Decomposing cardinal numerals: Cross-linguistic evidence

Introduction. It is commonly assumed that basic cardinal numerals such as English *five* are simplex expressions. In this paper, we explore cross-linguistic marking patterns suggesting that cardinals in fact lexicalize complex syntactic and semantic structures. We propose a unified morpho-semantic account for the typological variation in form and meaning of cardinals. We argue that it is possible to identify cross-linguistically stable semantic ingredients, which compositionally provide the attested types of numerals. We adopt the framework of Nanosyntax (Starke 2009) as a model of morphology which, when applied to the semantic primitives we propose, delivers the relevant marking patterns. Our model is broadly based on the idea that the meaning components are uniformly structured across languages, and they must all be pronounced, though languages differ in how they pronounce them. All cardinals share an underlying scale of natural numbers but differ in a number of operations subsequently applied to that scale. The key evidence comes from Austronesian.

Two functions. Cardinals can have two different functions which we will refer to as **abstract counting**, i.e., reference to a number concept as in (1), and **object counting**, i.e., quantification over individuals as in (2) (e.g., Bultinck 2005). In order to account for the difference, the mainstream approach is to derive the abstract-counting meaning by a special shifting operation (Rothstein 2017) or a null suffix (Ionin & Matushansky 2018) whose function is to yield the number corresponding to an object-counting cardinal.

(1) Ten divided by **five** equals two.

(2) **five** roses

The asymmetry. Interestingly, languages often distinguish formally between the two flavors (Hurford 1998, Fassi Fehri 2018). For instance, in Japanese a form used to refer to mathematical entities, see (3), differs from the one conveying the cardinality of a particular set of objects in (4). Though both expressions contain a common core, e.g., *go*, the object-counting function requires an additional morpheme, e.g., *ko*, usually referred to as a classifier (a general classifier in this case). Cross-linguistically, this asymmetry is a relatively frequent pattern (e.g., Mandarin, Vietnamese, Thai, Bhojpuri, Adang) which suggests that the abstract-counting function is basic whereas the object-counting function is derived from it both morphologically and semantically.

(3) juu waru **go-(*ko)**-wa ni-da.
ten divide.by five-CL-TOP two-COP
'Ten divided by five is two.'

(4) **go-*(ko)**-no ringo
five-CL-GEN apple
'five apples'

Note that the distribution of the classifier is not due to the syntactic position of the numeral, i.e., present in an NP-internal use and absent in an NP-external use. In predicate position, the numeral is not an attribute to a noun, yet it has to be accompanied by a classifier (Sudo 2016). It is, thus, there to mark object counting.

Complex abstract-counting numerals. However, there is cross-linguistic evidence that abstract-counting numerals can be morphologically complex. In Hawaiian (Oceanic), the numerals 1-9 consist of a root, e.g., *-lua*, always accompanied with an obligatory additional morpheme, i.e., the classifying prefix *e-* as in (6) (Elbert & Pukui 1979). What makes this case different from (3)–(4) is that Hawaiian numeral roots can never appear bare. Thus, the prefix is required even in an abstract-counting form such as (5) (Pukui & Elbert 1986).

(5) ‘e-lua ā me **‘e-lua**, ‘e-hā.
PRFX-two and with PRFX-two PRFX-four
'Two plus two is four.'

(6) **‘e-lua** i‘a
PRFX-two fish
'two fish'

In Vera'a (Oceanic), the numeral always includes the obligatory prefix *vō-* along with the root. This is the case when reference to a number concept is made as in (7). Interestingly, in object-counting contexts an additional marker, the so-called ligature *ne*, is required showing evidence of a tripartite structure (Schnell 2011).

(7) vēvē-gi ne lukun ēn naw, din ēn **vō-’ōl**.
mother-3SG ASP count ART wave reach ART NBR-three
'His mother counted the waves reaching (the number) three.'

(8) ēn woqe’enge **ne vō-ru**
ART tree LIG NBR-two
'two trees'

Typology. To classify morphological patterns, we employ the following terminology. **Symmetric numerals** have one form for both functions, **asymmetric numerals** employ an additional morpheme in the object-counting function. **Simplex numerals** are monomorphemic in the abstract-counting function, while **complex numerals** consist of two morphemes. The properties intersect: English 5 is simplex symmetric (1)–(2), Hawaiian 2 is complex symmetric (5)–(6), Japanese 5 is simplex asymmetric (3)–(4), and Vera'a 2 is complex asymmetric (7)–(8). The categories are not properties of languages, but rather of a particular numeral. Individual languages may contain different classes of numerals as in Chol and Mi'gmaq (Bale & Coon 2014).

