Universal semantic features and the typology of numerals

Pavel Caha & Marcin Wągiel, Masaryk University in Brno

Introduction. The paper proposes a unified morpho-semantic account for the typological variation in form and meaning of cardinal numerals across languages. In particular, we investigate the morphological marking of different types of cardinals and argue that despite an apparent morphological chaos, 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 et seq.) as a model of morphology which, when applied to the semantic structures we propose, delivers the relevant marking patterns. The model we develop 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 asymmetry. Cardinals can have different functions including what we will refer to as ABSTRACT COUNTING, i.e., reference to a number concept, and OBJECT COUNTING, i.e., quantification over individuals (e.g., Bultinck 2005, Rothstein 2017). Interestingly, languages often distinguish formally between the two flavors (Hurford 1998). For instance, in Japanese a form used to refer to mathematical entities, see (1-a), differs from the one conveying the cardinality of a particular set of objects in (1-b) (Sudo 2016). Though both expressions contain a common core, e.g., *yon*, the object-counting function requires an additional morpheme, e.g., *ko* or *rin*, usually referred to as a classifier. Such an asymmetry is a cross-linguistically relatively frequent pattern which suggests that the abstract-counting function is basic whereas the object-counting function is derived from it both morphologically and semantically.

(1) a. ni tasu ni-wa **yon-(*ko)**-da
two plus two-TOP four-CL-COP
'Two plus two is four.'

b. **yon-*(rin)**-no hana
four-CL-GEN flower
'four flowers'

Symmetric numerals. In a number of languages, however, we observe no such asymmetry as in (1). For instance, in English both functions are expressed by the same formal exponent, see (2), suggesting that the bare numeral itself incorporates a classifier semantics (Krifka 1995). In other words, the form of *four* is ambiguous. In one use, it is semantically equivalent to *yon-rin*, in another use to *yon*.

(2) a. Two plus two is **four**.

b. **four** roses

Importantly, the (a)symmetry is not a property of a language as a whole, but rather of a particular numeral since languages such as Chol and Mi'gmaq have both types of cardinals (Bale & Coon 2014).

Inverse numerals. The most intriguing morphological facts come from Arabic. In this language, abstract counting is expressed by a morphologically more complex form than object counting, see (3) (Fassi Fehri 2018). This pattern seemingly implicates a reverse asymmetry, i.e., that compared to the object-counting function the abstract-counting function has some extra meaning which needs to be introduced by an additional morpheme, e.g., a gender marker. However, admitting this would jeopardize a morpho-semantic explanation of the widespread asymmetry illustrated in (1) as well as any unified typology of numerals. What we need to capture is the fact that Arabic exists, but that it is very rare.

(3) a. **ṯaalaṯ-*(at)**-un tusawii ʔiṯnayni zaʔid waaḥid
three-FEM-NOM equals two plus one
'Three equals two plus one.'

b. **ṯaalaṯ-(*at)**-u banaatin
three-FEM-NOM girls
'three girls'

Universal semantic features. In order to account for the data, we propose the ingredients in (4)–(6) to be part of the universal underlying structure of numerals. We assume three syntactic heads and the standard function application operation. SCALE is a lower bounded scale, e.g., a set of natural numbers in the interval $[4, \infty)$. NUM (for 'number') takes a set of integers and yields the smallest number from that set, i.e., forges a proper name of a mathematical entity. Finally, CL (for 'classifier') 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 actual objects.

(4) $[\![\text{SCALE}]\!]_{\langle n,t\rangle} = \lambda m_n[m \geq n]$ (5) $[\![\text{NUM}]\!]_{\langle\langle n,t\rangle,n\rangle} = \lambda P_{\langle n,t\rangle}[\text{MIN}(P)]$

(6) $[\![\text{CL}]\!]_{\langle n,\langle\langle e,t\rangle,\langle e,t\rangle\rangle\rangle} = \lambda n_n \lambda P_{\langle e,t\rangle} \lambda x_e[{}^*P(x) \wedge \#(P)(x) = n]$

Composition. Combining the ingredients introduced above in a compositional fashion leads to the structures in (7)–(8). For instance, (7-a) can be interpreted as (7-b), i.e., due to application of MIN the interval $[4, \infty)$ is turned into the integer 4. The resulting expression is, thus, of type n and can be used as a name of a number concept. On the other hand, (8-a) is an object-counting modifier interpreted, e.g., as (8-b). After the number slot in (6) is saturated by 4, 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 4.

