Classifiers, group counting and the internal structure of number concept expressions

Lars Dalen[1], Jakob Majdič[1] , Marcin Wągiel[1,2,3]
[1]Masaryk Univesrity in Brno, [2]Universit of Wrocław, [3]ZAS Berlin

Most research on numeral classifiers has been focused on individual sortal classifiers (e.g. Borer 2005, Bale et al. 2019, Chierchia 2010). We hence provide data on classifiers dedicated to counting *groups* of individuals (Krifka 1995). Our cross-linguistic comparison of numeral constructions in (i) unit counting, (ii) group counting and (iii) abstract uses will reveal a pattern that supports an analysis on which number concepts are internally complex and derived in the syntax.

Background. Group classifiers are attested in multiple classifier languages. Below we present group classifier constructions in Mandarin Chinese (1) (Zhang 2013), Uzbek (Levy-Forsythe & Kagan 2022), Bwamu (3) (Hölzl & Bognana 2024) and Yucatec Maya (4) (Blaha Pfeiler & Skopeteas 2024). Complex group numerals have been reported, e.g. in Finnish (5) (Hurford 1998) and Czech (6) (Grimm & Dočekal 2021). Group classifiers often differ syntactically from individual classifiers (Chao 1968, Li 2011, Zhang 2013). While the syntactic status of individual classifiers is debated, it is typically assumed that **group classifiers form a constituent with the numeral** (Zhang 2013), which **suggests a parallel** structure and semantics **to complex group numerals**. Investigating the environments in which different types of numeral constructions occur therefore informs our understanding of the components of various numeral constructions.

Patterns. *Unit and group counting.* The pattern exhibited in the examples below show that in Mandarin and Uzbek only the choice of classifier differentiates between unit counting (**Num+A**) and group counting (**Num+B**). Bwamu, Yucatec, Finnish and Czech exhibit a bimorphemic numeral construction in unit counting (**Num+A**) and a trimorphemic one in group counting (**Num+A+C**). In Bwamu and Yucatec Maya CLF$_1$ appears both in unit and group counting.

(1)	a.	sān **tóu** niú three CLF$_1$ cow ‘three cows’	b.	sān **qún** niú three CLF$_2$ cow ‘three herds of cows’	(Mandarin)
(2)	a.	ikki **dona** kitob two CLF$_1$ book ‘two books’	b.	ikki **guruh** bola two CLF$_2$ child ‘two groups of children’	(Uzbek)
(3)	a.	livru-ßɛ **bo** tĩ book-PL CLF$_1$ three ‘three books’	b.	hã-ßɛ **gnũ-nɛ bo** tĩ woman-PL CLF$_2$-PL CLF$_1$ three ‘three groups of women’	(Bwamu)
(4)	a.	jum-**p’éel** kib one-CLF$_1$ candle ‘one candle’	b.	jun-**p’èel cháach** iib one-CLF$_1$ CLF$_2$ bean ‘one bunch of beans’	(Yucatec Maya)
(5)	a.	vii-**si** saapasta five boot.PTV ‘five-SFX$_1$ boots’	b.	vii-**de-t** saappaat five-SFX$_1$-SFX$_2$ boots ‘five pairs of boots’	(Finnish)
(6)	a.	tř-**i** karty three-SFX cards ‘three cards’	b.	tr-**oj-e** karty three-SFX$_1$-SFX$_2$ cards ‘three sets of cards’	(Czech)

Abstract uses. In other environments, numeral constructions are not involved in any counting operation at all, but instead semantically contribute only an abstract number concept/a numerical value (e.g. in arithmetic statements, countdowns, reference to numbers *qua* numbers). This contrast is reflected linguistically: across all languages we are aware of, abstract numeral constructions occur *without a noun*, even if the semantics of the noun is near inert, (7)–(8) (Rothstein 2017).

(7) Four (*things) times two (*things) is . . . (8) My favorite number is one (*thing).

Our data show that numeral constructions denoting abstract number concepts differ across languages in interesting ways: in Mandarin and Uzbek abstract uses are incompatible with classifiers; i.e. the numeral construction in abstract uses is *monomorphemic* (9)–(10). By contrast, Bwamu and Yucatec pattern with

Finnish and Czech, where abstract numeral constructions are *bimorphemic*. In Bwamu and Yucatec abstract numeral constructions we see the same classifier (CLF_1) as in the counting constructions above (11)–(14).

(9)	sān	[three]	(Mandarin)	(10)	ikki	[two]	(Uzbek)
(11)	**bo** na	[CLF_1 four]	(Bwamu)	(12)	jum-**p'éel**	[one-CLF_1]	(Yucatec)
(13)	vii-**si**	[five-SFX_1]	(Finnish)	(14)	tř-**i**	[three-SFX_1]	(Czech)

The set of languages with bimorphemic numeral constructions in abstract uses precisely corresponds to the set of languages that behave differently than Mandarin and Uzbek with respect to counting constructions. The pattern exhibited by Mandarin and Uzbek across all three environments is given in (15), whereas (16) illustrates the other pattern. These data present strong evidence that **number concepts are internally complex and derived syntactically** (Wągiel & Caha 2021). In languages exhibiting Pattern 1, the internally complex structure is realized by a portmanteau form that spells out multiple syntactico-semantic components (**Num**), while in languages exhibiting Pattern 2 the complexity is reflected overtly (**Num+A**).

