Classifying classifiers:
Two kinds of numeral classifiers across languages

Carol-Rose Little[†], Mary Moroney[†] & Justin Royer[‡]
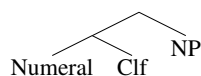[†]Cornell University, [‡]McGill University

94th Annual Meeting of the Linguistic Society of America
January 3, 2020

. . . .
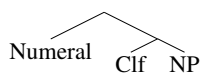
1 INTRODUCTION

What are numeral classifiers for?

- Answer 1: classifiers needed for numerals (Krifka 1995; Bale & Coon 2014); (1).

- Answer 2: classifiers needed for nouns (Chierchia 1998; Cheng & Sybesma 1999); (2).

(1) *Classifiers-for-numerals*

(2) *Classifiers-for-nouns*

- Numeral classifiers can be found in typologically diverse languages.

- For example, in Ch'ol (Mayan) and Shan (Tai) classifiers obligatorily appear with numerals:

(3) CH'OL

*ux-**kojty*** *ts'i'*
three-CLF dog

'three dogs'[1]

(4) SHAN

*măa$_i$ săam **tǒ*** *t$_i$*
dog three CLF

'three dogs'

- It's not immediately clear how to make a principled choice between the two theories.

ROADMAP

- §2 Go over two types of theories for classifiers discussed in Bale et al. 2019, which we call '**classifier-for-numeral**' theories and '**classifier-for-noun**' theories;

- §3 Discuss the semantic predictions that the two theories make, situating Ch'ol and Shan w.r.t. to these predictions. Main claim: classifiers fall into two categories, supporting both theories.

- §4 Bring in data from Chuj (Mayan) and draw connections between its two types of classifiers and the classifier systems found in Ch'ol and Shan.

---

[1]CLF = classifier; COMP = complementizer; DEM = demonstrative; IPFV = imperfective aspect; N.CLF = Chuj noun classifier; PROG = progressive aspect; #.CLF = Chuj numeral classifier

---

2 BACKGROUND

- Using the noun denotation in (6), we show how each theory derives the meaning of *two dogs* in a numeral-classifier language.[2]

(6) $[\![\text{DOGS}]\!] = \lambda x.[\text{DOGS}(x)] = \{a, b, c, ab, ac, bc, abc\}$

- (6) denotes a set containing atomic and plural dog entities.

2.1 Classifier-for-numeral theories

- These accounts argue that classifiers are needed for numerals because the numeral requires an extra semantic argument in order to compose with the noun, as in (7) (Krifka 1995; Bale & Coon 2014; Bale et al. 2019; Hall 2019).

- The classifier in (8) saturates the first argument of the numeral in (7), where $\mu_{\#}$ is a variable over measure functions.[3]
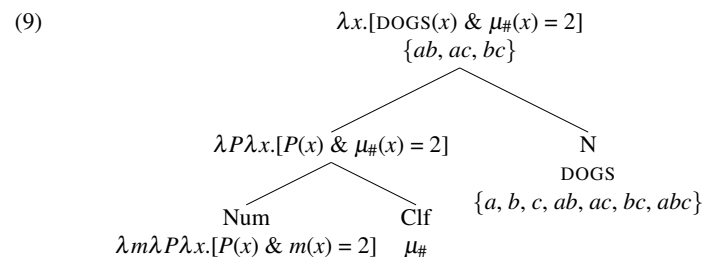
(7) $[\![\text{TWO}]\!] = \lambda m \lambda P \lambda x.[P(x) \ \& \ m(x) = 2]$ (8) $[\![\text{CLF}]\!] = \mu_{\#}$

- (7) gives the set of individuals x such that for predicate P, x has the property of P and the measure of x is 2.

- (8) is a measure function which gives the number of atoms in a plurality x (Wilhelm 2008: 55).

- The numeral in (7) takes the classifier in (8) as its first argument.

