Proceedings of SALT 34: 713–733, 2024

Accounting for counting (crosslinguistically)*

Ushasi Banerjee
Yale University

Abstract This paper proposes a universal account of counting constructions. The observable variation in such constructions can be categorized into four possible systems resulting from parametric variation of two linguistic properties - NP[+/-pred] (whether the NP starts out as a predicate or a kind term) and Card/CL (whether cardinal and classifier heads are fused). The four systems show different patterns vis-à-vis overt occurrence of number-marking plurals (of the English type) and numeral classifiers - they are either obligatorily complementary, or obligatorily absent, or they obligatorily co-occur. The first two systems have been extensively studied, and the third has been argued to be impossible. This paper introduces novel data from Khasi, which exhibits such a system. Independent support for the proposed typology comes from languages where the cardinal and classifier form a complex morphological unit. In one such language, Ch'ol, numeral classifiers are not grammatically compatible with borrowed Spanish cardinals, supporting that they too are cardinal/classifier units but morphologically non-transparent.

Keywords: typology of counting strategies, Khasi, co-occurring plurals and classifiers, parametric variation, morphological complexity

1 Introduction

It has been long observed in classifier languages that nouns in counting constructions lack plural marking (Greenberg et al. 1963; Sanches & Slobin 1973; Aikhenvald 2000). This paper sheds light on the nature of the interaction between number

* Acknowledgment: The Khasi data presented in this paper have been taken from on-going joint work with Bianca Tara Faith Nongkinrih, a linguist at North East Hill University, India. The project is aimed at documenting the nominal system of Khasi, a linguistically understudied language and probing questions on definiteness, indefiniteness and genericity adopting the questionnaire developed in dayalhitch. This project would not have been possible without the support and guidance of my mentors Veneeta Dayal, Zoltan Szabo, and Ayesha Kidwai. I wish to thank the three anonymous SALT reviewers for their comments. Sincere gratitude for the for the consistent help and valuable discussions due to my amazing peers - Bianca Nongkinrih, Ka Fai Yip, Catarina Suarez, Comfort Ahenkorah, Yusuke Yagi. I also wish to thank the members of Yale Semantics Reading Group for their comments and feedback, in particular Lydia Newkirk for her generosity with time, Pauline Jackobson, Mitcho (Michael Yoshitaka Erlewine), Helen Koulidobrova for insightful discussions

marking plurals (henceforth #-PL) and classifiers (henceforth CL), drawing primary and novel data from Khasi and extends it to other systems of counting that have been discussed in the literature. Specifically, it investigates the systems that underlie counting constructions that cause crosslinguistic variation. As shown in (1) and (2), English employs the plural *-s* while Mandarin employs the classifier *ge* in counting constructions.

(1) a. *three student

b. three student**s**

(2) a. *san xuesheng
three student

b. san ge xuesheng
three CL student
'three students'
(Mandarin, Jiang 2020)

The theories that account for this pattern can be broadly divided into two. The primary claim for both is the same - that #-PLs and CLs ***do not co-occur***. One offers a syntactic explanation of the complementarity, arguing that the two functional heads occupy the same syntactic position (Borer 2005). The other theory offers a semantic account, claiming that kind terms are not compatible with the PL function (Chierchia 1998). Both approaches predict that languages either have plural marking, or they use numeral classifiers. While plural morphemes have been attested in CL languages, they exhibit special properties. For example, Mandarin *men* has been analyzed as the realization of the D head (Li 1999); Bangla *ra* has been analyzed as an associative plural (Biswas 2013); as an animacy classifier (Dayal 2014), as a classifier that turns singular kinds to plural kinds (Saha 2023). Some other cases include the plural *eet*/*aat* in Lebanese Arabic (Borer & Ouwayda 2010, 2021); *tul* in Korean, *tachi*/*tati* in Japanese (Nemoto 2005).

#-PL	CL	Languages
✔	✘	English, Hindi, Spanish
✘	✔	Mandarin, Bangla, Korean
✘	✘	Dëne Sųłiné, Yoruba
✔	✔	?

Table 1 Distribution of #-PL and CL and (non-exhaustive) corresponding languages

However, Khasi poses a challenge to the claim of complementarity between numeral classifiers and plurals. The plural *ki* in Khasi can and ***must*** co-occur with numeral classifiers; ***both*** the classifier morpheme and the plural morpheme are obligatory when counting plural entities (4). The fact that number marking plurals

and numeral classifiers can co-occur is important not just for explaining the data in Khasi[1] but also has theoretical implications.

(4) a. ar *(tylli) *(ki)-kot
two CL PL-book
'Two books.'

b. phra*(ngut) *(ki)-khynnah
eight CL PL-girl
'Eight girls.'

The question of markedness, when it comes to number marking, has been an important topic for debate. There are two dominant views that exist. One is that plurality is semantically contentful and predicates are pragmatically inferred to represent singularity in the absence of plural marking (Link et al. 1983). The other is the opposite - plural predicates are the unmarked form and appear whenever the singularity condition is not met (Krifka 1989, Sauerland 2003, Sauerland, Anderssen & Yatsushiro 2005). While there are justified arguments on both accounts, the question of number marking remains an unsettled issue. Neither singular nor plural nominals in Khasi are the 'unmarked'. They come in three forms - the unmarked, the plural marked, and the gender marked (which signals a singularity). Therefore, a predicate like *shynrang* 'boy' appears as a nominal in either the unmarked form, or in the singular form *u-shynrang* 'MASC-boy', or in the plural form *ki-shynrang* 'PL-boy'. The singular form comes packaged with gender information, while the plutal form specifically encodes number information. Based on this pattern, I argue that the plural morpheme in Khasi is a pluralizing function.

The role of a numeral classifier has also been debated upon. Under one view, the numeral and the classifier behave as a unit that composes with the nominal (Krifka 1995, Bale & Coon 2014, Bale, Coon, Arcos & López 2019). Another view considers nominals and classifiers form a unit which combines with the numeral (Chierchia 1998, Cheng & Sybesma 1999). A third view considers that the directionality of composition varies between languages, drawing empirical evidence from counting constructions in Ch'ol and Shan (Little, Moroney, Royer & Little 2022). In this paper, I argue that classifiers are universally required for the function of counting since they function as atomizers. In some languages they are morphologically overt while in others they are a part of a complex head Card/CL where the features of

1 This pattern is consistent across the three subcategories of Khasic languages - Khasi, Pnar (Anissa Lamare p.c.), War (Molly Varghese p.c.) as shown in (3) .

