‘and-a-half’ Numeral constructions in Hindi

SHAUNAK PHADNIS[1], *Independent*

ABSTRACT

Complex numerals combine via addition and multiplication in the syntax from a sequence of simplex numerals. In this paper, we investigate a novel class of numerals labeled ‘and-a-half’ numerals which can combine with simplex numerals via addition resulting in a simplex numeral. But across languages, the presence of ‘and-a-half’ limits the formation of complex numerals to only via multiplication. Further addition of another numeral in this structure considerably degrades the construction. This paper focuses on Hindi data and seeks to explain this restriction placed by ‘and-a-half’ by investigating its pragmatic role in setting standards of precision. The analysis presented here predicts that the planning component in communicating standards of precision is encoded at the phrasal level where once you set a low standard of precision you cannot arbitrarily raise it - which is exactly what happens when an additive component is introduced in the structure.

1 Introduction

While there are comprehensive accounts of Number systems in languages of the world (Hurford 1975 & 1987) as well as syntax-semantics of complex numerals (Ionin & Matushansky 2006 & 2018), there is a class of complex numerals across languages that has gone unnoticed. Present paper introduces this class in Hindi which will be called ‘and-a-half’ numerals *i.e.* complex numerals that typically feature an ‘and-a-half’ component as in 1a where *sāṛhe* is the corresponding ‘and-a-half’ expression in Hindi. These expressions are typically marked for counting and tend to have an approximate interpretation. Interestingly, ‘and-a-half’ numerals follow the syntax of complex numerals but they resist an additive component 1b.

(1) a. sāṛhe=tīn hazār
three=half thousand
‘3,500’

b. #sāṛhe=tīn hazār saat
half=three thousand seven
‘3,507’

In section 2, I present examples from various languages and show that blocking behavior in 1b is exhibited by ‘and-a-half’ numerals across languages. I also show that the existing account for ruling out illicit complex numeral constructions via PACKING STRATEGY as

[1]alephnaught18@pm.me

described in Hurford (1975) cannot explain this blocking behavior. In section 3, I make the case that *sāṛhe* stereotypically gets an approximate interpretation. I explain the ill-formedness of constructions like 1b via RNRI principle by Krifka (2007) as well as the inferences regarding precision/vagueness arising at the Sem/Prag interface. The analysis presented here applies to 'and-a-half' numerals that denote a precise number. Section 3.4 discusses the limitations of this analysis. Section 4 concludes the paper by summarizing the account presented. In the next subsection I discuss the syntax and semantics I will be assuming for complex numerals throughout the paper.

1.1 Our assumptions about Complex numerals

Following Zabbal (2005), I assume that a complex numeral consists of a sequence of simple numerals with optional or obligatory intervening material. The internal organization of these simplex numerals determines the meaning of a complex numeral. Numbers are infinite, so it is tedious and impossible to have an idiosyncratic name for each number - given this, it is natural that our grammar incorporates recursive rules to generate possible expressions for numbers. Throughout this paper, we will follow Hurford (1975) and assume the following PS rules that generate numerals across languages sufficiently:

$$\text{NUM} \rightarrow \text{DIGIT}$$
$$\text{NUM} \rightarrow \text{NUMP}(\text{NUM})$$
$$\text{NUMP} \rightarrow \text{NUM M}$$

NUM here represents the category of all possible numerical expressions in a language. DIGIT represents any single numeral word up to the value of the base number *e.g.*, 'one', 'two',..., 'nine'. M represents the category of noun-like numeral forms that can be used as multiplicational bases like 'hundred', 'thousand' etc.

Bylinina & Nouwen (2020) present a semantics for numerals where they argue that numerals denote in the domain $\mathscr{D}_N \subset \mathbb{N}$. Given the presence of fractional cardinalities across languages that we are considering in this paper, I take the view that numerals denote in $\mathscr{D}_Q \subset \mathbb{Q}$ while assuming the same semantics. Thus, semantic composition simply corresponds with regular arithmetic operations like addition and multiplication:

$$[\![\text{seventeen}]\!] = 17$$
$$[\![\text{plus}]\!] = \lambda d \lambda d.d + d$$
$$[\![\text{times}]\!] = \lambda d \lambda d.d \times d$$

2 'and-a-half' numerals

Languages employ various means to construct complex numeral expressions which exhibit a regular pattern for counting. The same numeral can be expressed in a variety of ways as well. Consider 2 in Hindi and 3 in English where 3,500 can be expressed differently:

(2) Hindi

a. tīn hazār pānch sau
three thousand five hundred
'3,500'

b. sāṛhe=tīn hazār
half=three thousand
'3,500'

