Gendered numerals in Slavic, Arabic and Abkhaz

Marcin Wągiel & Pavel Caha (Masaryk University in Brno)

Introduction. Recent research reveals a non-trivial relationship between grammatical gender and quantification (Arsenijević 2016, Fassi Fehri 2018). In this paper, we explore the interaction between different uses of cardinals and gender marking patterns in Slavic, Arabic and Abkhaz. We propose a unified morpho-semantic account for the typological variation in form and meaning. We adopt Nanosyntax as a model of morphology which, when applied to the semantic primitives we propose, delivers the relevant marking patterns. The 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.

Two functions of cardinals. Cardinals have at least two different functions: ABSTRACT COUNTING (AC), used in contexts like (1), and OBJECT COUNTING (OC), as in (2) (Bultinck 2005, Rothstein 2017). In a number of languages, a single form, e.g., English *five*, can be used to express both functions. However, many languages distinguish the two functions morphologically (Hurford 1998). 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 OC function requires an additional morpheme, e.g., *ko*, usually referred to as a classifier (a general classifier in this case). Such an asymmetry is a cross-linguistically relatively frequent pattern which suggests that the AC function is basic whereas the OC function is derived from it both morphologically and semantically.

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

(2) **five** cats

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

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

Interaction with gender. Interestingly, gender marking on cardinals is often correlated with the AC/OC distinction. Let us first consider the non-virile/virile marking (i.e., +MASC +HUMAN vs. everything else) in Slavic. For instance, in Bulgarian the numerals 2–10 agree with the noun in virility. The unmarked non-virile form *pet* ('five') is incompatible with virile NPs, see (5), whereas the marked virile form *petima* ('five') is bad with non-virile NPs, see (6) (Cinque & Krapova 2007). Crucially, only the unmarked form *pet* can appear in an AC context, see (7). The same pattern is also attested in Polish and Slovak (Wągiel 2018).

(5) **pet-(*ima)** ženi
five-V women.NV
'five women'

(6) **pet-*(ima)** mâže
five-V men.V
'five men'

(7) Deset deleno na **pet-(*ima)** e dva.
ten divided on five-V is two
'Ten divided by five is two.'

However, this is not the only way gender marking can interact with the AC/OC distinction. A seemingly inverse pattern is attested in Standard Arabic. This language is known for the gender mismatch between cardinals 3–10 and modified NPs (so-called 'gender polarity'). As demonstrated in (8)–(9), when the numeral combines with a feminine NP, it has the unmarked masculine form, but when combined with a masculine NP, the numeral requires the feminine suffix *-at*. Crucially, only the marked feminine form, i.e., *ṯalaaṯat* ('three'), can be used as an AC expression, see (10)–(11) (Fassi Fehri 2018).

(8) **ṯalaaṯ-(*at)**-u banaat-in
three-F-NOM girls.F-GEN
'three girls'

(9) **ṯalaaṯ-*(at)**-u ʔawlaad-in
three-F-NOM boys.M-GEN
'three boys'

(10) **ṯalaaṯ-at**-un t-usawii ʔitnayni zaʔid waaḥid.
three-F-NOM F-equals two plus one
'Three equals two plus one.'

(11) ***ṯalaaṯ-at**-un y-usawii ʔitnayni zaʔid waaḥid.
three-NOM M-equals two plus one

The final pattern concerns cardinals in Abkhaz (Northwest Caucasian). In the Abkhaz gender system, nouns are grammatically classified as human vs. non-human and, accordingly, the numerals 2–10 occur in two forms

both of which are morphologically marked, see (12)–(13) (Hewitt 1979, Chirikba 2003). Yet, only one of the two marked forms, i.e., the non-human form, can be used in an AC environment, see (14)–(15).

(12) a-c'°a-k°a **aa-bá** / ***aa-j°ə́(k')**
ART-apple.NH-PL eight-NH eight-H
'eight apples'

(13) a-č'k°'ən-c°a **aa-j°ə́(k') / *aa-bá**
ART-boy.H-PL eight-H eight-NH
'eight boys'

(14) x°ba jacəwc'ar xpa jəq'alojt' **aa-bá**.
five if-you-add three becomes eight-NH
'Five plus three equals eight.'