(3) a. hynru*(tylli) *(ki)-kot
two CL PL-book
'Two books.' (*Pnar*)

b. u *(rbe) *(i)-hun-tirme
two CL PL-child-male
'Two sons.' (*War*)

the cardinal and those of the classifier are bundled. The need for a classifier is facilitated by the internal composition of the nominal and the atomicity requirement of cardinals - that applies to both kind terms and plural predicates. The variability in terms of the presence/absence of overt classifiers and number marking is determined by certain parametric settings that include both semantic and syntactic features. The proposed typology can straightforwardly account for the long disputed co-occurrence of #-PL and CL as found in Khasi. This paper is organized as follows. §2 makes arguments for a *true* co-occurrence of classifiers and plurals in Khasi, demonstrating that both number marking and classifiers follow the typical properties found in number marking and classifier languages respectively. §3 proposes a new typology that integrates the co-occurrence of Khasi plurals and classifiers with other counting mechanisms that have been discussed in the literature and provides a compositional analysis of the four predicted language types. §4 discusses two independently attested linguistic phenomena that are predictions of the given typology and its theoretical assumptions - suffixed numerals and syntactic-semantic asymmetry in number features. Finally, §5 concludes with future directions for this project.

2 Obligatory co-occurrence of number marking plural and counting classifiers in Khasi

In this section, I claim that Khasi exhibits co-occurrence of individuating classifiers and a typical number-marking plural. I start by introducing relevant data regarding Khasi grammar - the three noun forms, number agreement with VP, and obligatory use of classifiers in counting constructions. Khasi is a Khasi-Palaungic language of the Austroasiatic family spoken predominantly in the Khasi hills of Meghalaya, India. Nouns in Khasi are morphologically marked for number and gender features in argument positions. Singular feminine nouns are marked with the morpheme *ka* (5a), masculine singular nouns with *u* (5b), and plural nouns are underspecified for gender and marked with the morpheme *ki* (5c).

(5) a. ka-sngi
SG.F-sun
'sun'

b. u-tiew-kulap
SG.M-flower-rose
'rose'

c. ki-tiew-kulap
PL-flower-rose
'roses'

The following examples show that Khasi exhibits subject-verb agreement. In (6a), the singular noun 'dog' and the VP 'barking' *both* must be marked with the masculine prefix *u*. Similarly, in its plural form, both the noun and the VP should be prefixed with the plural morpheme *ki*, as in (6b). The unmarked bare noun by

itself can make number-neutral reference, but its distribution is restricted to object positions, exemplified in (7) with the unmarked noun *kot* 'book'[2].

(6) a. **u**-ksew u-dang-wiar
MASC-dog MASC-PROG-bark
'The dog is barking.'

b. **ki**-ksew ki-dang-wiar
PL-dog PL-PROG-bark
'The dogs are barking.'

(7) a. *(ka)-kynthei *(ka)-thied
FEM-girl FEM-buy
kot
book
'The girl bought a book/books.'

b. *um ka long kaba
water FEM MOD that
donkam bha
necessary very
Intended: 'Water is essential for life'

As mentioned, both CL and #-PL are employed in counting constructions in Khasi(8).

(8) a. ar *(tylli) *(ki)-kot
two CL PL-book
'Two books.'

b. phra*(ngut) *(ki)-khynnah
eight CL PL-girl
'Eight girls.'

Crucially, Khasi cannot be called an ordinary number marking language given that it has numeral classifiers, nor can it be called an ordinary CL language because it has productive plural-marking. In the following I lay out three arguments in favor of the claim that the plural in Khasi is a number marking plural *and* its classifiers exhibit properties typical of CL languages.

2.1 Khasi *ki* is not a special plural such as Mandarin *men*

Several facts about Khasi grammar indicate that *ki* is a regular number-marking plural. Firstly, *ki* establishes *Agree* relations both within the nominal structure (9) and the VP (10-11).

(9) a. **ka**-ne **ka**-kynthei
FEM-DEM.PROX FEM-girl
'This girl.'

b. **ki**-ne **ki**-kynthei
PL-DEM.PROX PL-girl
'These girls.'

In (10), the plural noun shows overt agreement with the VP, marking the plural morpheme *ki* on both the noun and the VP. In (11) we see the same for the masculine

2 Unmarked bare nouns exhibit several properties typical of pseudo-incorporation when used as an argument; such as restriction to object position, lack of case-marking (Dayal 2011), but that will not be discussed separately in this paper

singular using the morpheme *u*. This system is morphologically analogous to Hindi (10b, 11b), a number marking language. Plural morphemes attested in classifier languages do not interact with Agree relations. They have very restricted distribution and impart 'special' meanings to the noun they modify, rather than patterning with #-PLs which function as inflectional morphemes (Wiltschko 2008). The plural *ki* patterns as a functional head rather than a modifier.

(10) a. **ki**-ksew **ki**-dang-wiar
PL-dog PL-PROG-bark
'The dogs are barking.'
(*Khasi)*

b. kutt-**e** bhõk rah-**e**
dog-PL bark PROG-PL
hãi
AUX.PL
'The dogs are barking.'
(*Hindi)*

(11) a. **u**-ksew
MASC-dog
u-dang-wiar
MASC-PROG-bark
'The dog is barking.'
(*Khasi)*

b. kutt-**a** bhõk
dog-MASC bark
rah-**a** hai
PROG-MASC AUX.SG
'The dog is barking.'
(*Hindi)*

Secondly, a robustly common feature among special plurals in classifier languages is that they facilitate *associative* uses (Nemoto 2005 for Japanese *tati* and Korean *tul*, Jiang 2020 for Mandarin *men*, a.o). In such cases, the plural can attach to proper names and create a group with the property 'x and associates of x', where x is the proper name. Associative uses are not available with Khasi *ki* as shown in (12). The only available reading of the construction is that a contextually determined set of girls all named Molly bought cars, such as 'the Mollys of the class'

(12) ki-Molly ki-thied-kali
PL-Molly PL-buy-car
Available: ✔'The Mollys bought cars.'
✘ 'Molly and her associates bought cars.'

