Approximation on Numerals
Insights from the Grammar of Approximating Number Pairs (ANPs) in Mandarin Chinese

Xuehuai He
Pomona College

Approximating number pairs (ANPs), like in 'twenty or thirty people', appear cross-linguistically and usually indicate an approximate range rather than a precise disjunction of either twenty or thirty. This paper explores the syntax and semantics of well-formed ANPs in Mandarin Chinese, proposing generalizations in semantic types and scopes of approximation across various approximative numeral expressions in Mandarin Chinese. I propose that approximative expressions on numerals are independent of the specific numerals chosen and the only variable parameter in the construction would be the range specific to the concrete approximative expression used.

0. Introduction

Mathematically well-defined numerals display different semantic behaviors in natural languages to convey more complex and nuanced meanings (Bylinina and Nouwen, 2020). Given the typical uncertainties of discourse, languages employ strategies to express approximate quantity and measure. One such strategy is to 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); in Mandarin, expressions like 二三十 ‘twenty or thirty’, 六七百万 ‘six or seven million’ present similar effects. However, constructions of ANPs 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 structurally further from the aforementioned languages. It is notable that these constraints are not only mathematical but also grammatical, as mathematically equivalent number pairs do not necessarily produce the same meaning. For instance, Solt (2018) observed that ANPs are not interchangeable with sentential disjunctions (e.g. *There were twenty people there or there were thirty people there*).

This paper will begin in section 1 with an overview of relevant literature on the syntax and semantics of numerals, theories of approximation, and characterizations of ANPs in other languages. Section 2 will focus on the grammar of the number-generation system and nominal classifiers in Mandarin Chinese. Then, in section 3, I will make key observations on the features of ANPs in Mandarin Chinese that will support an analysis of the grammatical structure of well-formed ANPs in section 4. From the grammatical structures, the meanings of ANPs and how they connect with other approximative numeral expressions in Chinese will be highlighted in sections 4-7. Finally, I will present some limitations of this paper and explore future directions of research in sections 8-9.

1. Current analyses of the construction of ANPs

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 and/or Corpus of Contemporary American English (COCA; Davies 2008- as cited) also largely corroborated the above descriptions. However, we note that such rules focus on the mathematical properties of numbers such as divisibility, but not the grammatical construction of ANPs. In addition, these corpus analyses are all carried out in number-marking languages (Scontras, 2013) that do not systematically make use of nominal classifiers. Due to the unclear composition of the classifier with the numeral in [Num Cl NP] constructions (Jiang et al., 2022), it is of interest to see if classifier languages like Chinese exhibit any deviations to these rules (Her et al., 2022).

In addition to the construction, current semantic analysis of ANPs characterized mainly by Solt (2018) 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. This assumption is questionable due to ANPs not being interchangeable with sentential disjunctions. 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}. I demonstrate 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

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

Example: 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.
Non-example: Say 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 did not provide justifications for how the unit *gran* is chosen other than divisibility concerns and social conventions. Moreover, the rule for determining whether a quantity would be considered true involves complicated calculations with all alternatives. Whether there could be a grammar-motivated and simpler explanation for the semantics of ANPs is still to be investigated.

Finally, ANPs are far from the only form of approximative numeral expressions; the English expression of ***Thousands** of people* also denotes an approximate order of magnitude rather than precise multiples of the number 1000. Similarly, Mandarin expressions like 七百多 [seven 10^2 DUO] also simply vaguely denote more than 700. Current analyses of ANPs are, to my knowledge, not connected with other approximative expressions. Therefore, it would be meaningful to also look into the grammatical commonalities of ANPs with other approximative numeral expressions.

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. Features of Mandarin Chinese numerals and classifiers

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 digit **multiplier** before the numerical base

multiplicand 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).

(6) a. 五十
wu shi
5 10
fifty (50 = 5 × 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 ($4358 = 4 \times 10^3 + 3 \times 10^2 + 5 \times 10 + 8$)

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. 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 (e.g. 四十 [4 10] 'forty' is felicitous, but *四七 [4 7] is not due to 7 not allowed in multiplicand positions).

Moreover, Chinese is also a classifier language, 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 investigations 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 is 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 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 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, which makes the [Num CL] constituency favorable. 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 a classifier's type-converting property acts on the NP and ensures that the combination of numerals and nouns is 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 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.

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 classifiers as in (7).

(7) Multiplication continues with container/measurement/group classifiers:
二十打鸡蛋
er-shi da jidan
two-10 dozen egg
twenty dozen 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 following nouns from *kinds* to atomic predicates to be counted. In contrast, bases do not have such function and cannot directly precede nouns in modern Mandarin without a classifier. This difference is illustrated in (8).