(3) English

a. three thousand five hundred
'3,500'

b. three and a half thousand
'3,500'

sāṛhe and *and a half* in 2 & 3 are examples of 'and-a-half' expressions that we will be concerned with in this paper. Recall that in section 1, I briefly mentioned that 'and-a-half' expressions are present across languages. Here, I present examples of the same. Consider the following contrast for Hindi, English, Malayalam, and Russian featuring their respective 'and-a-half' expression. Notice that the 'and-a-half' numerals are degraded when there's an additional syntactic component as in b in each example:

(4) Hindi

a. sāṛhe=tīn hazār
three=half thousand
'3,500'

b. #sāṛhe=tīn hazār saat
half=three thousand seven
'3,507'

(5) English

a. three and a half thousand
'3,500'

b. #three and a half thousand eighteen
'3,518'

(6) Malayalam

a. mūn-ara laksham
three-half lac
'3,50,000'

b. #mūn-ara lakshathi irupathi
three-half lac twenty
'3,50,020'

(7) Russian

a. tr'i s polov'in-oj tys'ac-i
three with half-INS.SG thousand-GEN.SG
'3,500'

b. #tr'i s polov'in-oj tys'ac-i sorok sest
three with half-INS.SG thousand-GEN.SG forty six
'3,546'

Why are constructions in b in each example above unacceptable? Given the distribution discussed so far, *'and-a-half'* complex numerals form using the same syntax as their counterparts. The difference between a & b in the paradigm above is that constructions in b feature an additional layer of complexity in the structure. We can fine-tune our observation by noting the following contrast in Hindi:

(8) a. sāṛhe=tīn sau hazār
three=half hundred thousand
'3,50,000'

b. #sāṛhe=tīn hazār saat
half=three thousand seven
'3,507'

8a involves stacking of multipliers exploiting the multiplicative syntax but the construction itself is grammatical. As opposed to this, 8b involves an additive component. Stacking multipliers results in a number that is round thus, one can say that 'and-a-half' expressions are licensed in complex numerals that only exploit the multiplicative syntax and denote a round number but they are anti-licensed in expressions that involve an additive component. This is the contrast we want to explain in this paper.

2.1 Properties of Hindi *sāṛhe*

Hindi has two lexical entries *sāṛhe* and *ādhā* that stand for 'half'. [2] Distribution of *sāṛhe* is very limited in that it occurs only in numeral phrases combining with a NUM as in 9a while *ādhā* occurs in measure phrases (see 10a) or as a nominal modifier as in 10b, 10c. Thus, they are in complementary distribution.

(9) a. sāṛhe=tīn
half=three
'3.5'

[2] *ādhā* has its origins in Sanskrit (much like the rest of the Hindi numeral system) and it is likely that *sāṛhe* was incorporated into Hindi at a later point in history from Prakrit *saḍḍhe* - which itself perhaps came from Sanskrit *sārdh*.

b. *ādhā tīn
half three
'3.5'

(10) a. ādhā litre dūdh
half litre milk
'half a litre milk'

b. ādh-e kāgaz
half.PL paper.[PL]
'half of the papers'

c. ādh-i kitāb
half.F book.[F]
'half of the book'

As seen in 9a, *sāṛhe* prefixes to a numeral root and it is bound to it. It cannot occur independently like *ādhā*. Moreover, *sāṛhe* does not inflict for number or gender whereas, *ādhā* will often inflict for number and gender (see 10b & 10c). The combination *sāṛhe* + NUMX results in a NUM but *sāṛhe* itself doesn't seem to fit in any of the syntactic categories given by Hurford[3]. *sāṛhe* has an extremely low degree of selectivity. It only attaches to a syntactically simplex NUM. Even among these, the combination of *sāṛhe* with a simplex NUMX is acceptable for NUMX till *nineteen* but expressions for simplexes beyond that are uncommon and speaker judgments for them vary. Safely speaking, the upper bound is *ninety-nine*. In Hindi complex numerals, all multipliers can occupy the complement position of *sāṛhe* + head NUMX in a NUMP. This is significant given that not all languages allow this. See for example, in English *#three and a half hundred* is quite degraded. Now, consider the cross-linguistic paradigm in 11. We see that there is variation in how 'and-a-half' expression combines with the head NUMX. In Malayalam, there seems to be suffixation while Russian seems to resort to adpositioning and English confirms with the known coordination pattern in its numeral system. The puzzling observation is that in all the languages, there's a uniform pattern of 'and-a-half' following the head NUMX but in Hindi, this pattern is reversed. Why is this the case?:

(11) a. Hindi
sāṛhe=tīn hazār
three=half thousand
'3,500'

b. Malayalam
mūn-ara laksham
three-half lac
'3,50,000'

[3]The issue regarding the exact syntactic location of *sāṛhe* is noted but will not be addressed in this paper.

c. Russian
tr'i s polov'in-oj tys'ac-i
three with half-INS.SG thousand-GEN.SG
'3,500'

d. English
three and a half thousand
'3,500'

In the next section I will argue that *sāṛhe* is a pro-clitic thus explaining this deviation.