Thirdly, *ki*-marked NPs are not instances of *pluralia tantum* (or fake mass nouns). Typical mass nouns in Khasi have very different properties than the plural forms. Mass nouns cannot be plural-marked, and can only combine with gender morphemes *ka-* or *u-*. If *ki*-marked NPs were instances of fake mass nouns, they would be expected to behave the same as regular mass nouns in the grammar, which they don't.

(13) a. *ki-um/ *ki-beer/
PL-water/ PL-beer/
*ki-phlang
PL-grass
Intended: 'Water'/ 'Beer'/ 'Grass'

b. ka-um/ u-beer/
FEM-water/ MASC-beer/
ka-phlang
FEM-grass
'Water'/ 'Beer'/ 'Grass'

These properties of the plural morpheme *ki* are typical of #-PLs. I claim that Khasi *ki* is in fact a true plural, and it does not exhibit the properties that have been reported of special plurals in CL languages.

2.2 Classifiers in Khasi

Classifiers in Khasi also exhibit typical properties, like ones found in Mandarin. There are two individuating classifiers - *tylli* which functions as a default (14a), and *ngut* which is specific to human nouns such as *girl*, *boy*, *woman*, *man* (14b). Along with these two, there are several group classifiers that combine with count nouns as shown in (15).

(14) a. saw **tylli** ki-bilor
four CL PL-bottle
'four bottles'

b. saw **ngut** ki-kynthei
four CL PL-girl
'four girls'

(15) a. nga thied lai **jur** ki-juti
I bought three CL_{pair} PL-shoe
'I bought three pairs of shoes.'

b. ki don lai **synduk** ki-kot
PL BE three CL_{box} PL-book
'There are three boxes of books.'

c. shi **kynkhun** ki-kynthei ki ieng
one CL_{group} PL-girl PL stand
'A/One group of girls were standing.'

Khasi also has mensural classifiers that combine with mass nouns (16). Thus, semantic features of the referent such as animacy or human/non human distinction plays a role in the choice of the classifier morpheme, a widely attested property of classifier languages.

(16) a. saw **bilor** u-beer
four CL_{bottle} SG.M-beer
'four bottles of beer'

b. lai **klas** ka-um
three CL_{glass} SG.F-water
'three glasses of water'

Another typical property of classifiers in Khasi is that they do not recur (Jiang 2020). That is, there can only be a single classifier used within a cardinal projection (17). This indicates that the classifier *tylli* is not a vacuous morpheme, and has a well-defined function in the grammar.

(17) a. *lai tylli synduk ki-kot
three CL_{indiv} CL_{box} FEM-book
b. *lai tylli tylli ki-kot
three CL_{indiv} CL_{indiv} FEM-book

To sum up, I have presented novel Khasi data that indicate that the plural *ki* in Khasi shows the same properties as #-PLs in languages like English, Hindi. I have also shown that classifiers in Khasi exhibit properties typical of CL languages such as Mandarin or Korean. In other words, Khasi exhibits co-occurrence of #-PL of the English/Hindi type and CL of the Mandarin/Korean type. This gives us an updated distribution of #-PLs and CLs as shown below.

#-PL	CL	Languages
✔	✘	English, Hindi, Spanish
✘	✔	Mandarin, Bangla, Korean
✘	✘	Dëne Sųłiné, Yoruba
✔	✔	Khasi

Table 2 Updated distribution of #-PL and CL and (non-exhaustive) corresponding languages

I conclude from this that the #-PL morpheme and CLs in Khasi indeed co-occur. This pattern cannot be accounted for by the current theories of plurality or classifiers. In fact, these theories hinge on the assumption that #-PLs and CLs cannot co-occur. That is, the implausibility of co-occurrence of #-PL and CL is not a bug but a *feature* of the existing theories, which motivates a revision. In the following section, I propose a new typology for counting constructions based on parametric settings that determine the variation in counting strategies.

3 A universal CLP analysis

I begin by introducing the parameters, which involve both semantic and syntactic features, that determine the form of counting constructions in a language, and discuss each in turn. Following that, I lay out the definitions that I assume for the relevant functional projections, and some theoretical assumptions.

3.1 Parametric variation

I propose that there are two parameters that determine the structure of counting constructions crosslinguistically - (i) **Nominal Mapping** (NP[±pred, ±arg]), (ii) **Morphological fusion** (Card/CL). The first parameter is due to Chierchia (1998), who argues that the semantic composition of an NP is dependent on the combination of features [±pred] and [±arg]. Languages set to the parameter **NP[+pred,±arg]** have nouns that originate as ordinary predicates of <e,t> type. This would cover both cases like English, Hindi, or Khasi (NP[+pred, +arg]) as well as French, Italian (NP[+pred, -arg]). On the other hand, languages set to the parameter **NP[-pred, +arg]** have nouns that donot originate as predicates. This would represent *typical* classifier languages like Cantonese, Korean, or Bangla. This parametric setting determines the structure of the nominal spine of the relevant language, which I spell out in the next section. The second parameter concerns the morphological output of the Card(inal) head and CL head. I posit that the Card and CL heads undergo *Fusion* in some languages and remain separate heads in others. Fusion refers to a post-syntactic PF operation which creates a single syntactic node for lexical insertion of separate heads, which results in suppletion (Noyer 1992, Halle, Marantz, Hale & Keyser 1993, Halle & Marantz 1994). In languages that fuse the Card and CL heads, there is cumulative exponence of the two heads at Vocabulary Insertion. Conversely, in languages that don't fuse these heads, the features of the Card and the CL remain to be in separate heads and therefore get separate exponents - *cardinal* and *classifier*. Different settings of these two parameters give us four possibilities for counting systems crosslinguistically as shown in (18) below.

(18) A *new* typology

parameters	NP[+pred, ±arg]	NP[-pred, +arg]
✘*Card/CLP*	(i) ✔#-PL; (ii) overt CLs Eg.***Khasi***	(i) ✘#-PL; (ii) overt CLs Eg.***Mandarin***
✔*Card/CLP*	(i) ✔#-PL Eg.***English***	(i) ✘#-PL Eg. ***Dëne Suliné***

In the following, I provide the definitions that have been adopted for cardinals (Card), plurals (#-PL), and classifiers (CL) in this analysis. Following that, I show the derivations for counting constructions in the four types of languages I have posited above.