- The noun combines directly with the numeral classifier, as in (9), to yield the set of groupings of two dogs:

(9)

$$\lambda x.[\text{DOGS}(x) \ \& \ \mu_{\#}(x) = 2]$$
$$\{ab, ac, bc\}$$

$\lambda P \lambda x.[P(x) \ \& \ \mu_{\#}(x) = 2]$ N
                                                                  DOGS
                                                          $\{a, b, c, ab, ac, bc, abc\}$

Num   Clf
$\lambda m \lambda P \lambda x.[P(x) \ \& \ m(x) = 2]$ $\mu_{\#}$

---

[2](6) could also be written:

(5) $[\![\text{DOG}]\!] = \lambda x.[{}^*\text{DOG}(x)]$

(5) gives the set of individuals in the complete join-semilattice formed from the atomic set of dogs (Link 1983).

[3]All of the types of classifiers discussed in this talk vary depending on properties of the noun (e.g., whether the noun is human, animal, etc.). For simplicity, we are not representing this in the semantics today.
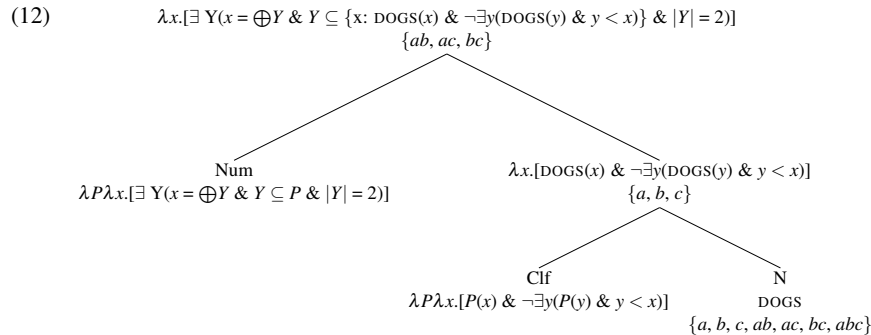
2.2 *Classifier-for-noun theories*

- These accounts argue that classifiers are needed for nouns to mediate between a noun, as in (6), and a numeral, as in (10) (Chierchia 1998; Cheng & Sybesma 1999).[4]

- We are assuming a classifier-for-noun theory following Bale et al. (2019).[5]

- The numeral needs to combine with an atomic predicate and returns the set containing all sums with property that have a cardinality of 2:

(10) $[\![\,\text{TWO}\,]\!] = \lambda P \lambda x.[\exists\, Y(x = \bigoplus Y \ \& \ Y \subseteq P \ \& \ |Y| = 2)]$ (based on Bale et al. 2019)

- (10) denotes the x such that x is the sum of entities in each Y that is a subset of the property P that includes exactly two members.

- (10) measures the set. This is different from (7) which measures the entities in the set.

- Since the noun in (6) is not atomic, classifiers, like in (11), are needed to atomize the members in the set denoted by the NP predicate.

(11) $[\![\,\text{CLF}\,]\!] = \lambda P \lambda x.[P(x) \ \& \ \neg \exists y(P(y) \ \& \ y < x)]$ (Nomoto 2013; Bale et al. 2019)

- (11) gives the set of x, such that x has the property P and there is no y with the property P that is a subpart of x.

- The classifier in (11) first combines with a noun allowing for the numeral in (10) to then combine with the classifier-noun complex, as shown in (12).

(12) $\lambda x.[\exists\, Y(x = \bigoplus Y \ \& \ Y \subseteq \{x:\ \text{DOGS}(x) \ \& \ \neg \exists y(\text{DOGS}(y) \ \& \ y < x)\} \ \& \ |Y| = 2)]$
          $\{ab, ac, bc\}$

               Num $\lambda x.[\text{DOGS}(x) \ \& \ \neg \exists y(\text{DOGS}(y) \ \& \ y < x)]$
$\lambda P \lambda x.[\exists\, Y(x = \bigoplus Y \ \& \ Y \subseteq P \ \& \ |Y| = 2)]$ $\{a, b, c\}$

                                      Clf   N
                      $\lambda P \lambda x.[P(x) \ \& \ \neg \exists y(P(y) \ \& \ y < x)]$ DOGS
                                             $\{a, b, c, ab, ac, bc, abc\}$

In sum: Though derivationally distinct, each theory produces the same meaning for *two dogs*:

  (i) For classifier-for-numeral theories, the numeral takes the classifier as a measure function, and then combines with the noun.

