Approximating number pairs (ANPs) and what they tell us about numeral phrases

Background. Approximating number pairs (ANPs) takes a number pair and expresses a quantity close to them, e.g. 'there are **twenty or thirty people**". Constraints are present, as expressions like '#**eleven or fifty** people" are either infelicitous or disjunctive. Sentential disjunctions with the same numbers do not have approximative effects. Such constructions and restrictions exist cross-linguistically. Existing analyses are mostly statistical generalizations on number-marking languages (Pollmann and Jansen, 1996); formal semantic analysis by Solt (2018) on ANPs **(i)** treats the 'or' in ANPs as ordinary disjunctions, and **(ii)** estimates the size of the range through considering the set of alternatives based on an arbitrary *granularity*. However, justifications for how the granularity unit is chosen, other than divisibility concerns, were not provided. Moreover, the rule for determining whether a quantity would be considered true involves complicated calculations with all alternatives. The highly generative numeral system of in Mandarin Chinese (MC), a classifier language, provides a new perspective on the morphosyntax and semantics of ANPs. ANPs can also potentially shed light on the syntax of numeral phrases [Num CL NP], whose internal structure is still under heavy debate (Jiang et al., 2022).

Claim. MC ANPs require the approximated digits to be immediately concatenated without bases or disjunctions intervening, suggesting that ANPs are not two numbers connected by disjunction but one number with a digit modified into a range. Further, the range of the function is independent of specific numbers but applicable to all ANP constructions. Such framework allows us to characterize the semantics of many other approximative expressions in MC through solely varying the range of the approximation function, while still generating surface forms consistent with aforementioned findings in other languages.

MC numeral-noun structure. MC numerals employ the decimal system with multiplicative and additive generation. Ten simple digits 0–9 serve as **multipliers** and bases / anchors (Barlow, 2025) {10, 10^2, 10^3, 10^4, 10^8, 10^{12}} as **multiplicands**. Numerals are formed in the fashion like in *siqian sanbai wushi ba* [$4 \times 10^3 + 3 \times 10^2 + 5 \times 10 + 8$] '4358'. MC also requires nominal classifiers in order to **type-shift** kinds into countable atomic predicates (Chierchia, 1998), i.e. type $\langle k, et \rangle$. They can also carry an internal quantity (Jiang et al., 2022) that serves as a **multiplicand**, like in *er shi da dan* [2 10 CL_{12} egg] '20 dozen eggs (240 eggs)'. The structure [Num [CL NP]] is supported by aforementioned type conversion (Scontras, 2013), but [[Num CL] NP] is better supported from historical word order (Her, 2017); Jiang et al. (2022) argued that both analyses are needed.

ANP in MC is a single number without disjunction. MC ANPs **prohibits the duplication of bases**, as in (1a). Duplicating the base in (1b) results in ungrammaticality. Adding the disjunction in (1c) diminishes the approximative meaning, in contrast to English. In fact, ANPs cannot be well-formed whenever there is intervention between the two simple digits in the number pair. We refer to the two immediately concatenated simple digits before a base a **concatenated pair** (circled in 1a) and denote by $(n, n+1)$.

(1) a. liang-san -**bai** zhi mao
two-three-10^2 CL cat
Two (hundred) or three hundred cats.

b. *liang-**bai** san-**bai** zhi mao
two-10^2 three-10^2 CL cat
Intended: Two (hundred) or three hundred cats.

c. liang **huo** san-bai zhi mao
two or three-10^2 CL cat
(Precisely, either) two hundred or three hundred cats.

As shown, concatenated pairs (e.g. two three) can further combine multiplicatively with bases (e.g. 10^2)

and classifiers to generate bigger numbers with coarser ranges (e.g. liang san bai da [two three 10^2 dozen] 'two or three hundred dozens'). An ANP functions as a **multiplier**.