2.2 Explaining the word order difference in Hindi

In the previous section, we saw that *sāṛhe* doesn't confirm to the cross-linguistic pattern of 'and-a-half' following the NUMX. In Hindi, *sāṛhe* precedes the Head NUMX. See the representative examples in Hindi and Malayalam reproduced here in 12:

(12) a. Hindi
sāṛhe=tīn hazār
three=half thousand
'3,500'

b. Malayalam
mūn-ara laksham
three-half lac
'3,50,000'

I will now argue that *sāṛhe* has clitic-like properties. It doesn't seem to belong to any discernible syntactic category given by Hurford at all. Distribution-wise it falls with syntactically simplex but morpho-phonologically complex numerals motivating the position that it is perhaps a pro-clitic.

I will use criteria provided by Zwicky (1985) as well as Zwicky & Pullum (1983) to do so. They provide a host of criteria to distinguish a clitic from a particle as well as from an affix. The criteria are not meant to be exhaustively satisfied as no lexical item could exhaustively satisfy all of them but rather serves as a diagnostic to discern whether something is a clitic or an affix or an independent word.

sāṛhe forms a prosodic unit with the head NUMX. Stress and prosody pattern dictates that in a complex numeral, *sāṛhe* cannot carry stress (13a). Only the head NUMX (13b) or the multiplier (13c) can be stressed.

(13) a. #SĀṚHE=tīn hazār
half=three thousand
'3,500'

b. sāṛhe=TĪN hazār
half=three thousand
'3,500'

c. sāṛhe=tīn HAZĀR
half=three thousand
'3,500'

Earlier we noted that *sāṛhe* cannot feature independently without a host and it is bound to the head NUMX. Thus, it cannot move independently of its host either and it does not block further affixation or clitics from attaching to the host numerals as seen in 14.

(14) sāṛhe=tīn=hī
half=three=EMPH
'three and a half only'

Word order for *sāṛhe* + Head NUMX is fixed and does not change. In the previous section, we noted that *sāṛhe* exhibits extremely low selectivity for its host numerals - it only attaches to a limited set of syntactically simplex NUM. Even among these, the combination of *sāṛhe* with a simplex NUMX is acceptable for NUMX till *nineteen* but expressions for simplexes beyond that are uncommon and speaker judgments for them vary. We see in 15b that *sāṛhe* cannot modify the embedded complex numeral at all and the intended interpretation is unavailable.

(15) a. [sāṛhe=tīn] hazār
half=three thousand
'three and a half thousand'

b. #sāṛhe [tīn hazār]
half three thousand
'three thousand and a half'

Given these facts, it suffices to say that *sāṛhe* is perhaps a clitic. Now, to explain word order deviation in Hindi, Note the following pattern for syntactically simplex but Morpho-phonologically complex Hindi Numerals.

(16) a. sāṛhe=tīn
half=three
'three and a half '

b. ekkīs
one-twenty
'twenty-one'

c. chautīs
four-thirty
'thirty-four'

We see that Hindi simplexes have an underlying 'small before big' order. This is the same order found in Sanskrit where simplexes are made out of base numerals and the smaller numeral precedes the bigger one *e.g., dvādash* (twelve) = *dve* (two) + *dasham* (ten). From the point of distribution, *sāṛhe* + NUMX falls in line with this pattern - lending weight to the hypothesis that *sāṛhe* is a pro-clitic thus, explaining the word order deviation in Hindi. In the next section, we will see that the linguistic universal proposed by Hurford (1975) fails to explain why *sāṛhe* resists an additive component in Hindi complex numerals.

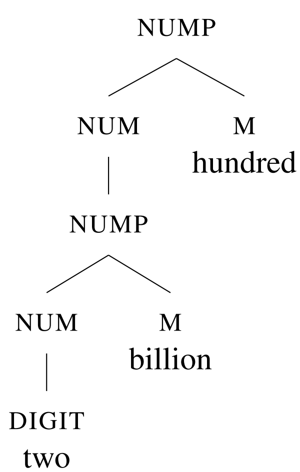
2.3 Packing Strategy and 'and-a-half' numerals

The PS rules given in section 1.1 of course might over-generate possible numeral expressions in a language. To account for this, Hurford (1975) introduces a constraint called PACKING STRATEGY as a way to block illicit complex numeral constructions in a language. The constraint states that

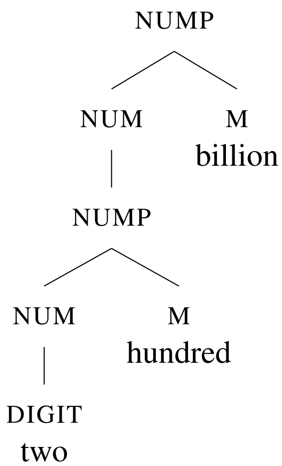
> 'Within any part of a numeral structure, the sister of NUM node must have the highest possible value given the denotation of the node that immediately dominates it.'