  (ii) For classifier-for-noun theories, the numeral cannot directly combine with the noun, and so a classifier is needed to individuate the members of the nominal predicate into a set of atoms.

---

[4] $\bigoplus$ comes from Champollion & Esipova 2018, following Sharvy's (1980) analysis of definite descriptions. For $\bigoplus$ to apply to a set, it might be defined: $\lambda Q_{\langle e,t \rangle}.\iota x(\forall y, z(Q(y) \wedge Q(z) \wedge y \neq z \rightarrow y, z \leq x))$.

[5] In some classifier-for-noun theories, such as Chierchia 1998, the noun denotes a kind and the classifier mediates a type mismatch between the classifier and the noun. We assume $\langle e,t \rangle$ type denotations for nouns.

3 TWO TYPES OF CLASSIFIERS

Despite producing similar meanings, the two theories make different predictions:

PREDICTIONS   (spelled out in more detail in §3.2)

If a classifier first forms a constituent with the...

1. NUMERAL, we might expect to find idiosyncrasies in whether or not a numeral requires a classifier, as argued in Bale & Coon 2014.

2. NOUN, we might expect to find idiosyncrasies in whether or not a noun requires a classifier, as argued in Simpson 2005 and Simpson & Ngo 2018.

3. NOUN, we might expect to find it with the noun in places other than with numerals.

4. NUMERAL, we might expect to find it with the numeral when it is not combining with a noun.

We compare two unrelated languages, Ch'ol and Shan, which have been described as having numeral classifiers, and show that while Ch'ol shows evidence for predictions 1 and 4, Shan shows evidence for predictions 2 and 3, supporting our proposal in (13):

(13) PROPOSAL
      We argue that there are two types of numeral classifiers across languages: *classifiers-for-numerals* (CLF-for-NUM) and *classifiers-for-nouns* (CLF-for-N)—supporting two types of classifier theories.

3.1 *Background*

Ch'ol is a Mayan language of the Ch'olan-Tseltalan branch, spoken in southern Mexico by $\approx$ 222,000 speakers.

Shan is a Tai language of the Southwestern Tai branch, spoken in Myanmar and surrounding countries by $\approx$ 3 million speakers.

- In both languages, classifiers obligatorily appear in the presence of a numeral:
- Which classifier appears depends on properties of the NP referent.

(14) CH'OL
      *cha'-**kojty** ts'i'*
      two-CLF   dog
      'two dogs'

(15) SHAN
      *măa sɔ̌ŋ **tǒ***
      dog two CLF
      'two dogs'

3.2   Predictions

> PREDICTION 1 (CLF-for-NUM): If a classifier is a measure function required by a numeral, there might be idiosyncrasies in whether or not a numeral requires a classifier (some numerals might have the measure function in their lexical semantics, others not).

- This is the case in Ch'ol, as shown in Bale & Coon 2014. Mayan-based numerals require a classifier, whereas those borrowed from Spanish do not:

(16) a. *ux*(-kojty) ts'i'*   b. *ocho(*-kojty) ts'i'*
         three-**CLF** dog   SP:eight-**CLF** dog
         'three dogs'   'eight dogs'   (Ch'ol)

- No such idiosyncrasies are found in Shan.

> PREDICTION 2 (CLF-for-N): If a classifier is used to create an atomic set from the noun predicate, there might be idiosyncrasies in whether or not a noun must combine with it (e.g. some nouns might only denote a set of atoms), as argued in Simpson 2005 and Simpson & Ngo 2018 for Vietnamese and other East/Southeast Asian languages.

- In Shan, some nouns do not need to combine with a classifier:

(17) *săam mɤ́ŋ*
       three country
       'three countries'   (Shan)

- This is different from Ch'ol, which always requires classifiers to combine with numerals, irrespective of the noun.