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

b. 三十本书
san-shi ben shu
three-10 CL book
30 books

c. 三打书
san da shu
three dozen book
3 dozen 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	×	×	√

[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)

3. Empirical observations

Based on four European languages, Pollmann and Jansen (1996) came up with restrictions on the constructions of ANPs as formulated in (2). While making statistically successful predictions in the select number-marking languages, the rules generally govern the mathematical properties of numbers rather than their syntactic and semantic roles. 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 as discussed, incurring many distributional constraints for the simple numbers and bases. As a start, we make some empirical observations of under which circumstances well-formed and invalid ANPs would form.

The first thing we note is that the construction of Mandarin ANPs **prohibits the reduplication of bases**, as the repeated base in (9b) results in ungrammaticality. Adding the disjunction 或 'or' in (9c) and (9d) restores the grammaticality, but subsequently eliminates the intended approximation meaning and results in the meaning '**precisely** 200 or 300', which makes the phrase no longer an ANP construction. Hence, the two numbers also need to be immediately **concatenated without conjunctions**.

(9) 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

I will refer to the two immediately concatenated simple digits before a base as a **concatenated pair**. The only well-formed concatenated pairs in Mandarin occur when the two numbers chosen are consecutive and increasing. That is, the concatenated pairs **must**

be of the form $(n, n + 1)$ with n being a simple digit. For example, *十五二十 'fifteen or twenty' is ill-formed in Mandarin since it could not be written as a pair of the form $(n, n + 1)$ multiplied to an existing base.

One might point out that certain common expressions like the one in (10a) do not follow the $(n, n + 1)$ rule, as the second number is greater than the first one by 2. However, we note that (10a) is in fact not an ANP: the meaning is roughly to 'a few' in general and not correlated with the magnitudes of the numbers. As far as I know, such expressions are also no longer productive: shifting both numbers in the pair as in (10b) would not result in a similar meaning. Idiomatic expressions as such are therefore outside of the scope of this analysis.

(10) 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

However, having a concatenated pair of the form $(n, n + 1)$ alone does not guarantee a well-formed ANP. It also seems to be required that the concatenated pair replaces the **multiplier of the final summand** in a number. If we observe the examples in (11), only (11c) is unnatural.

(11) a. 四百六七十万
si-bai liu-qi-shi wan
four-10^2 six-seven-10 10^4
four million, six or seven hundred thousand[2]

b. 两三千万
liang-san-qian wan
two-three-10^3 10^4
twenty or thirty million

c. *两三百二十一
*liang-san-bai er-shi yi

[2] 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 a problem is not present in Mandarin and the phrase sounds natural.

two-three-10^2 two-10 one
*two or three hundred twenty one

Using the base-final multiplicative and additive generation mechanism as discussed in section 2, we could rewrite the ANPs in (12) into a numerical form representing the generation mechanism:

(12)
- Number in (12a): $(4 \times 10^2 + \mathbf{[\![(6, 7)]\!] \times 10^1}) \times 10^4$
- Number in (12b): $\mathbf{[\![(2, 3)]\!] \times 10^3 \times 10^4}$
- #Number in (12c): $\mathbf{[\![(2, 3)]\!] \times 10^2} + 2 \times 10^1 + 1 \times 10^0$

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. Each of these terms to be added can be called a **summand**. In (11a), 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 (11b), there is only one summand $[\![(2, 3)]\!] \times 10^3 \times 10^4$ due to there being only multiplicative structures. In (11c), 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. We thereby make an observation that the two concatenated numbers seem to have to **occupy the final summand** in the internal structure of a numeral.

Based on the empirical observations, we can start to analyze the structure of well-formed Mandarin ANPs.

4. ANP as one inseparable numeral

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 meanings of the 'or'-inserted concatenated pair in (9d) and the 'or'-absent pair in (9a) are **contrastive**, with only the 'or'-absent pair having the approximative 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, …, 9} and then add a base as needed. Effectively, the two concatenated simple numerals **replace 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 concatenated pair as a whole 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 concatenated pair 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. Consequently, we conclude that a Mandarin ANP is not two separate numbers connected by a disjunction, but simply one number where the last multiplier is a concatenated pair.