3.2 Theoretical assumptions: definitions

I. Cardinals are modifiers

Following Ionin & Matushansky (2006), I argue that cardinals are modifiers that have an atomic requirement. That is, they need the nominal they combine with to strictly consist of atoms. A cardinal *n* is a function from a predicate *P* to a set of entities *x* such that *x* can be partitioned into *n* parts that each have the property *P* (19).

(19) **Definition of n_{card}**

a. $[\![n_{card}]\!] = \lambda P \lambda x . \exists S[\Pi(S)(x) \wedge |S| = n \wedge \forall s \in S\, P(s)]$

b. Π(S)(x) = 1 iff S is a cover of x, and ∀z,y ∈ S[z=y ∨¬∃a[a ⩽ z ∧ a ⩽ y]] (Forbidding that cells of the partition overlap ensures that no element is counted twice.)

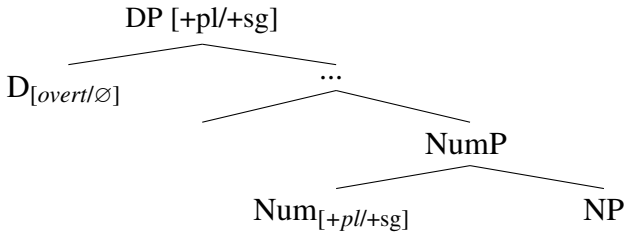
c. A set of individuals C is a cover of a plural individual X iff X is the sum of all members of C: ⊔C = X

(Ionin & Matushansky 2006)

II. Num introduces *i[+pl]* and *i[+sg]* features

NPs in number marking languages need to be determined for number - i[+pl] or i[+sg]. The Num head does exactly that (Kramer 2016). The [+pl] feature or [+sg] feature of the Num percolates up to DP, thereby facilitating agreement with the VP (20).

(20)



III. Two possible structures of CardP

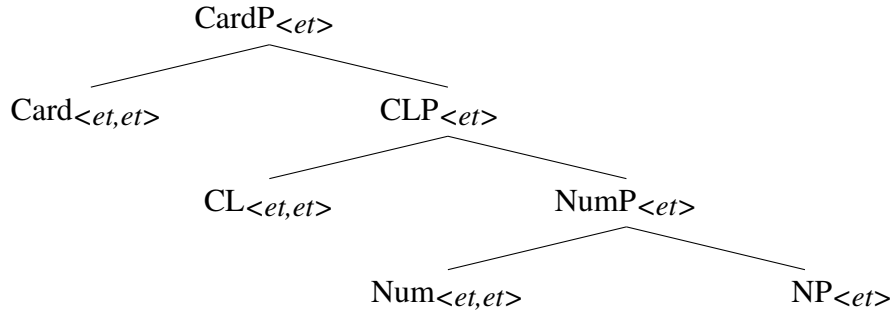
I argue that the setting of the Nominal Mapping parameter determines the structure of a CardP fragment in a language. NP[+pred, ±arg] languages show systematic number agreement. That indicates that the nominal spine consists of a Num head which carries number features i[+pl] and i[+sg]. NP[-pred,+arg] languages, on the other hand, do not show number agreement.

(21)

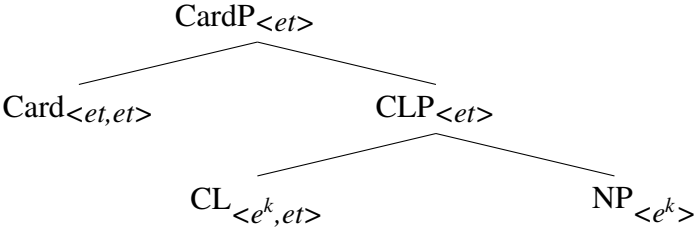
parameters	**NP[+pred, ±arg]**	**NP[-pred, +arg]**
✘*Card/CL*	(i) ✔#-PL; (ii) overt CLs Eg.***Khasi***	(i) ✘#-PL; (ii) overt CLs Eg.***Mandarin***
✔ *Card/CL*	(i) ✔#-PL Eg.***English***	(i) ✘#-PL Eg.***Dëne Sųłiné***

The following trees illustrate the two possible CardP structures that have been motivated above.

(22) NP[+pred] languages with NumP



(23) NP[-pred] languages without NumP



IV. Number marking plurals (#-PLs) are semantically contentful

I adopt the view that plural morphemes are not vacuous. A #-PL that occupies Num is interpreted as the *-operator following Link et al. (1983).

(24) **Definition of #-PL**

a. $[\![\text{\#-PL}]\!]^{g,c} = \lambda P.\ {}^*P$

b. ${}^*P(X) = 1$ iff there is a cover C of X with $P(x) = 1$ for every x in C

c. A set of individuals C is a Cover for X iff $\oplus C = X$

(adapted from Sauerland 2003:263)

The singular noun, on the other hand, is semantically null. It is an identity function over sets that carries a presupposition that the set be atomic (25). While #-PL carries interpretable [+pl] feature, #-SG carries interpretable [+sg] feature.

(25) **Definition of #-SG**

$[\![\text{ \#-SG }]\!] = \lambda P_{<e,t>} : \forall x[P(x) \rightarrow AT(x)].\ P_{<e,t>}$

V. #-PLs and CLs play different roles in grammar

CardP cannot combine directly with a plural-marked NP owing to the atomic requirement of cardinals (Ionin & Matushansky 2006). CLP is an intermediary projection which facilitates composition of the cardinal and the nominal. Following Borer (2005), I assume that CLPs are universal in counting constructions. However, I argue that they serve opposing purposes, drawing evidence from the obligatory co-occurrence of the two heads in Khasi. Number marking plurals have both syntactic and semantic significance. They are necessary for establishing plural agreement *and* they are functions that adds sums to predicates (Link et al. 1983). Individuating classifiers, on the other hand, are functions that atomize a predicate (Krifka 1995, Chierchia 1998, Bale, Gagnon & Khanjian 2010).