> PREDICTION 3 (CLF-for-N): If a classifier is used to create an atomic set from the noun predicate, we might expect to find it in environments other than with numerals.

- This is the case in Shan, which can have a classifier occur with quantifiers (18), with demonstratives (19), and with relative clauses (20), even in the absence of a numeral.

- In (18), the quantifier function like a numeral. In (19)-(20), the classifier atomizes the noun, giving rise to a singular interpretation.

(18) *măa ku   tŏ*   (19) *măa tŏ   nâj*
       dog every CLF   dog CLF DEM
       'every dog'   'this dog'

(20) *măa tŏ   [_RC ʔăn   nɔ́n   jù   ]*
       dog CLF   COMP sleep IPFV
       'the dog that is sleeping'   (Shan)

- This is not true in Ch'ol. Classifiers only ever occur in the presence of a numeral (or with the quantifier *jay-* 'how many').

> PREDICTION 4 (CLF-for-NUM): If a classifier is a measure function required by a numeral, we would expect it to *always* appear with that numeral.

- This is the case in Ch'ol: classifiers are *always* required, even when counting (21) and referring to directly to the number (22).

(21) CONTEXT: Students are practicing counting.
       jum-(p'ej), cha'-*(p'ej), ux-*(p'ej)* . . .
       one-CLF   two-CLF   three-CLF
       '1, 2, 3'   (Ch'ol)

(22) CONTEXT: A teacher is pointing at the number three and says:
       Ili   jiñ   ux-(p'ej).*
       this DET three-CLF
       'This is three.'   (Ch'ol)

- In contrast, Shan classifiers are not *always* required with numerals. They cannot appear with numeral when counting (23) or referring to the number (24).

(23) CONTEXT: Students are practicing counting.
       nɯ̱ŋ (?tŏ), sɔ̌ŋ (?tŏ), săam (?tŏ) . . .
       one CLF two CLF three CLF
       '1, 2, 3'   (Shan)

(24) CONTEXT: A teacher is pointing at the number three and says:
       *nâj pěn măaj   săam (*tŏ) .*
       this COP number three CLF
       'This is the number three.'   (Shan)

In sum: While Ch'ol shows evidence for the classifier-for-numeral theories, Shan shows evidence for the classifier-for-noun theories.

Table 1: Summary

		Ch'ol	Shan
CLF-for-NUM	Prediction 1	✓	✗
CLF-for-N	Prediction 2	✗	✓
CLF-for-N	Prediction 3	✗	✓
CLF-for-NUM	Prediction 4	✓	✗

Conclusion: There is evidence that both kinds of classifiers exist, as proposed in (13).

4 TWO KINDS OF CLASSIFIERS IN ONE LANGUAGE? THE CASE OF CHUJ

> QUESTION: Can a language have both classifier types at the same time?

Chuj is a Mayan language of the Q'anjob'alan branch, spoken in Guatemala and Mexico by ≈ 70,000 speakers (Piedrasanta 2009, Buenrostro 2013). It has two types of classifiers we call (following the literature) "numeral classifiers" (#.CLF) and "noun classifiers" (N.CLF):

- They can co-occur, but N.CLFs are always optional when they occur with numerals (see Craig 1986, Buenrostro et al. 1989, Zavala 2000, Hopkins 1970, 2012).

(25) *ox-wanh   (nok') tz'i'*
      three-#.CLF N.CLF dog
      'three dogs'   (Chuj)

- Chuj's #.CLFs pattern with Ch'ol CLF-for-NUMs, and not with CLF-for-Ns (see appendix).

*What about Chuj's "noun classifiers"?*

Similar syntactic distribution: Chuj *noun classifiers*, which mark definiteness and specificity (see Buenrostro et al. 1989, García Pablo & Domingo Pascual 2007, Royer to appear), pattern like Shan and unlike Ch'ol classifiers in their syntactic distribution:

- Prediction 2 (CLF-for-N): Like Shan, not all Chuj nouns can combine with a noun classifier, such as abstract nouns like *ib'* 'strength'.