Hurford introduces this constraint not as a principle or a rule but rather as a linguistic universal - thus, its grammatical status is not obvious. He provides ample cross-linguistic evidence to make the case for such a universal. Packing strategy explains why expressions such as *three billion hundred* are blocked in favor of *three hundred billion*. We have the following two contesting structures corresponding to these expressions:

(17) a.



b.

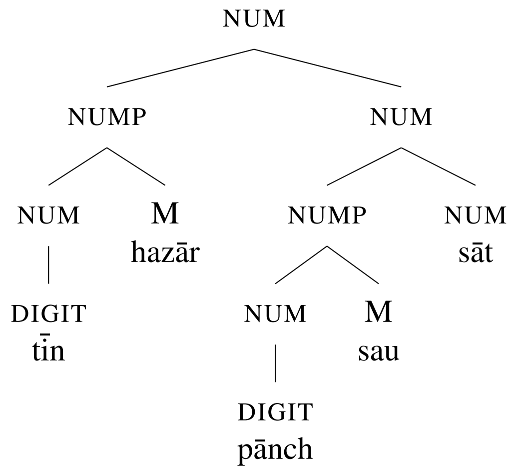


Note that packing strategy can be applied if the node immediately dominating the NUM node has the same denotation. In this case, we will be checking the constituents [NUM M] of the highest NUMP which has the same denotation in both structures. The M node in 17b has a higher value (billion) than the M node in 17a (hundred) thus the expression that corresponds to structure in 17a *i.e. three billion hundred* is ruled out. The intuition is that the strategy to construe possible numeral expressions is similar to the strategy that is used to stack books of different sizes. It would be desirable to stack the books in ascending or descending order according to their sizes. Moreover, One would club together books that are of the same size in chunks and stack the chunks according to the preferred order. Now, let's see if packing strategy helps us block constructions as in 18b against their unblocked counterparts in 18a. The corresponding structures are represented in 19b and 19a respectively:

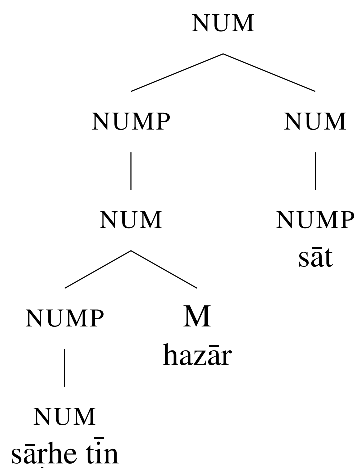
(18) a. tı̄n hazār pānch sau sāt
three thousand five hundred seven
'3,507'

b. #sāṛhe=tı̄n hazār saat
half=three thousand seven
'3,507'

(19) a.



b.



Given that only the highest NUM node in both structures has the same denotation (3,507), we will evaluate at the immediate constituent [NUMP NUM] in both structures. The [NUMP [tīn hazār]] node in 19a has the denotation 3000 while the [NUMP [sāṛhe tīn hazār]] node in 19b evaluates to 3500. Thus, the structure in 19b wins out as the denotation of the NUMP is higher than the denotation of the contesting NUMP in 19a. This is contrary to what we expect and it would seem that packing strategy makes wrong predictions. This is not ideal as the constraint itself has wide cross-linguistic empirical support and it makes sense that a constraint such as ‘packing strategy’ exists (cf. Hurford 1987 & 2007).

Hurford (2007) argues that the universal is justified by performed practice of counting objects. He provides two conceptual guidelines or maxims that are utilized - *go as far as you can with the resources you have* and *minimize the number of entities you are dealing with*. For a numeral system in a language, the basic numerals would form the resources at hand with the rules to combine them. Now, while counting one would try to use this resource exhaustively. we club objects in tens or hundreds and leave out the remainder, counting using the lexical sequence. A shift occurs when we encounter ten groups of hundreds and so on which motivates forming larger chunks of ten *hundreds* calling for a separate lexical entry - which reflects the guideline of 'minimize the number of entities you are dealing with'. Assuming that *sāṛhe* entered into the numeral system in Hindi much later it stands to reason that 'and-a-half' numerals in Hindi fall outside the explanatory domain of packing strategy. By the time they entered the Hindi lexicon, the counting system was already in place and well-developed. as a performative practice, 'and-a-half' numerals are employed in contexts where high standards of precision are not required whereas counting is a precise activity. Thus, packing strategy does not apply to them meaning we must look for an alternative explanation. In the next section, I argue that *sāṛhe* has a stereotypically approximate interpretation and explain the blocking phenomenon.