VI. Why add sums just to take them out?

In the system proposed here, counting constructions with cardinals ≥ 2 require adding sums to the NP set at NumP and taking them out subsequently at CLP, which is not an economic strategy. However, I argue that a different principle is at competition with Economy here. This principle favors a positive correspondence between syntactic and semantic features. NPs in number-marking languages need to be specified for the number features of the NP - i[+pl] or i[+sg], which can then undergo *Agree* with an u[+pl] or u[+sg] feature on the VP. Which means that #-PL serves two purposes - **(a)** it carries *agreement feature* [+pl] necessary for VP agreement and **(b)** it is interpreted as the *-operator which reflects *semantic plurality*. Crucially, there needs to be a correspondence between the two. That is, if a modified NP is semantically plural (has sum reference), it also carries a [+pl] feature.

VII. Classifiers are type flexible

As for classifiers, I argue that their function is the same across languages - whether they are number marking or *typical* classifier languages; they atomize number neutral predicates. However, their semantic type varies between the two language categories - they can be functions from either kinds (Chierchia 1998; Dayal 2012; Jiang 2012) or predicates Krifka 1995, Bale et al. 2019 to sets containing atoms; either <e^k,et> or <et,et>.

(26) **Definition of CL**

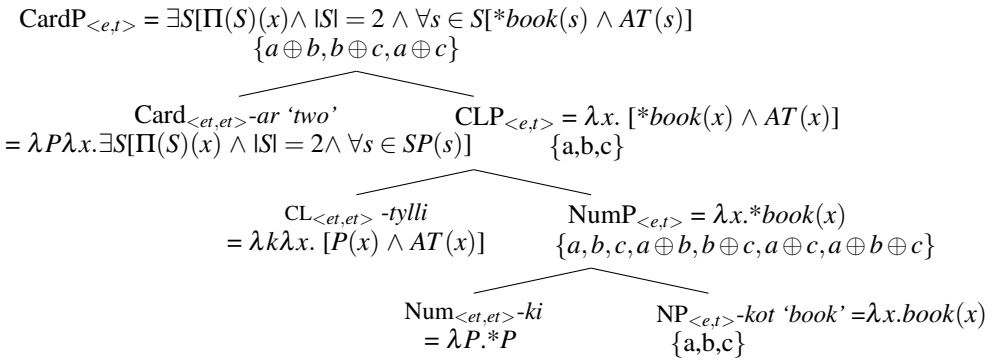
a. $[\![\mathbf{CL}_{<e^k,et>}]\!] = \lambda k \lambda x.\ [^{\cup}\mathrm{k(x)} \wedge AT(x)]$

b. $[\![\mathbf{CL}_{<et,et>}]\!] = \lambda P \lambda x.\ [\mathrm{P(x)} \wedge AT(x)\ ]$

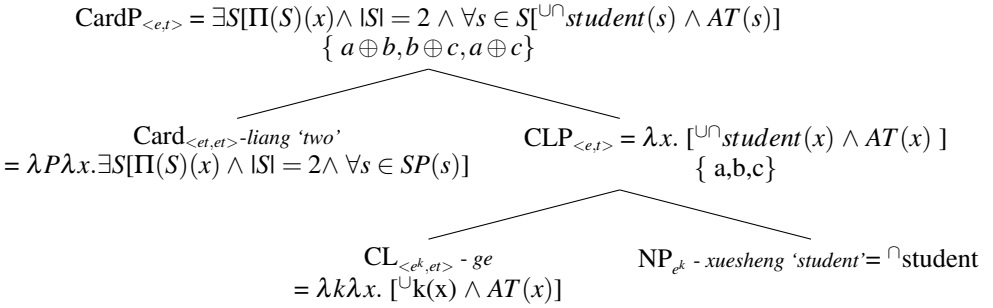
3.3 Deriving counting constructions

With these definitions in place, let's look at how counting constructions compose in the four types of languages - (i) Khasi which is set to **NP[+pred,±arg]** and ***✘Card/CLP***; (ii) Mandarin which is set to **NP[-pred, +arg]** and ***✘Card/CLP***; (iii) English which is set to **NP[+pred,±arg]** and ***✔Card/CLP***; and (iv) Dëne Sųłiné which is set to **NP[-pred, +arg]** and ***✔Card/CLP***.

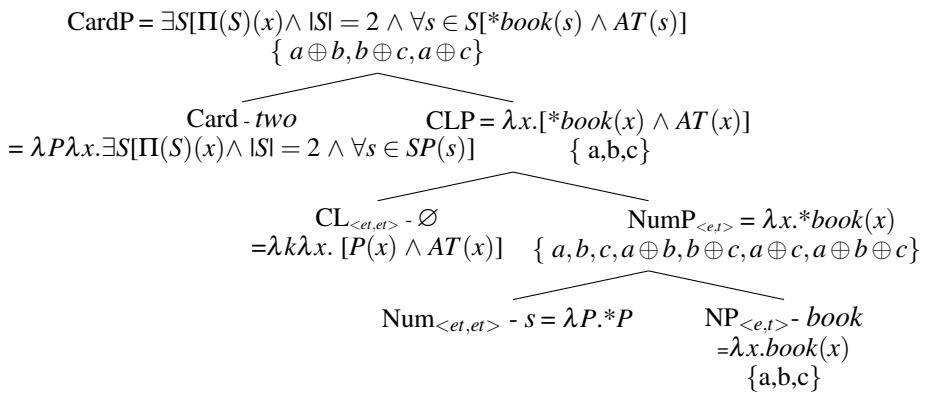
(27) The composition of *ar tylli ki-kot* 'two books' in Khasi



(28) The composition of *liang ge xuesheng* 'two students' in Mandarin



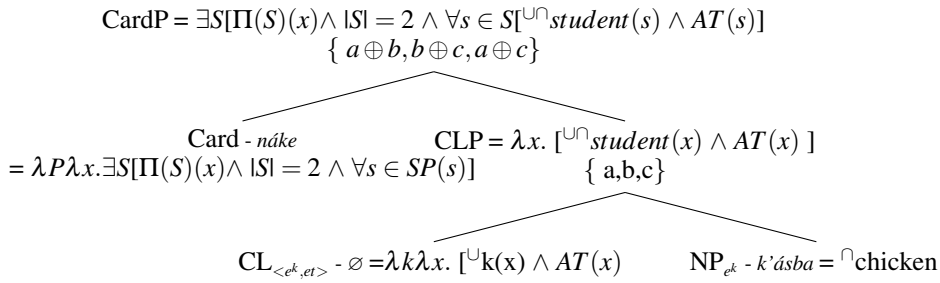
(29) a. The composition of *two books* in English



b. Vocabulary insertion

[Card: 2, CL: +] $\leftrightarrow$ *two*

(30) a. The composition of *náke k'ásba* 'two chickens' in Dëne Sųłiné



b. Vocabulary insertion

[Card: 2, CL: +] $\leftrightarrow$ *náke*

4 Some predictions

In what follows, I discuss the implication of the proposed typology, drawing evidence from linguistic data that have been attested in the literature. The first one concerns numeral constructions in certain languages that indicate *Fusion* of the Card and CL heads. The second one concerns singular agreement in counting constructions with cardinals above 2.