- Prediction 3 (CLF-for-N): Appear in environments other than with numerals, such as alone with nouns (26), with demonstratives (27), and with relative clauses (28):

(26) **nok'** *tz'i'*   (27) **nok'** *tz'i'* **chi**
      CLF dog   CLF dog that
      'the dog'   'that dog'

(28) *nok' tz'i'* [$_{RC}$ **nok'** *lan   s-way-i*   ]
      CLF dog   CLF PROG A3-sleep-IV
      'the dog that is sleeping.'   (Chuj; cf. similar Shan examples above)

- Perhaps Chuj N.CLFs share the same syntactic position as Shan classifiers, and Chuj #.CLFs share the same syntactic position as Ch'ol classifiers:

(29)
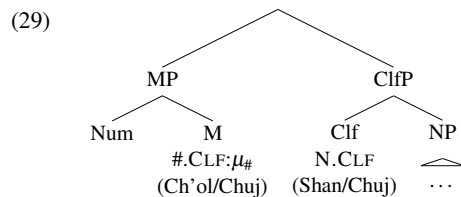
- The structure in (29) follows Bale et al. (2019) in positing a Measure phrase (MP) that contains the numeral and CLF-for-NUM.

- And follows Cheng & Sybesma (1999), Simpson (2005), a.o. in positing a Classifier Phrase (ClfP) that takes the nominal as complement.

Different semantics: But even though Chuj noun classifiers seem to pattern with Shan classifiers syntactically, they cannot fulfill the exact same semantic function.

- Crucially, unlike Shan classifiers, Chuj's *noun* classifiers are **never** required in presence of numerals (only Chuj's *numeral* classifiers are).

- Therefore, it seems that while Chuj has Ch'ol-style numeral classifiers, the noun classifier is semantically distinct from the Shan-style classifier.

What would it mean semantically for a language to have both types of classifiers?

Answer: The result would be semantically ill-formed. CLFs-for-Ns generate a set of atoms, so when the CLF+N combines with the numeral, there are no pluralities in the set to measure.

(30)
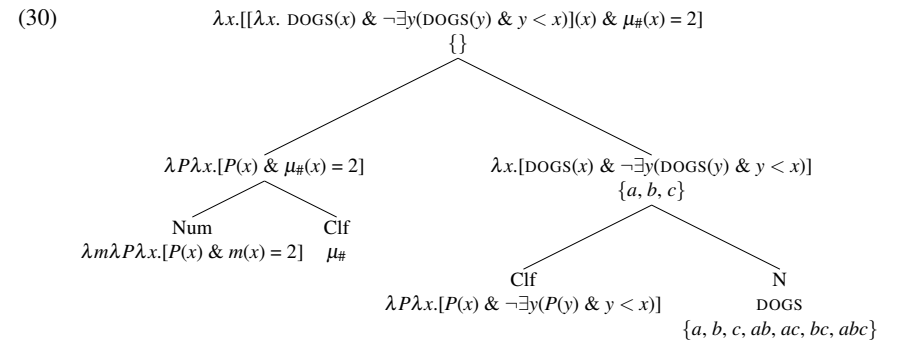
Result: We don't expect to find a language with the semantics of CLFs-for-NUMs (which measure pluralities) and CLFs-for-Ns (which atomize) at the same time.

But we might expect to find a language with exponents of both syntactic heads. Perhaps Chuj is this kind of language.

- The presuppositions associated with Chuj noun classifiers are shared among other cases of classifier-for-nouns across languages.

- For instance, it is well-known that Cantonese (Cheng & Sybesma 1999) and Vietnamese (Simpson 2005) classifiers (which are CLF-for-Ns) are associated with definiteness.

- This isn't surprising if Chuj noun classifiers and CLFs-for-Ns share a syntactic position.[6]

---

[6]Aikhenvald (2000) observed that in many Southeast Asian languages, noun classifiers appear to be a subtype of numeral classifiers, whereas in some Mayan languages, numeral and noun classifiers are separate lexical items. This fits with the current study.