(7) a. [NUM SCALE] ABSTR.COUNT
b. $[\![\text{(7-a)}]\!] = 4$

(8) a. [CL [NUM SCALE]] OBJ.COUNT
b. $[\![\text{(8-a)}]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e[{}^*P(x) \wedge \#(P)(x) = 4]$

The non-terminal lexicalization model. 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 (9-a) can also pronounce the structure in (9b) since this structure is its sub-constituent. Furthermore, we adopt the Elsewhere Principle, which states that when multiple items match a particular semantic structure, the more specific one, i.e., having fewer superfluous features, is chosen (Kiparsky 1973). Finally, we assume that there are no cardinals pronouncing only [SCALE], but we will independently justify its relevance based on the morphological evidence from Czech and Vurës (Malau 2016) as well as from suppletive forms and semantics of ordinals and multipliers.

(9) a. [CL [NUM SCALE]] b. [NUM SCALE]

Typology. The proposed system is able to derive all the attested variation by treating different types of numerals as lexicalizations of different structures derived from the universal semantic components, see Table below. Symmetric numerals are stored as complete structures pronouncing all the three heads, which allows them to cover both the abstract-counting and the object-counting function, e.g., English *four*. A special case of symmetry is represented by idiosyncratic numerals, which have suppletive forms for the two functions, e.g., Maltese *tnejn* ~ *żewġ* (the choice of *tnejn* as the abstract-counting form is due to the Elsewhere Condition). On the other hand, 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 *yon*. The fact that the inverse pattern is scarce is because it can only arise in a very particular configuration. Specifically, a numeral needs to be stored simply as SCALE, NUM needs to have an overt exponent, and [CL NUM] needs to be lexicalized as a null morpheme. As a result, abstract- and object-counting numerals are spelled out according to the inverse pattern. Finally, the system predicts two more very rare types of numerals PRED_1 (inverse numerals with β overtly realized) and PRED_2 (numerals with two affixes).

ABSTRACT			OBJECT		
SCALE	NUM		SCALE	NUM	CL
four		ENG 4	*four*		
tnejn		MLT 2	*żewġ*		
yon		JPN 4	*yon*		*rin*
ṯaalaṯ	*at*	ARA 3	*ṯaalaṯ*	Ø	
X	α	PRED_1	X	β	
X	α	PRED_2	X	α	β

The potential candidates for PRED_1 and PRED_2 are Abkhaz numerals 2–10 (Hewitt 1979, Chirikba 2010) and numerals in languages that allow for classifier stacking such as Akatek (Zavala 2000, Aikhenvald 2000), respectively.

References. Aikhenvald (2000) *Classifiers* • Bale & Coon (2014) *Classifiers are for numerals, not for nouns* • Bultinck (2005) *Numerous meanings* • Chirikba (2010) *Abkhaz* • Fassi Fehri (2018) *Constructing feminine to mean* • Hewitt (1979) • *Abkhaz* • Hurford (1998) *The interaction between numerals and nouns* • Kiparsky (1973) *'Elsewhere' in phonology* • Krifka (1989) *Nominal reference, temporal constitution and quantification in event semantics* • Krifka (1995) *Common nouns* • Link (1983) *The logical analysis of plurals and mass terms* • Malau (2016) *A grammar of Vurës, Vanuatu* • Rothstein (2017) *Semantics for counting and measuring* • Starke (2009) *Nanosyntax* • Sudo (2016) *The Semantic role of classifiers in Japanese* • Zavala (2000) *Multiple classifier systems in Akatek (Mayan)*