4.1 Evidence for Card/CL: suffixed numerals

Numerals in various languages exhibit morphological complexity. For example, Ch'ol numerals are always followed by a morpheme specific to the noun class it is modifying, as in (31). Little et al. (2022) argue that these are *numeral* classifiers since they exhibit a closer connection with numerals. Similar morphologically complex numerals have been attested in Mizo (32) and Yoruba (33).

(31) Complex numerals in Ch'ol

1	jum-p'ej	6	wäk-p'ej
2	cha'-p'ej	7	wuk-p'ej
3	ux-p'ej	8	waxäk-p'ej
4	chäm-p'ej	9	bolom-p'ej
5	jo'-p'ej	10	lujum-p'ej

(López 2009, Little et al. 2022)

(32) Complex numerals in Mizo

pa-khat	'one'
pa-nhiʔ	'two'
pa-thuml	'three'
pa-liil	'four'
pa-ngaal	'five'
pa-ruk	'six'

(Chhangte 1986)

(33) Complex numerals in Yoruba

	Base	m-form	Output	Gloss
a.	èjì	ḿ + èjì	méjì	'two'
b.	ẹ̀ta	ḿ + ẹ̀ta	mẹ́ta	'three'
c.	ẹ̀rin	m- + ẹ̀rin	mẹ́rin	'four'

(Ajíbóyè & Déchaine 2004, Ajiboye 2010)

Krifka (1995) has argued (for English) that the denotation of a numeral has a predicativising *and* an atom-accessing function (a "built-in classifier"). Extending this notion, Wilhelm (2008) has claimed that the difference between English or Dëne Sųłiné and Mandarin lies in the semantics of numerals in the respective languages, drawing evidence from data such as in (34).

(34)

basic	*human*
ʔįłághe 'one'	ʔįłághį 'one'
náke 'two'	nádëne 'two'
tághe 'three'	tághį/ tanį 'three'
dįghį 'four'	dįghį/ dįnį 'four'
sǫlághe 'five'	sǫlághį 'five'

(Cook 2004, Wilhelm 2008)

Phenomena such as these, across language families, show that the Card and CL heads interact morphologically. This prediction is borne out within the proposed typology - we would expect to find morphologically complex numerals in languages that bundle Card and CL. Additional evidence comes from Ch'ol where counting can involve

either native cardinals or ones borrowed from Spanish. While numeral classifiers are obligatory when counting with native cardinals, they are ungrammatical when used with Spanish ones (Bale & Coon 2014). This suggests that Spanish cardinals are also morphologically complex, though not transparently, and what gets borrowed in Ch'ol is the cardinal-classifier complex morpheme.

In light of this data, 'classifier-for-numeral' theories argue that the noun combines with a complex constituent formed by Card and CL (Little et al. 2022). Adopting this view would straightforwardly annul the Economy violation discussed earlier in §3.2. However, there are two possible confounds of that view. First, even in languages like Ch'ol, the noun class affects the choice of classifier that is attached to the numeral; *-p'ej* for inanimate and generic nouns, *kojty* for animals, *tyikil* for humans, among others (López 2009). Similarly, Dëne Sųłiné doesn't show overt classifiers in default cases, but counting [+ human] nouns involves one (34). Thus, even though the primary function of an atomizing classifier is to facilitate counting, it is sensitive to, at least, the animacy features of the noun. This indicates that the CL layer must exist *between* NP and CardP to allow this interaction. Second, Ionin & Matushansky (2006) comment on the challenges of assuming that cardinals occupy the Spec position on an extended NP spine considering languages like Russian, or Inari Sami where lower cardinals assign case on sister nouns (35) and (36) (see the following section for a discussion of the implications of this data on the syntax proposed in this paper).

(35) a. četyre šagà
four step-PAUC
'Four steps'

b. šest' šagov
six step-GEN.PL
'Six steps'
(Russian, Ionin & Matushansky 2006)

(36) a. kyehti/ kulmâ/ nelji/ vittâ/ kuttâ päärni
two/ three/ four/ five/ six child-ACC.SG
'two/ three/ four/ five/ six children'

b. čiččâm/ kávci/ ovce/ love/ ohtnubáloh/ kyehtnubáloh/ čyeti... pärnid
seven/ eight/ nine/ ten/ eleven/ twelve/ hundred child-PART.SG
'seven/ eight/ nine/ ten/ eleven/ twelve/ hundred children'
(Inari Sami, Ionin & Matushansky 2006)

Thus enough crosslinguistic data indicates that the principle of Economy is overruled by various syntactic and semantic requirements of both Card and NP/NumP.

4.2 Semantic plurality, syntactic singularity

The proposed typology also provides some insights on counting systems found in languages such as Hungarian, Finnish or Welsh where nouns show singular marking in counting constructions (Ionin & Matushansky 2006).

(37) a. Yhdeksän omena-a puto-si maa-han
nine-NOM apple-PART.SG fall-PAST-3SG earth-ILL
'Nine apples fell to earth.' (*Finnish*; Nelson & Toivonen 2000)

b. három gyerek/ *három gyerekek
three child/ three child.PL
'three children' (*Hungarian*; de Swart & Farkas 2010)

Arguments in Hungarian/Finnish type languages are *semantically plural* but *syntactically singular*. de Swart & Farkas (2010) have proposed an OT approach to deal with the crosslinguistic distinction between Hungarian and English counting strategies. I suggest that there are two markedness constraints that are at competition:

- (*REDUNT) which favors the least number of operations if composition is already possible.
- (*M-SYN/SEM) which favors syntactic plurality (+pl) of the NP to match with semantic plurality (sum reference) of the NP.