5 CONCLUSIONS

In this talk, we reviewed two types of theories on the cross-linguistic status of classifiers, as described in Bale et al. 2019:

1. Classifier-for-numeral theories (Krifka 1995, Bale & Coon 2014, Bale et al. 2019, Hall 2019)

2. Classifier-for-noun theories (Chierchia 1998, Simpson & Ngo 2018, a.o.)

- We concluded, based on predictions that each theory makes and on the distribution of classifiers in three languages from two language families (Ch'ol, Chuj, Shan), that both theories are needed.

- There are at least two kinds of classifiers that mediate between numerals and nouns across languages: *classifiers-for-numerals* and *classifiers-for-nouns*.

- Drawing on previous work, we've provided some diagnostics to distinguish between these different kinds of classifiers.

- We also proposed that the semantics of numerals can vary across languages, and that this is connected to which classifier is needed:

  - *Classifiers-for-numerals* arise to saturate an extra argument of the numeral required to count (in other languages, this could be lexically encoded in the numeral); Such numerals measure (atomic and plural) entities denoted by the noun.

  - *Classifiers-for-nouns* arise because the numeral is set counting, and so the classifier is needed to extract the atoms out of the nominal predicate (and thus exclude plural entities).

REFERENCES

AIKHENVALD, ALEXANDRA Y. 2000. *Classifiers: A typology of noun categorization devices: A typology of noun categorization devices*. OUP Oxford.

BALE, ALAN, and JESSICA COON. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry*.

BALE, ALAN; JESSICA COON; and NICHOLÁS ARCOS. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa: a journal of general linguistics* 4.

BUENROSTRO, CRISTINA. 2013. *La voz en Chuj de San Mateo Ixtatán*. Mexico City: El Colegio de México dissertation.

BUENROSTRO, CRISTINA; JOSÉ CARMEN DÍAZ; and ROBERTO ZAVALA. 1989. Sistema de clasificación nominal del Chuj. *Memorias del segundo coloquio internacional de Mayistas*, vol. II. Mexico City: UNAM.

CHAMPOLLION, LUCAS, and MARIA ESIPOVA. 2018. Integrating compositional semantics and event semantics. Lecture notes for the NASSLLI 2018 course, June 23-29, 2018.

CHENG, LISA LAI-SHEN, and RINT SYBESMA. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30.509–542.

CHIERCHIA, GENNARO. 1998. Reference to kinds across language. *Natural Language Semantics* 6.339–405.

CRAIG, COLETTE GRINEVALD. 1986. Jacaltec noun classifiers. *Lingua* 70.241–284.

GARCÍA PABLO, GASPAR, and PASCUAL MARTÍN DOMINGO PASCUAL. 2007. *Stzolalil sloloni-spaxtini heb' chuj: Gramática descriptiva chuj*. Academia de Lenguas Mayas de Guatemala.

HALL, DAVID. 2019. Licensing D in classifier languages and "numeral blocking". *Definiteness across languages*, ed. by Ana Aguilar-Guevara, Julia Pozas Loyo, and Violata Vázquez-Rojas Maldonado, 221-257. Language Science Press.

HOPKINS, NICHOLAS. 2012. Noun classifiers of the Chuchumatán Mayan languages: A case of diffusion from Otomanguean. *International Journal of American Linguistics* 78.411–427.

HOPKINS, NICHOLAS A. 1970. Numeral classifiers in tzeltal, jacaltec and chuj (mayan). *Papers from the sixth regional meeting of the chicago linguistic society*, 23–35.

KRIFKA, MANFRED. 1995. Common nouns: A contrastive analysis of English and Chinese. *The generic book*, ed. by Gregory Carlson and Francis Jeffry Pelletier, 398–411. Chicago: University of Chicago Press.

LINK, GODEHARD. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. *Formal semantics: The essential readings*, 127–146.

NOMOTO, HIROKI. 2013. *Number in classifier languages*. University of Minnesota dissertation.