5. Uniqueness of the approximation range

Based on the constituency analysis, we observe that a precise numeral and an ANP only differ on whether the multiplier of the final summand is a simple digit or a concatenated pair. Since the bases and classifiers themselves are not approximative, the approximative meaning is created by the concatenated pair. For each simple digit $n \in \{1, \ldots, 9\}$, the only concatenated pair it could form is $(n, n + 1)$. Hence, there is only one available size of the range of approximation by concatenated pairs independent of number choice. Denote this range as R_{cp}. Then, the numbers that would be considered in the range of $(n, n + 1)$ would be the range $\{n + r \mid r \in R_{cp}\}$. 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 1. Thus, this assumption is sufficient for the purpose of the following analysis but could have some limitations. I will discuss this more in the future work section.

We also note that, although the range of approximation R_{cp} is created by concatenated pairs and is independent of the specific number chosen, bigger numbers still have larger approximation ranges due to further multiplication of the R_{cp} with the base. That is, since the approximation only scopes over the concatenated pair, the numeral is converted into a range before multiplication by bases or classifiers. The meaning of the entire numeral is, nonetheless, always interpreted as a whole on the surface level, so the range magnifies and/or shifts accordingly after multiplications and further additions. Hence, we now need to derive how the pair $(n, n + 1)$ scoped by the approximation range semantically combines with the rest of the segments in the numeral.

6. Connections to other approximative numeral expressions

As discussed in section 1, ANPs are far from the only form of approximative numeral expressions. Nonetheless, current analyses of ANPs, such as the four rules by Pollmann and Jansen (1996), are idiosyncratic to ANPs and not connected with other approximative expressions. Therefore, while we are investigating the semantics of ANPs, it is also of interest to look at the semantics of other approximative numeral expressions that also map precise numbers to ranges in general.

We approach this by looking at the example of the behavior of the particle 多 *duo* 'more than' in expressions like 七百多 [seven 10^2 DUO]. Luo (2018) observed the puzzle that constructions involving the particle *duo* exhibit different meanings when the particle attached before or after classifiers such as 杯 *bei* 'cup', 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 a 1-interval when after is that the *duo* particle means a *subportion* and can only scope over the single unit-morpheme immediately preceding it. Luo called this the *immediacy constraint*. This implies that (13a) means 10 cups plus a *subportion* of 10 cups, while (13b) is 10 cups plus a *subportion* of one unit cup. The ill-formed example in (14a) further 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', or a 'multiplicand' in the terminology of this paper. This constraint is, therefore, 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; m must be a 'round number'.)

To prepare for generalization, we could rewrite this denotation by letting the set $R_{duo} = [0, 1]$ be a set of quantities between 0 and 1; in this case, we could write an denotation equivalent to (15) in (16):

(16) $[\![duo]\!] = \lambda m.(\lambda n.(\lambda x.\mu_{CARD}(x) \in (n + R_{duo}) \times m))$
(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. $(n + R_{duo}) \times m = \{(n + r) \times m \mid r \in R_{duo}\}$.)
"The cardinal quantity of the referent is between $n \times m$ and $(n + 1) \times m$."

Here, we see that the proposed formulation for ANPs in this paper is compatible and generalizable to that for the *duo* particle. Like how ANPs are approximating concatenated pairs combined with any chosen multipliers and multiplicands, the denotation of the *duo* particle here reveals that *duo* is also simply a unique approximating range R_{duo} combined with any multipliers n and any multiplicands m. The final range of approximation, as seen on the surface form of the whole number, would also be scaled and shifted by n and m like ANPs. To go from ANPs to *duo*-containing expressions, we only need to substitute $R_{cp} = [-\frac{1}{2}, 1\frac{1}{2}]$ with $R_{duo} = [0, 1]$. That is, the only variable parameter in the construction would be the range or set R specific to the concrete approximative expression used, whether it be concatenated pairs, the *duo* particle or other such constituents.

7. Formal denotation of ANPs

From the analyses so far, ANPs should convert simple-digit multipliers into a fixed range to then multiply by respective multiplicands; in addition, this operation should be independent of the specific number chosen.

Hence, we could first attempt to formulate the denotation of the concatenated pair $(n, n + 1)$ in the same way as that for the *duo* particle:

(17) $[\![(n, n + 1)]\!] = \lambda m.(\lambda n.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m))$
(where the set $(n + R_{cp}) \times m$ is defined as the collection of $(n + r) \times m$ for all $r \in R_{cp}$, i.e. $(n + R_{cp}) \times m = \{(n + r) \times m \mid r \in R_{cp}\}$.)
"The quantity of the referent is between $(n - \frac{1}{2}) \times m$ and $(n + \frac{1}{2}) \times m$."