(15) *x°ba jacəwc'ar xpa jəq'alojt' **aa-j°ə́(k')**.
five if-you-add three becomes eight-H

In the examined cases, both morphologically marked and unmarked cardinals have the OC meaning. In Bulgarian, only the unmarked cardinals whereas in Arabic only the marked cardinals have the AC function. In Abkhaz, there are two marked forms but only one has the AC meaning. Our aim is to explain this variation.

Semantic features. We propose three universal syntactic heads in (16)–(18). SCALE_m is a closed interval, e.g., the set of natural numbers in $[0, 5]$. NUM takes a set of integers and yields the greatest number from that set, i.e., forges a proper name of a number concept. CL takes a number and returns a counting device, i.e., a predicate modifier equipped with the pluralization operation * (Link 1983) and the measure function $\#(P)$ (Krifka 1989). In addition to CL, the numeral may also include additional features, i.e., semantically interpreted gender heads, that introduce special presuppositions concerning the nature of the referents of the modified NP (cf. Sudo 2016), e.g., the contribution of ANIM in (19) is that animate individuals are required.

(16) $[\![\text{SCALE}_m]\!]_{\langle n,t\rangle} = \lambda n_n[0 \leq n \leq m]$

(17) $[\![\text{NUM}]\!]_{\langle\langle n,t\rangle,n\rangle} = \lambda P_{\langle n,t\rangle}[\text{MAX}(P)]$

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

(19) $[\![[\text{ANIM } [\text{CL } [\text{NUM SCALE}_m]]]]\!]_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle} = \lambda P_{\langle e,t\rangle} \lambda x_e : \text{ANIMATE}(x)[^*P(x) \wedge \#(P)(x) = m]$

Composition. Combining the ingredients introduced above yields complex structures. In the AC structure (20), the application of MAX turns the interval $[0, 5]$ into the integer 5. The resulting expression is, thus, of type n and can be used as a name of a number concept. On the other hand, (22) is an OC modifier augmented with a rich gender structure. It is an expression which, when applied to a predicate denoting virile entities, yields a set of pluralities of such entities that have the relevant property and whose cardinality is 5, see (23).

(20) $[\text{NUM SCALE}_5]$ AC

(21) $[\![(20)]\!] = 5$

(22) $[\text{VIR } [\text{HUM } [\text{ANIM } [\text{CL } [\text{NUM SCALE}_5]]]]]$ OC

(23) $[\![(22)]\!] = \lambda P_{\langle e,t\rangle} \lambda x_e : \text{VIRILE}(x)[^*P(x) \wedge \#(P)(x) = 5]$

Lexicalization. To account for the morphological patterns, we adopt the view that lexical entries link morphemes to non-trivial syntactic structures. Following Starke (2009), the Superset Principle (SP) allows a given morpheme to pronounce *any sub-constituent* contained in its entry. For instance, (22) can also pronounce (20) since this structure is its sub-constituent. Furthermore, we adopt the Elsewhere Condition (EC), which states that when multiple items match a particular semantic structure, the more specific one, i.e., having fewer superfluous features, is selected (Kiparsky 1973).

NUMERAL	SCALE	NUM	CL	ANIM	V/H/F
Bulgarian 5	*pet*				*ima*
Abkhaz 8	*aa*	*bá*			–
		j°ə́(ḱ)			
Arabic 3	*ṯalaaṯ*	*at*			–
		ø			

Analysis. The proposed system can derive the attested variation by treating different types of numerals as lexicalizations of different structures derived from the universal semantic components. Bulgarian *pet* pronounces SCALE, NUM, CL and lower gender heads, whereas *-ima* lexicalizes only the highest gender head, i.e., VIR. Due to the SP, *pet* can be used both in non-virile OC and AC contexts; *petima* is restricted to virile OC. On the other hand, Abkhaz *aa* pronounces only SCALE, and thus cannot appear bare either in AC, or in OC. The suffixes *-bá* and *-j°ə́(ḱ)* are both portmanteau forms allowing for OC. The only difference is that *-j°ə́(ḱ)* is bigger since it also includes the HUM head. Thus, due to the EC only *aabá* can be used in an AC context. Finally, in Arabic *ṯalaaṯ* is also stored as SCALE and *-at* is a portmanteau allowing for OC. To

account for the inverse asymmetry, we propose there is also a null morpheme pronouncing in addition the FEM head. In AC, it looses competition due to the EC, and thus *ṯalaaṯat* needs to be used as the AC form.