质数。
(sishi-wu-liu) bu shi zhishu.
forty-five-six N is prime-number.

Intended: Forty-five or forty-six is not a prime number.

Bylinina and Nouwen (2020) pointed out that numerals possess more than one kind of semantic type depending on their semantic environments. In broad strokes, most[4] empirical data could be accounted for if we look at numerals that behave quantifier-like, property-like or entity-like. Quantifier-like numerals have almost the same semantics as quantifiers with only the specific quantity being replaced by the cardinality of the number. We observe that such numerals are fully replaceable by other quantifiers, as shown in (28).

(28) Bylinina and Nouwen (2020), extracted from example (1) with emphasis by me:
a. Some students came to the party.
b. Twelve students came to the party.

This could be used to identify numerals with this semantic type. In fact, if we replace the numeral in (25) with a quantifier like 'a few', the sentence is still felicitous.

(29) 花园里有几只猫。

[4]That is, with some exceptions such as the issue of exhaustivity of numbers. More on this in the *Future directions* section.

Huayuan li you (ji) zhi mao.
garden inside have a-few CL cat.
There are a few cats in the garden.

However, like how we could not substitute ANPs, we also could not replace the numeral in (26) or (27) by quantifiers like 'some'. According to Bylinina and Nouwen (2020), this is because the number in (26) is characterizing the property of the predicates; in this case, numerals are modifiers giving the modified predicates a property of being that number. In (27), the non-prenominal number is simply the sheer mathematical concept associated with the number.

By these diagnostics, we note that the distribution of ANPs is only complementary with that of quantifier-like numerals but not other kinds of numerals.

Identifying the semantic type, we now proceed to look at how the denotation of ANPs is formed compositionally.

4.2 Compositional semantics of ANPs

I have previously discussed in section 2.2.1 that number-denoting segments in Mandarin are the simple numeral digits, bases and classifiers, which then compose numbers using the base-final multiplication and addition in orders of descending magnitudes. In this construction, simple numerals serve as multipliers, which then combine with a multiplicand of either a base or a classifier as discussed in section 2.2.3.

In addition, we could also recall from section 3.1.2 that Mandarin ANPs are formed by replacing the simple numeral of the final summand with a concatenated pair $(n, n+1)$. Various observations such as the lack of base reduplication and disjunction have led us to the conjecture that the approximation operation therefore only scopes over the concatenated pair and not the base or other disjoint summands. However, the meaning of the entire numeral is always interpreted as a whole on the surface level, and ANPs containing the concatenated pairs exhibit an approximate quantity. Hence, we need to derive how the pair $(n, n+1)$ scoped by the approximation range combines with the rest of the segments in the numeral.

4.2.1 Initial formulation based on *duo* approximatives

We could approach so by formulating a denotation $[\![(n, n+1)]\!]$ that is analogous to $[\![duo]\!]$. According to Luo (2018), the denotation is formalized as $[\![duo]\!] = \lambda m.(\lambda n.(\lambda x.(\mu_{CARD}(x) = (n+r) \times m)))$ with m being the multiplicand and $r \in [0, 1]$, as also shown in (15). To prepare for generalization, we could rewrite this denotation by letting $R_{duo} = [0, 1]$ be a set of quantities between 0 and 1; in this case, we could write an equivalent denotation:

(30) $[\![duo]\!] = \lambda m.(\lambda n.(\lambda x.\mu_{CARD}(x) \in (n + R_{duo}) \times m))$

where the set $(n+R_{duo})\times m$ is defined as the collection of $(n+r)\times m$ for all $r \in R_{duo}$, i.e. $(n+R_{duo})\times m = \{(n + r) \times m \mid r \in R_{duo}\}$.

From here, we might notice a shortcut to forming the denotation of the concatenated pair $[\![(n, n+1)]\!]$: it seems that we could simply replace the approximation range R_{duo} with the specific range allowed by the concatenated pairs in ANPs, say R_{cp}.