PIEDRASANTA, RUTH. 2009. *Los chuj, unidad y rupturas en su espacio*. Guatemala City, Guatemala: Amrar Editores.

ROYER, JUSTIN. to appear. Domain restriction and noun classifiers in Chuj. *Proceedings of the 49th annual meeting of the North East Linguistics Society*.

SHARVY, RICHARD. 1980. A more general theory of definite descriptions. *The philosophical review* 89.607–624.

SIMPSON, ANDREW. 2005. Classifiers and DP structure in Southeast Asia. *The Oxford handbook of comparative syntax*, 806–838.

SIMPSON, ANDREW, and BINH NGO. 2018. Classifier syntax in Vietnamese. *Journal of East Asian Linguistics* 27.211–246.

WILHELM, ANDREA. 2008. Bare nouns and number in Dëne Sułiné. *Natural Language Semantics* 16.39–68.

ZAVALA, ROBERTO. 2000. Multiple classifier sytems in Akatek (Mayan). *Systems of nominal classification*, ed. by Gunter Senft, 114–146. Cambridge: Cambridge University Press.
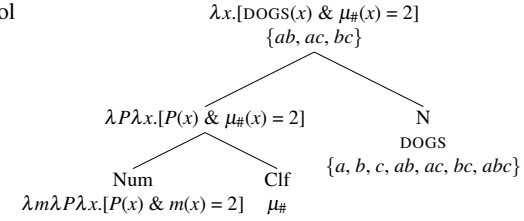
A DERIVATIONS

(31) Ch'ol

$\lambda x.[\text{DOGS}(x) \ \& \ \mu_\#(x) = 2]$
$\{ab, ac, bc\}$

$\lambda P \lambda x.[P(x) \ \& \ \mu_\#(x) = 2]$

N
DOGS
$\{a, b, c, ab, ac, bc, abc\}$

Num
$\lambda m \lambda P \lambda x.[P(x) \ \& \ m(x) = 2]$

Clf
$\mu_\#$

(32) Shan

$\lambda x. \ [\exists Y \ (x = \bigoplus Y \ \& \ Y \subseteq \{x : \text{DOGS}(x) \ \& \ \neg\exists y(\text{DOGS}(y) \ \& \ y < x)\}) \ \& \ |Y| = 2)]$
$\{ab, ac, bc\}$

Num
$\lambda P \lambda x.[\exists Y \ (x = \bigoplus Y \ \& \ Y \subseteq P \ \& \ |Y| = 2)]$

$\lambda x.[\text{DOGS}(x) \ \& \ \neg\exists y(\text{DOGS}(y) \ \& \ y < x)]$
$\{a, b, c\}$

Clf
$\lambda P \lambda x.[P(x) \ \& \ \neg\exists y(P(y) \ \& \ y < x)]$

N
DOGS
$\{a, b, c, ab, ac, bc, abc\}$

B CHUJ'S NUMERAL CLASSIFIERS AS CLASSIFIERS-FOR-NUMERALS

Chuj's numeral classifiers (#.CLF) pattern according to the classifier-for-numeral diagnostics established in §3.2 (i.e. like Ch'ol's classifiers).

- Prediction 1 (CLF-for-NUM): Obligatory when counting or referring directly to numeral:

(33) CONTEXT: Students are practicing counting.
ox-*(e'),   chanh-*(e'), hoy-*(e')
three-#.CLF, four-#.CLF, five-#.CLF
'3, 4, 5.'   (Chuj)

(34) CONTEXT: A teacher is pointing at the number three and says:
ha jun tik   ox-*(e').
TOP one DEM three-#.CLF
'This is three.'   (Chuj)

- Prediction 2 (CLF-for-N): Some numerals can't combine with a numeral classifier, including Mayan-based *jun* 'one' and numerals borrowed from Spanish:

(35) *jun-(*e')* ... *wentiyuno-(*e'), wentitres-(*e')*
one-#.CLF   21-#.CLF,   22-#.CLF
'1 ... 21, 22.'   (Chuj)