3 Analysis

Our task is to explain why 'and-a-half' in Hindi resists environments that feature an additive component but an additional multiplicative component is not resisted. We saw in section 2.3 that packing strategy cannot rule out 'and-a-half' numerals with an additive component in Hindi over their contesting counterparts with the same denotation. In this section, I will argue that *sāṛhe* stereotypically gets an approximate interpretation and the additive component is blocked due to contradictory inferences arising at the Sem-Prag interface owing to the approximative nature of *sāṛhe* and precise interpretation forced by denotation of numeral expressions. Moreover, forms containing *sāṛhe* are optimal among contesting forms to communicate a low standard of precision in a context. Semantically *sāṛhe* has the denotation 0.5 and its approximative meaning will be modelled via 'pragmatic halos' (Lasersohn, 1999). But before that, I will survey an important tool required to address the main puzzle at hand.

3.1 Krifka on approximate interpretation of number words

Krifka (2007) building on his previous work in Krifka (2002) develops a pragmatic theory of approximate interpretation of numbers in terms of strategic communication. The aim of these papers is to model approximate vs. precise interpretation of expressions such as *one hundred* over *one hundred and three*. To this end, he states an empirical generalization called RNRI principle which states *'round numbers in measuring contexts tend to receive round interpretation while precise numbers get interpreted precisely'*. Given that approximate and precise interpretations serve different roles in communication, it would be incorrect to say that the former is preferred over the latter as a general preference by speakers. Which of the two can be selected needs to be derived from more general pragmatic principles. As a consequence of the Maxim of Quality or Q-principle in the neo-Gricean framework (Horn 1984), the principle INRANGE is proposed which states:

> INRANGE: The true value of a measure must be in the range
> of interpretation of the measure term.

He posits conditional speaker preference for cognitively salient values where shorter and economic expressions are preferred over complex expressions even if there's no general bias towards them. Moreover, expressions that have approximate interpretation refer to more cognitively salient values - case in point, 'and-a-half' numerals. Thus, apart from simplicity of expressions, speakers also prefer simplicity of representations. Hence the preference for expressions such as *'the meeting lasted for an hour'* over stating the exact duration because an hour is a prominent conceptual unit. Simplicity of expressions and representations are both relevant to explain approximate interpretations where a bias for simpler representations can be correlated with a bias for coarse-grained scales - which refers to the level of granularity one assumes in measuring contexts. Krifka motivates a general principle named SER of which I will assume a modified and simplified version stated below - which, in the neo-Gricean tradition, correlates with R-principle.

SER : simple expressions/representation > complex expressions/representations

Simplicity of expressions can refer to phonological or syntactic simplicity. Ultimately it relates to the tendency to minimize cognitive load and Zipf's 'principle of least effort'. In the next section, we will see that INRANGE is built into our formalization and we do not need to assume it as a separate pragmatic principle. We will however require SER and RNRI to tackle the problem at hand.

3.2 Modelling approximate interpretation of *sāṛhe*

Recall that 'and-a-half' numerals in Hindi are typically used in contexts that allow for pragmatic slack and do not demand an exact answer. It is possible that in some contexts the speaker might have wrongly presumed that loose talk is warranted. In such cases, one can always demand more precision and a more informative answer can be provided - either by providing the actual measure as seen in 20a or by incorporating 'slack regulators' like *exactly* to reinforce a precise interpretation of the same 'and-a-half' numeral. We see in 20b that a similar discourse is infelicitous if a precise answer is provided and the questioner further presses on for more precision because rarely do we operate on precision levels in seconds.

(20) a. **Q:** anu-ke janam-kā samay kyā hai?
anu.GEN birth.GEN time what be
'what is anu's time of birth?'

A: dopahar sāṛhe=tīn
afternoon half-three
'three thirty in the afternoon'

Q': Nahi, sahi samay batao. kundali banani hai.
NEG exact time tell. star-chart make.F be
'no, tell me the exact time. I need it for the star chart'

A': tīn pachhīs ko
three twenty-five at

'at three twenty five'

b. **Q**: anu-ke janam-kā samay kyā hai?
anu.GEN birth.GEN time what be
'what is anu's time of birth?'

A: dopahar tīn pachhīs ko
afternoon three twenty-five at
'at three twenty five in the afternoon'

Q': #Nahi, sahi samay batao. kundali banani hai.
NEG exact time tell. star-chart make.F be
#'no, tell me the exact time. I need it for the star chart'

Lasersohn (1999) discusses several expressions that are not truth conditionally vague but are employed in contexts where exactitude is not necessary - as long as the actual value is *sufficiently close* to expressed value. Such expressions trigger a 'pragmatic halo' - which is a set containing values that are close to the denotation of the expression in pragmatically ignorable ways. Formally speaking, Given a context C, an expression α is assigned a partially ordered set $< H_C(\alpha), \leq_{\alpha,C} >$ called the pragmatic halo of α. Members of the Halo are objects which are of the same type as $[\![\alpha]\!]$ and differ from it in pragmatically ignorable ways. Moreover, it is necessary that $[\![\alpha]\!] \in H_C(\alpha)$.