For demonstration purposes, I *temporarily* assume $R_{cp} = [-\frac{1}{2}, 1\frac{1}{2}]$, as it roughly corresponds to the range of *slightly smaller than the lower number to slightly greater than the higher number*; moreover, we note that this is also mathematically consistent with the Ruler Model proposed by Solt (2018) as discussed in section 2.1.2. Thus, this assumption is sufficient for the purpose of the following analysis, but could have some limitations. I would discuss this more in the future work section.

Hence, the denotation $[\![(n, n+1)]\!]$ is simply:

(31) $[\![(n, n+1)]\!] = \lambda m.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m)$
"The quantity of the referent x is between $(n - \frac{1}{2}) \times m$ and $((n+1) + \frac{1}{2}) \times m$."

Like in (16), we could try to apply our formulation in (31) to a concrete example:

(32) qi ba bai
seven eight 10^2
seven (hundred) or eight hundred

$$\begin{aligned} [\![\textit{qi ba bai}]\!] &= ((\lambda n.[\![(n, n+1)]\!])([\![\textit{qi}]\!]))([\![\textit{bai}]\!]) \\ &\equiv ((\lambda n.[\![(n, n+1)]\!])(7))(100) \\ &= (\lambda n.(\lambda m.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m))(7))(100) \\ &= \lambda x.\mu_{CARD}(x) \in (7 + R_{cp}) \times 100 \\ &\equiv \lambda x.\mu_{CARD}(x) \in [6.5 \times 100, 8.5 \times 100] \\ &\equiv \lambda x.\mu_{CARD}(x) \in [650, 850] \end{aligned}$$

"The quantity of the referent x (to be given) is somewhere in the interval from 650 to 850."

This gives us a decent initial formulation of the denotation of ANPs. However, we note that this formulation lacks the ability to accommodate numbers with multiple summands. Consider:

(33) liang bai er san shi
two 10^2 two three 10
Two hundred twenty or thirty

If we attempt to derive the denotation of (33), we would run into problems:

(34) Denotations incur errors when applied to numbers with multiple summands:

$$\begin{aligned} [\![\textit{liang bai er san shi}]\!] &= ((\lambda n.[\![(n, n+1)]\!])([\![\textit{liang bai er}]\!]))([\![\textit{shi}]\!]) \quad (*) \\ &\equiv ((\lambda n.[\![(n, n+1)]\!])(220))(10) \\ &= (\lambda n.(\lambda m.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m))(220))(10) \\ &= \lambda x.\mu_{CARD}(x) \in (220 + R_{cp}) \times 100 \\ &\equiv \lambda x.\mu_{CARD}(x) \in [219.5 \times 10, 221.5 \times 10] \end{aligned}$$

First of all, the resulting final denotation is certainly an incorrect interpretation of what 'two hundred twenty or thirty' means; in addition, as early as in the step marked with (∗), we already observed that $[\![\textit{liang bai er}]\!]$ is not a proper Mandarin numeral construction in the first place, with the meaning '220' quite coerced. In fact, this problem is also present for the *duo* construction. Therefore, we need to make the denotation take another variable s denoting the *sum* of all other summands preceding the last summand, where the approximation scopes over.

4.2.2 Accommodating multiple-summand numerals

Hence, I revise the definition given in (31) as follows:

(35) Revised denotation of $(n, n+1)$:
$[\![(n, n+1)]\!] = \lambda m.(\lambda s.(\lambda x.\mu_{CARD}(x) \in s + (n + R_{cp}) \times m))$
"The quantity of the referent x is between $s + (n - \frac{1}{2}) \times m$ and $s + ((n+1) + \frac{1}{2}) \times m$."

Now we could properly re-derive the meaning of (33):

(36) New denotation of *liang bai er san shi* using the formalism given in (35):

$$\begin{aligned}
[\![\textit{liang bai er san shi}]\!] &= ((\lambda n.[\![(n, n+1)]\!])([\![\textit{er}]\!]))([\![\textit{shi}]\!])([\![\textit{liang bai}]\!]) \\
&\equiv ((\lambda n.[\![(n, n+1)]\!])(2))(10)(200) \\
&= (\lambda n.(\lambda m.\lambda s.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m))(2))(10)(200) \\
&= \lambda x.\mu_{CARD}(x) \in 200 + (2 + R_{cp}) \times 100 \\
&\equiv \lambda x.\mu_{CARD}(x) \in [200 + 1.5 \times 10, 200 + 3.5 \times 10] \\
&\equiv \lambda x.\mu_{CARD}(x) \in [215, 235]
\end{aligned}$$

which gives the desired meaning.