As argued before, the mismatched values of syntactic and semantic number information is a principle that overrides the Economy principle in languages like English or Khasi. The Hungarian/Finnish type of counting system shows that there can be variation in which principle is more valuable to a language. The distinction between Hungarian/Finnish and English/Khasi falls out of such a variation (in spirit of the unidirectional OT analysis in de Swart & Farkas (2010) for Hungarian plurals). Specifically, it results from (*REDUNT) being ranked higher than (*M-SYN/SEM) in the former. Classifiers do not get employed in the former type because the NumP denotation is already atomic. This is supported by the data in Russian and Inari Sami presented previously in (35) and (36). A syntactic-semantic asymmetry emerges; the noun remains to be marked as singular irrespective of the cardinal - whether one or higher. Case assignment requires structural immediacy between two syntactic nodes. The absence of an intermediate classifier projection in languages of this category straightforwardly explains the ability of Card to assign case in these languages. For brevity, detailed arguments for this claim is left for future work.

4.3 Khasi in the typology of definiteness

Khasi informs the typology of DP projection crosslinguistically, in that the *Nominal Mapping Parameter* does not interact with the existence of a DP in a language. There are two strategies of forming definite arguments in the system developed in Chierchia (1998) et. seq. The first is through an overt D/ ✔definite articles; and (b) ι type-shifting/ ✘definite articles (Chierchia (1998)). The first strategy is found in languages like English or Italian, which are set to the parameters NP[+pred, +arg] and NP[+pred, +arg] respectively. The second strategy is found in languages like Mandarin, which is set to NP[-pred, +arg], as well as Hindi or Russian, which are set to NP[+pred,+arg]. Subsequently, **?** (for Cantonese); Dayal 2012 (for Bangla) argue that some NP[-pred, +arg] languages also project D. They lack determiner articles but employ *CL to D movement* (Cantonese) or, *NP to Spec DP movement* (Bangla). Lastly, Jiang 2012, 2020 shows that determiner articles (overt D) can also be found in some NP[-pred, +arg] languages such as Nuosu Yi, expanding the typology further as in (38).

(38) i. Overt D/ ✔def. articles

- *English*, *Italian* (NP[+pred,±arg]); *Nyosu Yi* (NP[-pred, +arg])

ii. ι type-shifting/ ✘def. articles

- *Mandarin* (NP[-pred,+arg]); *Hindi*, *Russian* (NP[+pred, +arg])

iii. Covert D/ ✘def. articles

- *Cantonese*, *Bangla* (NP[-pred, +arg])

Khasi lacks determiner articles, but exhibits *NumP to Spec DP movement* in definite constructions with cardinals ≥ 2 (39b). Which implies that there's a covert D.

(39) a. ar tylli ki-kot
two CL PL-book
'Two books.'

b. ki-kot ar tylli
PL-book two CL
'The two books.'

We can infer that the *Nominal Mapping Parameter* does not show any correlation with strategies to form definite arguments, since both strategies have been attested in all parametrical possibilities.

(40)

parameters	**NP[+pred, ±arg]**	**NP[-pred, +arg]**
✘*Card/CL*	***Khasi*** - covert D	① ***Cantonese*** - covert D ② ***Nyosu Yi*** - overt D ③ ***Mandarin*** - ɪ type-shift
✔ *Card/CL*	① ***English*** - overt D ② ***Hindi*** - ɪ type-shift	***Dëne Sųłiné*** - ɪ type-shift

5 Conclusion and further directions

I have proposed four arguments in this paper. First, the long disputed co-occurrence of classifiers and number-marking plurals is brought to question by the nominal system in Khasi. Second, not only are classifiers and number marking plurals not complementary, individuating classifiers are *universally* required to mediate between cardinals and number neutral predicates. Although this has been claimed in the literature, Khasi provides empirical evidence for it. Third, a new typology integrating two parametric settings **Nominal Mapping** (NP[±pred, ±arg]) and **Morphological fusion** (Card/CL) can account for the variation in counting strategies crosslinguistically; languages with overt CLs and standard plural-marking (like Khasi); languages with overt CLs and special plurals (like Mandarin); languages with covert CLs and standard plural-marking (like English); languages with covert CLs and special plurals (like Dëne Sųłiné), all of which have been attested in the literature. Fourth, positive evidence for the typology comes from several unrelated languages (Ch'ol, Yoruba, Mizo, Dëne Sųłiné)) where fused Card/CL heads are morphologically transparent.

The proposed typology provides a concrete picture of how *individuating* classifiers and number marking interact crosslinguistically. Languages also employ more complex classifiers whose function is not simply tied to counting constructions, but modifying NPs with other information - plurality, definiteness, indefiniteness as has been observed for Bangla *gulo* (Dayal 2012, 2014), Mandarin *men* (Li 1999), among others. Moreover, certain languages have been claimed to have 'optional' classifiers such as Turkish *tane* (Sağ-Parvardeh 2019), Farsi *tā* (Gebhardt 2009), Western Armenian *ler* (Borer 2005, Bale & Coon 2014). Placing these languages within the typology is not straightforward. However, recent work has argued that these are lexical exponents of Measure Phrases that are inherently associated with numerals and not exponents of CL heads in these languages. Rather than functioning as regular classifiers, these morphemes interact with definite and indefinite expressions; for languages that do not have a definite determiner, they contribute to indefinite interpretations of numeral constructions (Sağ 2024). While a significant amount of work has been done on individuating classifiers, much more remains to be known

about plural classifiers or 'optional' classifiers. Understanding the role of these morphemes in grammar is fundamental to our understanding of numeral/counting constructions and the relationship between plurality and classifiers across languages. Lastly, the typology does not inform us about the low frequency of languages like Khasi which transparently show the co-occurrence of #-PL and CLs. It is possible that this is a result of insufficient documentation, but it remains an open question.

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. OUP Oxford.

Ajiboye, Oladiipo. 2010. Plural strategies in yoruba doi:https://doi.org/10.32473/sal.v39i2.107283.