This is the line of argument we adopt to model approximate interpretation of *sāṛhe*. Consider a context C where a person is looking to buy a piece of clothing in a shop. It is often the case that clothes are priced at prices like 999 Rs. or 3499 Rs. In such cases communicating the exact price is not necessary or even desired as long as we give a measure close to the actual value - in this case, *thousand* or *three and a half thousand* suffice. Thus approximating expressions like 'and-a-half' numerals are felicitous in a context if the actual measure is close to the measure expressed by the numeral expression. In terms of Halos, we have the following formalization which also ensures that INRANGE is obeyed as the requirement is built into the definition of pragmatic halos. Since we take the position that numeral expressions denote in the domain of Rational numbers, the pragmatic halo of *sāṛhe* will be an open interval containing rational numbers close to the denotation of *sāṛhe*.

$$[\![\textit{sāṛhe}]\!] = 0.5$$
$$[\![H_C(\textit{sāṛhe})]\!] = (..,0.485,..,0.486,..,0.499,..,0.5,..,0.501,..,0.53,..)$$

The resultant complex halo of *sāṛhe tīn hazār* can be obtained by point-wise composition with each element in the halo of *sāṛhe*. Since the semantic rules for the composition of complex numerals mentioned in Section 1.1 correspond to simple arithmetic operations of addition and multiplication, this composition becomes quite straightforward. One might argue that each element in the complex numeral might trigger a halo. This isn't wrong but I will gloss over this fact as it does not affect the analysis presented here in a significant way.

$$[\![\textit{sāṛhe tīn hazār}]\!] = 3500$$
$$[\![H_C(\textit{sāṛhe tīn hazār})]\!] = (..,3485,..,3486,..,3499,..,3500,..,3501,..)$$

We now have a **felicity condition for 'and-a-half' in Hindi**: Let α be an 'and-a-half' expression in Hindi being used in a context C. let x be the actual value that is being approximated. Then an utterance U containing α is felicitous in C in pragmatically ignorable ways iff $x \in H(\alpha)$.

This formalization also explains why 'and-a-half' numerals in Hindi are optimal forms for approximation among contesting forms that denote the same numeral. Consider a context where the actual measure x being approximated is 3,53,000. Now, in Hindi, we have the following forms available to approximate this apart from the form that expresses x.

(21) a. tīn lākh pachās hazār
three lac fifty thousand
'3,50,000'

b. sāṛhe=tīn lākh
half=three lac
'3,50,000'

Given the formalization sketched above, we see that $0.53 \in H(\textit{sāṛhe})$ and it follows that $3,53,000 \in H(\textit{sāṛhe tīn lākh})$. Moreover, following Krifka, SER predicts that 21b is a simpler expression/representation than 21a. Thus, 'and-a-half' expressions are optimal expressions to convey low standards of precision against their contesting counterparts in Hindi. We are now in the position to explain why *sāṛhe* blocks additive components in complex numerals.

3.3 Deriving blocking of additive components in Hindi 'and-a-half' numerals

We set out to explain the following contrast in Hindi:

(22) a. sāṛhe=tīn sau hazār
three=half hundred thousand
'3,50,000'

b. #sāṛhe=tīn hazār saat
half=three thousand seven
'3,507'

From the formalization discussed so far, we get the following denotation for 22b as well as the resultant complex halo. The presence of the additive component results in a kind of 'shifting' of the halo from where it starts in the composition.

$$[\![\textit{sāṛhe}]\!] = 0.5$$
$$[\![\textit{H(sāṛhe)}]\!] = (..,0.485,..,0.486,..,0.499,..,0.5,..,0.501,..,0.53,..)$$
$$[\![\textit{sāṛhe tīn hazār}]\!] = 3500$$
$$[\![H_C(\textit{sāṛhe tīn hazār})]\!] = (..,3485,..,3486,..,3499,..,3500,..,3501,..)$$
$$[\![\textit{sāṛhe tīn hazār sāt}]\!] = 3507$$
$$[\![\textit{H(sāṛhe tīn hazār sāt)}]\!] = (..,3492,..,3493,..,3506,..,3507,..,3508,..)$$

We established in section 3.2 that *sāṛhe* gets a stereotypically approximate interpretation. Therefore, the resultant complex halo of *sāṛhe tīn hazār sāt* gives us an R-implicature regarding the precision scale of the speaker *i.e. the speaker is operating with a low degree of precision*. But the semantics of the expression denotes a precise number. Recall that RNRI principle states 'round numbers in measuring contexts tend to get round interpretations while precise numbers tend to get precise interpretation'. Thus, given RNRI, we get a Q-implicature regarding the precision scale of the speaker *i.e. speaker is operating at a high degree of precision*. It is impossible for one to operate at both a low and high degree of precision. Moreover, a numerical measure cannot be both precise and vague simultaneously! Thus, we get two contradictory inferences at the Sem/Prag interface as the hearer concludes that the speaker is being uncooperative. This explains why 'and-a-half' expressions in Hindi block addition.