We see that this proposed formulation gives us multiple advantages. First, the entire additive and multiplicative structure of Mandarin Chinese numerals as described in section 2.2.1 is preserved, with each semantic argument having constituency and well-defined denotations. It also corroborates the 'last-summand' hypothesis, corroborating that (23c) gives an ill-formed ANP through semantics. By the formulation in (35), we would have the denotation of (23c) be $\lambda x.\mu_{CARD}(x) \in [21 + 1.5 \times 100, 21 + 3.5 \times 100]$. This is anomalous in construction, as summands in Mandarin numerals must be in descending magnitudes; in addition, it is also semantically anomalous, as the knowledge of 21 is way more precise than the approximation range of 200 according to our temporary assumption for the approximating range. Moreover, it suits the semantic type of ANPs being quantifier-like, favoring against the alternative semantic types of numerals, namely modifier-like and entity-like. Finally, this construction is highly generalizable to other approximative expressions in Mandarin Chinese; the only variable parameter in the construction would be the range or set R specific to the concrete approximative expression used.

5 Future directions

5.1 Limitations and unaccounted data

As an attempt to characterize the grammar of ANPs in Mandarin, this analysis still has a few limitations and some data that it could not account for.

5.1.1 Range of approximation

With the compositional properties of ANPs being determined in section 4.2 under the assumption that $R_{cp} = [-\frac{1}{2}, \frac{3}{2}]$, we are still left with the question of whether this is a suitable and justifiable approximation range of ANPs.

The current assumption is made as a special case of the granularity and Ruler Model proposed by Solt (2018); in Solt (2018), it is proposed that a granular unit is chosen for a given number, and the approximation range would be consisting of any number that is closer to the given number than other alternative numbers with the same granularity (see section 2.1.2 for the formal characterization and examples). However, Solt's proposal is based upon the fact that the approximation is done on the entire numeral instead of only on the simple numeral multipliers. As such, the granular units are chosen differently depending on the number itself. In this paper, I proposed in section 3.1.2 that the scope of approximation is only on the concatenated pair, and hence the granularity is always chosen to be 1 due to the pair consisting of integers 1-9. An arithmetic calculation would yield that the range $n + R_{cp} = [n - \frac{1}{2}, n + \frac{3}{2}]$ would give us the set of numbers closest to n and $n + 1$ than any other alternatives in the set of integers.

However, this range faces some empirical challenges. For one, the R_{cp} presented here is a very clear-cut closed interval, whereas we do not empirically have a precise judgement that e.g. it is definitely false to say 'there are twenty or thirty people' when there are 14, but it is definitely true when there are 15. This judgement also heavily depends on the specific contexts that subject different tolerances to imprecisions.

As such, an alternative theory is the Pragmatic Halo by Lasersohn (1999). This theory proposes that a *halo* of a numeral would be the set of values that are not *meaningfully* different from the value itself under a given context. Under this theory, we could make our R_{cp} be the set theoretic union of the halo of the two numbers n and $n + 1$ under the given context, as Lasersohn propsed that halos of complex expressions are derived compositionally from the halos of their constituents. As halos are pragmatic, they do not alter the truth value of the propositions, but a proposition could be considered felicitous if some element in its halo is true, even if the proposition itself is not.

Nonetheless, as Solt (2018) pointed out, this theory commits us to analyzing a large proportion of what speakers say using numerical expressions as strictly false. Solt used the example of "probably no rope in the world is 50 meters long without a deviation of even a few millimeters", showing that the Pragmatic Halo theory forces us to make the claim that most statements we consider felicitous are logically false, which is a big philosophical commitment. Hence, which theory of imprecision could be implemented to describe R_{cp} might still be worth some further investigations.