Universal semantic features. To account for the data, we propose the ingredients in (9)–(11) to be part of the universal underlying structure of numerals. We assume three syntactic heads and standard function application. The meaning of $\text{S}{\small\text{CALE}}_m$ is a closed interval, e.g., the set of natural numbers in $[0,5]$. The key intuition is that numerals are at their core scalar entities (Nouwen 2016). Following the set-theoretic characterization of natural numbers, we take $\text{S}{\small\text{CALE}}_m$ to be a complex set-theoretic object constructed syntactically by Merge (Chomsky 2008, Watanabe 2017). This motivates $\text{S}{\small\text{CALE}}_m$ being closed between 0 (corresponding to the empty set) and the lexically encoded upper bound m, e.g., 5. NUM takes a set of integers and yields the greatest number from that set, i.e., forges a proper name of an arithmetic concept. Finally, CL takes a number and returns a predicate modifier equipped with the pluralization operation * (Link 1983) and the measure function $\#(P)$ (Krifka 1989). Its goal is, thus, to form an expression that can be used for counting objects.

(9) $[\![\text{S}{\small\text{CALE}}_m]\!]_{\langle n,t\rangle} = \lambda n_n[0 \leq n \leq m]$

(10) $[\![\text{N}{\small\text{UM}}]\!]_{\langle\langle n,t\rangle,n\rangle} = \lambda P_{\langle n,t\rangle}[\text{MAX}(P)]$

(11) $[\![\text{C}{\small\text{L}}]\!]_{\langle n,\langle\langle e,t\rangle,\langle e,t\rangle\rangle\rangle} = \lambda n_n \lambda P_{\langle e,t\rangle} \lambda x_e[{*}P(x) \wedge \#(P)(x) = n]$

Composition. Combining (9)–(11) in a compositional fashion leads to the structures in (12) and (14). For $\text{S}{\small\text{CALE}}_5$, the tree in (12) will be interpreted as (13), i.e., application of MAX turns the interval $[0,5]$ into the integer 5. The result is, thus, of type n and can be used as a name of a number concept. On the other hand, (14) is an object-counting modifier interpreted, e.g., as (15). We obtain an expression which, when applied to a predicate, yields a set of pluralities of entities that have the relevant property and whose cardinality equals 5

(12) [NUM SCALE$_m$] ABSTR.COUNT

(13) $[\![(12)]\!] = 5$

(14) [CL [NUM SCALE$_m$]] OBJ.COUNT

(15) $[\![(14)]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e[{*}P(x) \wedge \#(P)(x) = 5]$

Lexicalization. To account for the morphological patterns, we adopt the view that lexical entries link morphemes to potentially complex syntactic/semantic structures. Following Starke (2009), we assume that the Superset Principle allows a given morpheme to pronounce *any sub-constituent* contained in its lexical entry. For instance, a lexical entry such as (16) can also pronounce (17) since this structure is its sub-constituent. To derive particular morpheme orderings we use the spellout driven movement technology (Starke 2018).

(16) [CL [NUM SCALE$_m$]]

(17) [NUM SCALE$_m$]

Analysis. The proposed system is able to derive the attested variation by treating different types of numerals as lexicalizations of different structures derived from the universal semantic components, see Table below. Simple symmetric numerals, e.g., English 5, are stored as complete structures pronouncing all the three heads, which allows them to cover both the abstract- and object-counting function. Simple asymmetric numerals lexicalize only the abstract-counting meaning, and thus require additional morphology in order to be able to be used as modifiers, e.g., a classifier in the case of Japanese 5. In complex symmetric numerals like Hawaiian 2, the root is stored as SCALE$_m$ while an additional affix is a portmanteau for CL and NUM. Finally, in complex asymmetric numerals such as Vera'a 2 each morpheme pronounces one of the three heads.

ABSTRACT			OBJECT		
SCALE	NUM		SCALE	NUM	CL
five		ENG 5	*five*		
go		JPN 5	*go*		*ko*
lua	*e*	HAW 2	*lua*	*e*	
ruō	*vō*	VER 2	*ruō*	*vō*	*ne*

References. Bale & Coon (2014) *Classifiers are for numerals* • Bultinck (2005) *Numerous meanings* • Chomsky (2008) *On phases* • Elbert & Pukui (1979) *Hawaiian grammar* • Fassi Fehri (2018) *Constructing feminine to mean* • Hurford (1998) *The interaction between numerals and nouns* • Ionin & Matushansky (2018) *Cardinals* • Krifka (1989) *Nominal reference, temporal constitution and quantification* • Krifka (1995) *Common nouns* • Link (1983) *The logical analysis of plurals and mass terms* • Nouwen (2016) *Making sense of the spatial metaphor for number* • Pukui & Elbert (1986) *Hawaiian dictionary* • Rothstein (2017) *Semantics for counting and measuring* • Schnell (2011) *A grammar of Vera'a* • Starke (2009) *Nanosyntax* • Starke (2018) *Complex left branches, spellout, and prefixes* • Sudo (2016) *The Semantic role of classifiers in Japanese* • Watanabe (2017) *Natural language and natural number*