This account also captures why stacking of multipliers as in 22a is not blocked by *sāṛhe*. As a numerical measure becomes higher and higher, one is bound to lower their expectations regarding precision in stereotypical contexts. Multipliers only aid in casting a wider pragmatic halo while the number stays round - an ideal environment for *sāṛhe*. Moreover, no contradictory inferences arise as RNRI dictates that the denotation anyway gets a round interpretation.

Another advantage of formalization in terms of pragmatic halos is that it also captures the gradation in judgments Hindi speakers have regarding 'and-a-half' expressions with an additive component. The difference between 23a and 23b presented below is that the former features a smaller additive component than the latter. the Hindi speakers have sharp judgments ruling out constructions like 23a (hence the '#' mark) but judgments for constructions like 23b vary across speakers (represented here with '??').

(23) a. #sāṛhe=tīn hazār sāt
three=half thousand seven
'3,507'

b. ??sāṛhe=tīn hazār pachpan
half=three thousand fifty-five
'3,555'

(Lasersohn, 1999, p. 545) argues that Halos are structured sets that have a central member - namely the denotation, where members of the halo are ordered according to their relative closeness to the central member. Moreover, the formalism does not provide a clear-cut distinction between members and non-members and there is no hard cap on the size of the halo. It varies from context to context whether certain distinctions are ignorable or not. Therefore, farther elements in the halo would give us looser judgments like 23b. For some speakers, the additive element in 23b is not that far and the distinction is pragmatically ignorable hence the construction is judged to be unproblematic but for others, this is not the case. Thus, the variation in judgment regarding 23b can be explained based on whether the speaker finds pragmatic distinctions ignorable or not.

Our analysis also predicts that *sāṛhe* only blocks additive component that gets added from *below*. Constructions such as 24 where the additive component is added from *above* are felicitous as their denotation is round - which, based on the analysis so far, does not give us any inference regarding high precision levels that contradicts with low precision inference associated with *sāṛhe*.

(24) chār lākh sāṛhe=tīn hazār
four lac three=half thousand
'4,03,500'

An important consequence of this is that the planning component in strategic communication seems to be encoded at the phrasal level *i.e.* even within a phrase, once you set a low degree of precision you cannot arbitrarily raise it. But in cases like 24, you only go to a lower degree of precision[4].

3.4 Problems with the account

One limitation of the account presented here is that expressions like 25 will be predicted as felicitous as they are round in their denotation even though they have an additive component. Thus, we need to improve our existing account to explain data like 25. In this section, I provide an informal sketch of another possible approach.

(25) ?sāṛhe=tīn hazār chālīs
three=half thousand forty
'3,540'

So far, we are using notions such as degree/standard of precision which can be either high or low. Note that Numerals have various scales available to them which might be used to approximate a value close to the numeral. What scale one chooses to operate on depends on the context. For a number like *thirty*, There might be a context where measurements turn out to be in decimals close to 30 and the speaker might choose to approximate them with 'thirty'. In another context, if the true measure is 28 or 29, it might be felicitous to

[4] Thanks to Ashwini Deo for bringing this to my attention

approximate that with *thirty*. Thus, a number like ‘thirty’ has at least two scales on which we might approximate. But a number like *fifty-seven* may only have one scale (decimals) on account of not being a cognitively salient measure. Round numbers like *hundred, thousand, five hundred, etc.* on account of being cognitively salient and round have many more scales available than *fifty-seven*. below I show some of the possible scales associated with *thousand*. Cognitively salient measures like *thousand* naturally operate on coarse-grained scales and have multiple scales available to them

$$...,999.7,999.8,999.9,1000,1000.1,1000.2,1000.3,...$$
$$...,997,998,999,1000,1001,1002,1003,...$$
$$...,970,980,990,1000,1010,1020,1030...$$

Multipliers play a role in casting a wider halo. The composition of the multiplier with *sāṛhe* + NUMX makes a wide range of scales available to *sāṛhe*. This is an ideal environment for sāṛhe to live in to fulfill its pragmatic role of providing inferences regarding low standards of precision. Additive components can be conceived as Slack regulators in the step-by-step composition of the tree and attenuation of the Halo. This means that the resultant denotation will have a lesser number of scales to vary on. Perhaps the role of the slack regulator here is to limit the number of scales available. A lower number of scales means higher precision.