5.1.2 The 'exact' vs. 'at least' interpretations of numerals

This paper focused on the case where numerals are quantifiers over degrees taking the "exactly" reading (type $\langle dt, t\rangle$). This denotation is formalized in (37a). However, as the paper by Solt (2018) pointed out in the footnote, numerals could also take a lower-bounded "at least" reading via a generally applicable type shift illustrated in (37b) formalized as in Bylinina and Nouwen (2020).

(37) a. $[\![twelve]\!] = \lambda P.\max(P) = \{12\}$ (type $\langle dt, t\rangle$)
Twelve denotes a set of degree properties, namely those properties whose maximal value is 12.

b. IOTA(BE($[\![twelve]\!]$)) = 12 (type d)
Denotes the number in the set of degrees that each interval in $[\![twelve]\!]$ shares.

Numerals systematically get both readings described above. This could be seen through the example conversation in (38):

(38) Bylinina and Nouwen (2020, extracted from examples (30) and (31)):
Q: Did John take ten biscuits?
A_1: Yes, he took eleven.
A_2: No, he took eleven.

We could observe that A_1 in (38) takes the *at least* meaning of the number, whereas A_2 takes the *exact* reading, both of which are felicitous. The scenario in (38) can be completely mirrored in Mandarin Chinese:

(39) Discourse (38) in Mandarin:
Q: 约翰拿了十块饼干吗?
yuehan na-le shi kuai binggan ma?
John take-PERF ten CL biscuits Q
Did John take ten biscuits?

A_1: 是的，他拿了十一块。
shide, ta na-le shi-yi kuai
yes, he take-PERF eleven CL
Yes, he took eleven.

A_2: 没有，他拿了十一块。
meiyou, ta na-le shi-yi kuai
no, he take-PERF eleven CL
No, he took eleven.

The question becomes, thus, whether the two readings could persist if we replace the numerals with ANPs. In particular, whether there exists a contrast between the available readings of precise numerals and that of ANPs in Mandarin Chinese discourses is yet to be discussed.

5.1.3 Generalizability to other languages

We observe that the analysis developed in this paper heavily depends on the analytical and generative nature of the internal structure of Mandarin numerals. In addition, each Mandarin numeral has only one precise way of being expressed up to phonological/morphological variations (e.g. 两 vs. 二 'two'). Due to these properties, claims such as the multiplier-multiplicand distinction (discussed in section 2.2.3) and that the approximation only scopes over the concatenated pair (discussed in section 3.1.2) stand.

For certain less analytical languages like Finnish, these claims might still apply with slight modifications such that we look at the internal structure of numbers on a morpheme level instead of word level for larger numbers, as shown in (40b).

(40) (Examples extracted from Kielikello with glossing added by the author)

a. Tavaran toimitus kestää **neljä viisi** päivää.
item-SG-GEN shipping take-3SG **four five** day-SG-PAR
The shipping of the item takes **four or five** days.

b. uutta elämäniloa **neli-viisikymppisille** naisille
new-SG-PAR exuberance-SG-PAR **four–five-10-PL-ILL** woman-PL-ILL
new joy of life for **fourty- or fifty-year-old** women.

However, it is questionable whether the proposed rules would be generalizable to languages with multiple ways to express the same number. For instance, consider the following examples in English:

(41) Intended meaning: a quantity somewhere around 1300 and 1400

a. Thirteen or fourteen hundred

b. # One thousand three or four hundred

We observe that (41a) and (41b) convey the same mathematical numbers, both obey the $(n, n+1)$ rule and the last-summand rule from this analysis in the English counterpart. However, only (41a) is a well-formed ANP. This problem does not occur in Mandarin, as there is only one way 一千三/四百 'one thousand three/four hundred' to express the numerals 1300 and 1400; however, this contrast between the ANPs formed from the two ways of expressing 1300/1400 indicates that there might be some other constraints on the well-formedness of ANPs in general. These rules might even be phonological or pragmatic, which are not described in this paper.

5.2 Other observations and potential directions

5.2.1 Mandarin base-only number pairs

In addition to the type of approximating expressions formed from a numeral containing a concatenated pair, there is another set of expressions completely formed by bases and absent of simple numerals that are also approximative. They are constructed in a way such that the second base must be the closest permitted base that is lower than the first one.