Ajíbóyè, ládiíp & R Déchaine. 2004. The syntax and semantics of yorùbáduplicative constructions. In *35th Annual Conference on African Linguistics, Harvard University*, .

Bale, Alan & Jessica Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry* 45(4). 695–707. doi:https://doi.org/10.1162/ling_a_00170.

Bale, Alan, Jessica Coon, Nicholás Arcos & Nicolás Arcos López. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa: a journal of general linguistics* 4(1). doi:https://doi.org/10.5334/gjgl.752.

Bale, Alan Clinton, Michaël Gagnon & Hrayr Khanjian. 2010. Cross-linguistic representations of numerals and number marking. In *Semantics and linguistic theory*, 582–598.

Biswas, Priyanka. 2013. Plurality in a classifier language: Two types of plurals in bangla. In *Generative linguists of the Old World in Asia (GLOW in Asia)*, 1–14.

Borer, Hagit. 2005. *Structuring sense: Volume 1: In name only*, vol. 1. Oxford University Press.

Borer, Hagit & Sarah Ouwayda. 2010. Men and their apples: Dividing plural and agreement plural. *Handout of a talk presented at GLOW in Asia VIII* .

Borer, Hagit & Sarah Ouwayda. 2021. Divide and counter. *Things and Stuff: The Semantics of the Count-Mass Distinction* 115. doi:https://doi.org/10.1017/9781108937979.006.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of np. *Linguistic inquiry* 30(4). 509–542. doi:https://doi.org/10.1162/002438999554192.

Chhangte, Lalnunthangi. 1986. *A Preliminary Grammar of the Mizo Language (Tibeto-Burman)*. The University of Texas at Arlington.

Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural language semantics* 6(4). 339–405.

Cook, Eung-Do. 2004. *A grammar of Dëne Sułiné (Chipewyan)*, vol. 17. Algonquian and Iroquoian Linguistics.

Dayal, Veneeta. 2011. Bare noun phrases. *Semantics: an international handbook of natural language meaning. Berlin: Walter de Gruyter* 149–76.

Dayal, Veneeta. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24(2). 195–226.

Dayal, Veneeta. 2014. Bangla plural classifiers. *Language and Linguistics* 15(1). 47–87. doi:https://doi.org/10.1177/1606822x13506151.

Gebhardt, Lewis. 2009. *Numeral classifiers and the structure of DP*: Northwestern University PhD dissertation.

Greenberg, Joseph H et al. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language* 2. 73–113.

Halle, Morris & Alec Marantz. 1994. Some key features of distributed morphology. *MIT working papers in linguistics* 21(275). 88.

Halle, Morris, Alec Marantz, Kenneth Hale & Samuel Jay Keyser. 1993. Distributed morphology and the pieces of inflection. *1993* 111–176.

Ionin, Tania & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of semantics* 23(4). 315–360. doi:https://doi.org/10.1093/jos/ffl006.

Jiang, Li. 2012. *Nominal arguments and language variation*: Harvard University, Cambridge, MA PhD dissertation.

Jiang, Li. 2020. *Nominal Arguments and Language Variation*. Oxford University Press.

Kramer, Ruth. 2016. A split analysis of plurality: Number in amharic. *Linguistic Inquiry* 47(3). 527–559.

Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event doi:https://doi.org/10.1515/9783110877335-005.

Krifka, Manfred. 1995. 11 common nouns: A contrastive analysis of chinese and english. *The generic book* 398.

Li, Yen-hui Audrey. 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 75–99.

Link, Godehard et al. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. *Formal semantics: The essential readings* 127. 147. doi:https://doi.org/10.1515/9783110852820.302.

Little, Carol-Rose, Mary Moroney, Justin Royer & Carol Rose Little. 2022. Classifiers can be for numerals or nouns: Two strategies for numeral modification. *Glossa: a journal of general linguistics* 7(1). doi:https://doi.org/10.16995/glossa.8437.

López, Nicolás Arcos. 2009. Los clasificadores numerales y las clases nominales en ch'ol. *Recurso impreso, recurso electrónico* .

Nelson, Diane & Ida Toivonen. 2000. *Counting and the grammar: Case and numerals in Inari Saami*. Citeseer.

Nemoto, Naoko. 2005. On mass denotations of bare nouns in japanese and korean doi:https://doi.org/10.1515/ling.2005.43.2.383.

Noyer, Robert Rolf. 1992. *Features, positions and affixes in autonomous morphological structure*: Massachusetts Institute of Technology PhD dissertation.

Sağ, Yağmur. 2024. On the (non) optionality of the turkish classifier tane. *Natural Language & Linguistic Theory* 1–58. doi:https://doi.org/10.1007/s11049-024-09626-4.

Sağ-Parvardeh, Yağmur. 2019. *The semantics of number marking: Reference to kinds, counting, and optional classifiers*: Rutgers The State University of New Jersey, School of Graduate Studies PhD dissertation.

Saha, Ankana. 2023. Reference to kinds: the perspective from Bangla. In *West Coast Conference on Formal Linguistics (WCCFL 41)*, .

Sanches, Mary & Linda Slobin. 1973. Numeral classifiers and plural marking: An implicational universal. *Working Papers on Language Universals* 11. 1–22.

Sauerland, Uli. 2003. A new semantics for number. In *Semantics and linguistic theory*, 258–275.

Sauerland, Uli, Jan Anderssen & Kazuko Yatsushiro. 2005. The plural is semantically unmarked. *Linguistic evidence: Empirical, theoretical, and computational perspectives* 413–434. doi:https://doi.org/10.1515/9783110197549.413.

de Swart, Henriëtte & Donka Farkas. 2010. The semantics and pragmatics of plurals. *Semantics and pragmatics* 3. 6–1. doi:https://doi.org/10.3765/sp.3.6.

Wilhelm, Andrea. 2008. Bare nouns and number in dëne sųłiné. *Natural Language Semantics* 16. 39–68. doi:https://doi.org/10.1007/s11050-007-9024-9.

Wiltschko, Martina. 2008. The syntax of non-inflectional plural marking. *Natural language & linguistic theory* 26(3). 639–694.

Ushasi Banerjee
Dow Hall, 370 Temple St
Yale University
New Haven, CT 06511
ushasi.banerjee@yale.edu