As *sāṛhe* + NUMX combines with the Multiplier in 25, we get a high number of scales to vary on. But further composition down the tree results in a number whose denotation (which is 3520) will give us a lesser number of scales to access. Thus, the Hearer fails to determine whether the speaker operates on High variation on scales or low variation on scales. Operating on two different scales simultaneously is not possible Thus we get the contradiction and perhaps this is a way to explain data such as 25 where having an additive component still results in a round number. There is a failure in being able to set the right expectations regarding standards of precision for a context. A key component of this approach would perhaps require establishing the exact syntactic relationship between *sāṛhe* and the immediate multiplier it will combine with. I leave this issue open for further research. As such, the account presented in this paper is only meant to explain why complex numerals that involve *sāṛhe* and denote a precise number are unacceptable which the present paper has been able to achieve.

4 Conclusion

A new class of Complex numerals labeled ‘and-a-half’ Numerals was introduced. ‘and-a-half’ numerals resist additive components in their structure but allow further multiplication. Present paper focuses on Hindi *sāṛhe* - which is shown to be a pro-clitic using diagnostics provided by Zwicky (1985) thus explaining the word order deviation in Hindi against the cross-linguistic pattern. It was shown that the universal named ‘packing strategy’ fails to rule away such constructions. ‘and-a-half’ numerals are marked for counting thus they fall outside the explanatory domain of packing strategy. We established that *sāṛhe* has a stereotypically approximate interpretation which is modelled via Pragmatic Halos (Lasersohn, 1999). Hindi ‘and-a-half’ expressions with an additive component give us two contradictory inferences. The pragmatics of sāṛhe triggers an R-inference that the speaker is being vague while, owing to RNRI principle given in Krifka (2007), the semantics of the ‘and-a-half’ numeral triggers a Q-inference that the speaker is being precise. This is contradictory as a number cannot be both precise and approximate at the same time. Thus explaining the contrast for the set of data on Hindi complex numerals that denote a precise number. The upshot of this analysis is that, planning component in communicating standards of precision is encoded at the phrasal level in ‘and-a-half’ numerals - once you set a low standard of precision you can only go lower, arbitrarily raising the standard of precision within a phrase results in illicit constructions and violates the cooperative principle.

References

Bylinina, Lisa & Rick Nouwen. 2020. Numeral semantics. *Language and Linguistics Compass* 14(8). e12390.

Halpern, Aaron L. 2017. Clitics. *The handbook of morphology* 101–122.

Horn, Laurence. 1984. Towards a new taxonomy for pragmatic inference: Q-and r-based implicature. *Meaning, form and use in context* .

Hurford, James. 2001. Numeral systems. insmelser, neil j. & baltes, paul b.(eds.), international encyclopedia of the social and behavioral sciences, 10756–10761. *Amsterdam: Elsevier* 10. B0–08.

Hurford, James R. 1975. *The linguistic theory of numerals*, vol. 16 Cambridge studies in linguistics. Cambridge University Press.

Hurford, James R. 1987. *Language and number: The emergence of a cognitive system*. Basil Blackwell Oxford.

Hurford, James R. 2007. A performed practice explains a linguistic universal: Counting gives the packing strategy. *Lingua* 117(5). 773–783.

Ionin, Tania & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of semantics* 23(4). 315–360.

Ionin, Tania & Ora Matushansky. 2018. *Cardinals: The syntax and semantics of cardinal-containing expressions*, vol. 79. MIT Press.

Krifka, Manfred. 2002. Be brief and vague! and how bidirectional optimality theory allows for verbosity and precision. *Sounds and systems. Studies in structure and change. A festschrift for theo vennemann* 439–458.

Krifka, Manfred. 2007. *Approximate interpretation of number words*. Humboldt-Universität zu Berlin, Philosophische Fakultät II.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 522–551.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Platts, John T. (John Thompson). 2015 [Online]. A dictionary of urdu, classical hindi, and english. Digital Dictionaries of South Asia. Accessed 6 Nov. 2022. https://dsal.uchicago.edu/dictionaries/platts/.

Rothstein, Susan. 2013. A fregean semantics for number words. In *Proceedings of the 19th amsterdam colloquium*, .

Sigurd, Bengt. 1988. Round numbers. *Language in Society* 17(2). 243–252.

Zabbal, Youri. 2005. The syntax of numeral expressions. *Ms., University of Massachusetts, Amherst* .

Zwicky, Arnold M. 1985. Clitics and particles. *Language* 283–305.

Zwicky, Arnold M & Geoffrey K Pullum. 1983. Cliticization vs. inflection: English n't. *Language* 502–513.