This gives us a good initial formulation of the denotation of ANPs. However, we note that this formulation gives us a small problem when it comes to numbers with multiple summands. Consider 两百二三十 [two 10^2 two three 10] ‘two hundred twenty or thirty’; if we attempt to derive its denotation by applying (17), we encounter problems:

(18) $[\![liang\ bai\ er\ san\ shi]\!] = ((\lambda n.[\![(n, n+1)]\!])([\![liang\ bai\ er]\!]))([\![shi]\!])$ (∗)
$\equiv ((\lambda n.[\![(n, n+1)]\!])(220))(10)$
$= (\lambda n.(\lambda m.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m))(220))(10)$
$= \lambda x.\mu_{CARD}(x) \in (220 + R_{cp}) \times 100$
$\equiv \lambda x.\mu_{CARD}(x) \in [219.5 \times 10, 221.5 \times 10]$

First, 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 $[\![liang\ bai\ er]\!]$ 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. Hence, I revise the definition given in (17) in (18) and re-derive 两百二三十 ‘two hundred twenty or thirty’ in (19), which gives the desired meaning provided that the assumption $R_{cp} = [-\frac{1}{2}, 1\frac{1}{2}]$ is correct:

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

(19) New denotation of *liang bai er san shi* using the proposed denotation:
$[\![liang\ bai\ er\ san\ shi]\!] = ((\lambda n.[\![(n, n+1)]\!])([\![er]\!]))([\![shi]\!])([\![liang\ bai]\!])$
$\equiv ((\lambda n.[\![(n, n+1)]\!])(2))(10)(200)$
$= (\lambda n.(\lambda m.\lambda s.(\lambda x.\mu_{CARD}(x) \in (n + R_{cp}) \times m))(2))(10)(200)$
$= \lambda x.\mu_{CARD}(x) \in 200 + (2 + R_{cp}) \times 100$
$\equiv \lambda x.\mu_{CARD}(x) \in [200 + 1.5 \times 10, 200 + 3.5 \times 10]$
$\equiv \lambda x.\mu_{CARD}(x) \in [215, 235]$

We see that this proposed formulation gives us multiple advantages. First, while the hypotheses in section 4 result in surface forms of well-formed ANPs largely congruent to the statistical generalizations described in (2), the entire additive and multiplicative structure of Mandarin Chinese numerals is preserved, with each semantic argument having constituency and well-defined denotations. The resulting denotation is also extendable to other approximative numerals.

8. Limitations

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. The most present among them being the assumption of the range R_{cp}, as 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 as explained in (4). 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 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 $R_{cp} = [-\frac{1}{2}, 1\frac{1}{2}]$ 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 judgment 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 judgment 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 proposed 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.

In addition, 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 (20a). 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 (20b) formalized as in Bylinina and Nouwen (2020).

(20) 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 discourse in (21):

(21) 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 (21) takes the *at least* meaning of the number, whereas A_2 takes the *exact* reading, both of which are felicitous. 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.

9. Other observations and future directions

From investigating ANPs, some other observations are made along the way. One future path of exploration would be how ANPs or approximative expressions in general play out in 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 data.

10. Conclusions

In this paper, I explored approximations as effects on numerals in Mandarin Chinese through the syntax and semantics of ANP expressions. Through looking at the features of ANPs, the constituency and structure of well-formed ANPs and the semantics of them, I presented an analysis on the grammar of ANPs that is generalizable to other approximative numeral expressions. Prior to this paper, existing studies are restricted to corpus studies of select European languages and general mathematical characterizations of the choice of numbers used in ANPs, but I observe that the construction and meaning of ANPs in Mandarin have deep roots in grammar and are not entirely mathematical.

Empirically, this paper records the observation and categorization that numerals in Chinese are generated additively and multiplicatively in general, with clear categorical differences between **multipliers** and **multiplicands**. I also observed that ANPs in Mandarin are formed by replacing the **multiplier** in the final summand of the numeral with a **concatenated pair** of form $(n, n+1)$ for some simple digit n, which is crucial for making the theory extendable to other approximative numeral expressions.

Analytically, I conjecture in this paper that ANPs are **not disjunctions**, unlike formerly hypothesized, and that the concatenated pairs in them are underlyingly an inseparable syntactic constituent. I hypothesize that the source of range approximation in Mandarin ANPs is concatenated pairs, and that there is only one kind of range R_{cp} specific to ANPs, but it could be translated and/or magnified in the surface form through the additive-multiplicative generation process of numerals. Finally, I note that the analyses across different approximating expressions in Mandarin Chinese should mostly only differ in the range R, which is specific to the kind of expression and not to the numbers. Upon these hypotheses, I claim that approximative numeral expressions are numerals **affected** by the approximation operation, as their approximative meanings and well-formedness are, in fact, independent of the specific numbers chosen.

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.

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.

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.

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

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.

THIJS POLLMANN AND CAREL 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.