(15)	Abstract use:	**Num** (Pattern 1)		(16)	Abstract use:	**Num + A** (Pattern 2)
	Unit counting:	**Num + A**			Unit counting:	**Num + A**
	Group counting:	**Num + B**			Group counting:	**Num + A + B**

Proposal. To account for the discovered patterns, we build on the syntactico-semantic system by Wągiel & Caha (2021), which assumes three basic meaning components. SCALE denotes a lexically encoded closed interval, i.e., a set of numbers. NUM denotes a function from intervals to numbers: it takes SCALE and yields the greatest value from that set. CL_μ represents a set of related heads all of which introduce counting semantics, i.e., they shift a number into a quantifying device, which allows for counting some kind of entities. We assume two such heads: CL_{OBJ} encodes the # measure function for counting object-level individuals (17-a) and CL_{GRP} encodes ||| (17-b) for counting groups. The arithmetical function of numerals is represented in (18), whereas (19) gives the general template for the quantifying function, where μ is a variable over the relevant measure functions. Based on the discovered meaning/form correspondences, we propose the lexicalization tables in **Table 1** for unit counting expressions and in **Table 2** for group counting expressions. The tables capture the meaning-form correspondences across the investigated languages.

(17) a. $\forall n_n \forall P_{\langle e,t\rangle} \forall x_e[\#(P)(x) = n \leftrightarrow \exists Y_{\langle e,t\rangle}[Y \in \text{PAR}_c(x) \ \& \ \forall y_e[y \in Y \rightarrow y \in P \ \& \ |Y| = n]]$

b. $\forall n_n \forall P_{\langle e,t\rangle} \forall x_e[|||(P)(x) = n \leftrightarrow \exists Y_{\langle e,t\rangle}[Y \in \text{PAR}_c(x) \ \& \ \forall y_e[y \in Y \rightarrow y \in {}^*P \ \& \ |Y| = n]]$

(where $\text{PAR}_c(x)$ is the set of salient partitions of x in context c;
a partition of any x_e is any set S of non-overlapping individuals st. $\bigsqcup S = x$)

(18) $[_{\text{NUMP}_n}\ [\text{NUM}_{\langle\langle n,t\rangle,n\rangle}\ \lambda P_{\langle n,t\rangle}[\text{MAX}(P)]]\ [\text{SCALE}_{3\langle n,t\rangle}\ \lambda n_n[0 \leq n \leq 3]]]$

(19) $[_{\text{CL}_\mu\text{P}_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle}}\ [\text{CL}_{\mu\ \langle n,\langle\langle e,t\rangle,\langle e,t\rangle\rangle\rangle}\ \lambda n_n \lambda P_{\langle e,t\rangle} \lambda x_e[{}^*P(x) \wedge \#(P)(x) = n]\ /\ \lambda n_n \lambda P_{\langle e,t\rangle} \lambda x_e[{}^*P(x) \wedge |||(P)(x) = n]]\ [_{\text{NUMP}_n}\ [\text{NUM}_{\langle\langle n,t\rangle,n\rangle}\ \lambda P_{\langle n,t\rangle}[\text{MAX}(P)]]\ [\text{SCALE}_{3\langle n,t\rangle}\ \lambda n_n[0 \leq n \leq 3]]]]$

Table 1.

	SCALE	NUM	CL_{OBJ}
Mandarin	sān		tóu
Uzbek	ikki		dona
Bwamu	tĩ	bo	
Yucatec Maya	jum	p'èel	
Czech	tř	i	
Finnish	vii	si	

Table 2.

	SCALE	NUM	CL_{GRP}
Mandarin	sān		qún
Uzbek	ikki		guruh
Bwamu	tĩ	bo	gnũnɛ
Bwamu	jum	p'èel	cháach
Czech	tr	oj	e
Finnish	vii	de	t

References. • Bale, A. et al. (2019) *Classifiers, partitions, and measurements*. • Blaha Pfeiler, B. & S. Skopeteas (2024) *Numeral classifiers in Yucatec Maya*. • Borer, H. (2005) *Structuring Sense I: In Name Only*. • Chao, Y. R. (1968) *A grammar of spoken Chinese.* • Chierchia, G. (2010) *Mass nouns, vagueness and semantic variation*. • Grimm, S. & M. Dočekal (2021) *Counting aggregates, groups and kinds*. • Hölzl, A. & M. Bognana (2024) *Classifiers in Bwamu*. • Hurford, J. R. (1998) *The interaction between numerals and nouns.* • Krifka, M. (1995) *Common nouns: A contrastive analysis of Chinese and English*. • Levy-Forsythe, Z. & O. Kagan (2022) *Classifiers and the mass-count distinction in Uzbek*. • Li, X.-P. (2011)

On the semantics of classifiers in Chinese. • Rothstein, S. (2017) *Semantics for counting and measuring*. • Wągiel, M. & P. Caha (2021) *Complex simplex numerals*. • Zhang, N. N. (2013) *Classifier structures in Mandarin Chinese*.