(42) '万千思绪，亿万观众，千百个人，百十块钱'
wan-qian sixu, yi-wan guanzhong, qian-bai ge ren, bai-shi kuai qian
10^4-10^3 thoughts, 10^8-10^4 audiences, 10^3-10^2 CL people, 10^2-10 CL money
tens of thousands of thoughts, millions and thousands of audience, hundreds and thousands of people, hundreds or tens dollars (*translated quantities not to scale*)

Whether this expression should be considered an ANP and follow the same set of rules have yet to be studied.

5.2.2 Non-decimal languages

By the proposed analysis, the approximation only scopes over the simple numerals that could then be combined with bases. Hence, it would be interesting to see if non-decimal (i.e. non-base-10) languages also exhibit the same patterns such that the ANPs would mathematically still have multiples-of-bases differences between them. One non-base-10 language that has a very similar numeral generation mechanism with Mandarin is Iñupiaq, an Inuit language with a base-20 number system (MacLean, 2014), where the number 380 is constructed as *akimiakipiaq sisamakipiaq* '15×20 + 4×20'. There also exist base-32 and base-60 languages (Comrie, 2021). Investigations of ANPs in these languages would require more documentations of the languages.

6 Conclusions

In this paper, I presented a characterization of the grammar of Approximating Number Pairs (ANPs) in Mandarin Chinese. The paper began with documenting empirical examples of well- and ill-formed ANPs in Mandarin, followed by reviewing relevant literature on the syntax and semantics of numerals, theories of approximation, and characterizations of ANPs in other languages. I then gave an overview of the grammar of the number-generation system and nominal classifiers in Mandarin Chinese. As such, I noted that much of the existing studies are restricted to corpus studies of European languages and general mathematical characterizations of the choice of numbers used in ANPs, but the construction and meaning of ANPs in Mandarin have deep roots in grammar and are not entirely mathematical.

Accounting for the unexplained empirical differences Mandarin ANPs display, this paper provides new analyses for the constraints of construction of ANPs in Mandarin. We observed that Mandarin ANPs do not reduplicate bases or have a disjunction 'or', leading to the hypothesis that he approximation in ANPs only scopes over simple numeral multipliers. This concatenated pair of simple numerals can only be of the ordered form $(n, n+1)$, and the pair can only occur on the last summand of each numeral construction, as elaborated in the 'last-summand' hypothesis. The new analysis in the paper not only accounts for the Mandarin data, but also applies to unexplained data of ANPs in English, such as why "30 or 45 minutes" can be approximating.

Then, I give an analysis on the semantic type and formal denotation of ANPs. I first pointed out that unlike ordinary numerals that could posess quantifier-like, property-like or entity-like semantic types, ANPs are always quantifier-like. Next, I formalized the denotation of ANPs, building upon and amending the denotation of other approximative expressions. The resulting denotation preserves the internal structures of numerals while being generalizable to other similar approximative expressions.

This study still has a few limitations regarding the choice of the range of approximation, the exhaustivity of numbers and cross-linguistic generalizability. Future research could be conducted on these issues, in addition to investigating whether the proposed analyses are applicable to non-base-10 languages.

Acknowledgements

I would like to express my most profound gratitude to my thesis reader, Professor Michael Diercks, for his invaluable guidance, mentorship, and encouragement during the entire process. This endeavor would also not have been possible without Professor Mary Paster, who offered insightful feedback and kept me on track. I am also grateful to Professor Peter Jenks at UC Berkeley for providing access to his inspiring book chapter.

In addition, I deeply appreciate the emotional support from my senior seminar classmates, current and former suitemates, friends, and all the cats that kept me company.

Lastly, my heartfelt thanks go to my family for enabling me to understand the intricacies of my beloved native language. It all started from there.

References

Lisa Bylinina and Rick Nouwen. Numeral semantics. Language and Linguistics Compass, 14(8): e12390, 2020. ISSN 1749-818X. doi: 10.1111/lnc3.12390.

Gennaro Chierchia. Reference to Kinds across Language. Natural Language Semantics, 6(4): 339–405, December 1998. ISSN 1572-865X. doi: 10.1023/A:1008324218506.