Semantics of ANPs. Numerals that can form ANPs are all quantifier-typed. The bases and classifiers (both functioning as **multiplicands**) themselves are not approximative, so the approximative meaning is created by the concatenated pair. Denote the approximation range as R_{cp}, noting that it is independent of the number choice. Then, the numbers that are considered to be in the range of $(n, n+1)$ are the set $R_{cp}(n) = \{n + r \mid r \in R_{cp}\}$. Assuming that Solt (2018) made the right (mathematical) predictions, we provisionally set $R_{cp} = [-1/2, 3/2]$. The concatenated pair **convert the simple-digit multiplier n into the range** $R_{cp}(n)$, and then multiply by respective multiplicands (assign to m below), scaling up the range. We propose the **denotation of concatenated pairs** in (2), where the argument for s should be the sum of all other summands preceding the approximated summand.

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

E.g. the denotation of *liang bai er san shi* [2 10^2 2 3 10] 'two hundred twenty or thirty' is $\lambda x.\mu_{CARD}(x) \in [215, 235]$ ('between 215 and 235'), where $n = 2, m = 10, s = 200$.

Broader implications. This formulation of MC ANPs has three broader implications. **First**, it generalizes over other approximative expressions in MC, with the only variable parameter in the construction being the range or set R specific to the concrete approximative expression used. For instance, Luo (2018) observed the puzzle that constructions involving the particle *duo* (*lit.*: 'many') exhibit different meanings when the particle attached before or after the classifier *bei* 'cup', as illustrated in (3):

(3) a. er shi duo bei shui
two 10 DUO CL_{cup} water
More than twenty cups of water (Between **20 and 30** cups of water)

b. er shi bei duo shui
two 10 CL_{cup} DUO water
More than twenty cups of water (Between **20 and 21** cups of water)

This can be accounted for with our formulation of approximation: we only need to substitute R_{cp} with the range of approximation of *duo* $R_{duo} = [0, 1]$. The range then multiplies with 10 in (3a) and with 1 (the internal quantity of CL_{cup}) in (3b). **Second**, the patterns we observed in ANPs and approximated expressions in general **require the two [Num CL NP] hypotheses simultaneously**: while the type-shifting function of CL needs to be applied to NP first, we need the inherent quantity of CL to combine with Num first as a multiplicand, as approximation functions can apply to this inherent quantity (c.f. (3)). If we assume a uniform type for the approximative expressions in (3a) and (3b) as well as in ANPs, this seems to suggest that CL has complex semantic functions, with its number-denoting, multiplicand-like portion $CL_{\langle CARD \rangle}$ composing with Num and type-shifting portion $CL_{\langle k \mapsto et \rangle}$ composing with NP separately, i.e. [[Num $CL_{\langle CARD \rangle}$][$CL_{\langle k \mapsto et \rangle}$ NP]]. **Finally**, the semantics given here is compatible with English ANPs that can be accounted for by the *granularity* analysis, while also applicable to those that could not be easily accounted for: while 'It lasted 30 or 45 minutes' is felicitous, '# She ate 30 or 45 biscuits' is not. Granularity units arbitrarily chosen by divisibility does not explain the difference in judgements. When we restrict the form of ANPs to be a concatenated pair followed by a multiplicand, since time operates in a base-60 numeral system with subbases of quarters ($60 \div 4 = 15$), '30 or 45 **minutes**' is a felicitous expression as it is underlyingly $(2, 3) \times 15$, displaying the $(n, n+1)$ pair combined with a valid anchor 15; conversely, normal entities are counted in English in the normal base-10 system, which means we could not find a multiplicand such that '# 30 or 45 **biscuits**' displays a $(n, n+1)$ pair underlyingly.

References

Russell Barlow. The notion of the numeral base in linguistics. Philosophical Transactions of the Royal Society B: Biological Sciences, 380(1937):20240208, October 2025. doi: 10.1098/rstb.2024.0208.

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

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.

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.

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.

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.