Bernard Comrie. Typology of Numeral Systems. May 2021.

Kimmo Eriksson, Drew H. Bailey, and David C. Geary. The grammar of approximating number pairs. Memory & Cognition, 38(3):333–343, April 2010. ISSN 1532-5946. doi: 10.3758/MC.38.3.333.

Matthias Gerner. Numeral classifiers in Chinese: The syntax-semantics interface by XuPing Li. Language, 90:971–974, January 2014. doi: 10.1353/lan.2014.0083.

Chuansheng He and Dandan Tan. On the constituenthood of numeral-verbal classifier combinations in Mandarin Chinese. Language and Linguistics, 20(3):336–361, 2019. ISSN 1606822X. doi: 10.1075/lali.00044.he.

Chuansheng He and Min Zhang. Numerical Expressions in Chinese: Syntax and Semantics. In Oxford Research Encyclopedia of Linguistics. Oxford University Press, November 2021. ISBN 978-0-19-938465-5. doi: 10.1093/acrefore/9780199384655.013.911.

One-Soon Her. Structure of numerals and classifiers in Chinese: Historical and typological perspectives and cross-linguistic implications. Language and Linguistics, 18(1):26–71, January 2017. ISSN 1606-822X, 2309-5067. doi: 10.1075/lali.18.1.02her.

One-Soon Her, Harald Hammarström, and Marc Allassonnière-Tang. Defining numeral classifiers and identifying classifier languages of the world. Linguistics Vanguard, 8(1):151–164, December 2022. ISSN 2199-174X. doi: 10.1515/lingvan-2022-0006.

Li Julie Jiang, Peter Jenks, and Jing Jin. The Syntax of Classifiers in Mandarin Chinese. In The Cambridge Handbook of Chinese Linguistics, number 24, pages 515–549. Cambridge University Press & Assessment, 2022. ISBN 978-1-108-42007-5.

Richard S. Kayne. A Note on the Syntax of Numerical Bases. 2005.

Kielikello. Numeroin vai kirjaimin? - Kielikello. https://www.kielikello.fi/-/numeroin-vai-kirjaimi-1.

Manfred Krifka. Approximate Interpretations of Number Words : A Case for Strategic Communication. 2005.

Peter Lasersohn. Pragmatic Halos. Language, 75(3):522–551, 1999. ISSN 0097-8507. doi: 10.2307/417059.

Xu-Ping Li. On the Semantics of Classifiers in Chinese. PhD thesis, Bar-Ilan University, Ramat Gan, Israel, 2011.

Qiongpeng Luo. Approximate Constructions Using duo ‘more’ in Chinese. In Yunfang Wu, Jia-Fei Hong, and Qi Su, editors, Chinese Lexical Semantics, Lecture Notes in Computer Science, pages 64–78, Cham, 2018. Springer International Publishing. ISBN 978-3-319-73573-3. doi:

10.1007/978-3-319-73573-3_5.

Edna MacLean. Iñupiatun Uqaluit Taniktun Sivuninit/Iñupiaq to English Dictionary. University of Alaska Press, October 2014. ISBN 978-1-60223-234-1.

Sharon Sui Ngan Ng and Nirmala Rao. Chinese Number Words, Culture, and Mathematics Learning. Review of Educational Research, 80(2):180–206, 2010. ISSN 0034-6543.

T. Pollmann and C. Jansen. The language user as an arithmetician. Cognition, 59(2):219–237, May 1996. ISSN 0010-0277. doi: 10.1016/0010-0277(95)00704-0.

Gregory Scontras. Accounting for counting: A unified semantics for measure terms and classifiers. Semantics and Linguistic Theory, pages 549–569, August 2013. ISSN 2163-5951. doi: 10.3765/salt.v23i0.2656.

Stephanie Solt. Granularity and approximating number pairs. In Proceedings of IATL 2015, October 2018.

Niina Ning Zhang. The Constituency of Classifier Constructions in Mandarin Chinese. Taiwan Journal of Linguistics, 9(1), January 2011. doi: 10.6519/TJL.2011.9(1).1.