Universal Semantic Features and the Typology of Cardinal Numerals[1]

Marcin Wągiel[ab] & Pavel Caha[a]

[a]Masaryk University in Brno, [b]University of Warsaw

marcin.wagiel@phil.muni.cz; pavel.caha@phil.muni.cz

Abstract: The paper proposes a unified morpho-semantic account for the typological variation in the form and meaning of cardinal numerals. In particular, we investigate the morphological marking of two different types of cardinals and 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 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 the number of operations subsequently applied to that scale.

1. Introduction

In this paper, we investigate two different functions of cardinal numerals. We call the first function OBJECT COUNTING. It amounts to quantification over individuals denoted by the modified noun in terms of their number. For instance, the examples in (1) indicate sets of pluralities of entities whose cardinality equals 5.

(1) a. **five** roses
     b. the **five** cats

The second function will be referred to as ABSTRACT COUNTING. In this use, cardinals simply denote a number concept in an arithmetic environment (e.g., Bultinck 2005, Rothstein 2013, 2017). For example, in (2a), *five* is not used to quantify over entities; it rather designates a

numeric value. Similarly, in (2b), the numeral denotes an abstract mathematical entity of which the property of being prime is predicated.

(2)   a. Two and **five** make seven.
       b. **Five** is prime.

So far, most research has focused on the object-counting use, and the abstract-counting function remains understudied (but see Rothstein 2017, Wągiel to appear). We aim to contribute to this area of research by asking the following question: What is the relationship between object-counting and abstract-counting uses of numerals? The answer we offer is that object-counting numerals are both syntactically and semantically derived from abstract-counting numerals.

The paper is organized as follows. In Section 2, we discuss the syntactic and semantic differences between abstract counting and object counting. In Section 3, we show that the two different types of numerals also show different morphological shapes across languages. In Section 4, we provide the essentials for the compositional analysis of both functions. Section 5 demonstrates how our system accounts for the typology of attested meaning-form correspondences. Section 6 extends the approach by proposing how gender interacts with object-counting numerals and Section 7 concludes.

2.   Object counting vs. abstract counting

The distinction between the object- and abstract-counting numerals is not merely a conceptual distinction. It is also reflected in different grammatical distribution. The differential distribution reflects the fact that the two kinds of numerals are objects of a distinct type.

Specifically, we understand number concepts to be linguistic/semantic objects corresponding to abstract numbers.[2] Because of this, they have different properties than pluralities of individuals (Rothstein 2017). For instance, we can attribute to number concepts special properties such as being a Fibonacci number, see (3a). (3b) shows that when number concepts are compared, the dimension of comparison is based on their relative ordering. Furthermore, as witnessed by (3c), there are special grammatical constructions calling for numeric arguments. Finally, abstract-counting numerals are compatible with arithmetical calculations such as those in (3d).

---

[2] The existence of linguistic expressions of fractions and decimals (Haida & Trinh 2019) as well as *zero* (Bylinina & Nouwen 2018) suggests that not only natural numbers are represented in natural language. In this paper, however, we will focus only on cardinal numerals associated with integers.

(3) a. **Five** is a Fibonacci number.

  b. **Five** is bigger than four.

  c. Jasna can count up to **five**.

  d. Ten divided by **five** equals two.

Let us now contrast (3) with the behavior of pluralities of individuals, which lack the properties mentioned above. The sentence (4a) is awkward, because being a Fibonacci number is not something that can be attributed to a collection of things. The truth conditions of (4b) are different than those of (3b), e.g., the sentence would not be true if one compared five pebbles with four boulders. Moreover, object-counting numeral phrases are illicit in constructions calling for numeric values, see (4c). Finally, as witnessed by the awkwardness of (4d) expressions denoting pluralities of entities are incompatible with mathematical statements.

(4) a. #**Five** things are a Fibonacci number.

  b. #**Five** things are bigger than four things.

  c. #Jasna can count up to **five** things.

  d. #Ten things divided by **five** things equals two things.

An additional difference is that only object-counting numerals allow for modification by comparative and superlative modifiers (5a–b), while abstract-counting numerals are odd with such modifiers (6a–b). One could imagine an intended interpretation of (6a) to mean something like there is a number which is bigger than five such that it is a Fibonacci number. But such a reading is not available. Similarly, (6b) is a funny statement.

(5) a. More than **five** cities were destroyed.

  b. At least **five** children got sick.

(6) a. #More than **five** is a Fibonacci number.

  b. #Ten divided by at least **five** equals two.

Yet another difference concerns scalarity. It has been known for a long time that object-counting numerals give rise to scalar implicatures. For instance, (7a) gets a lower bounded construal, i.e., it is interpreted in a way that you must take at least five cards. Similarly, the

main clause in (7b) can be felicitously continued by *if not more* which has been standardly assumed to be the evidence for the lower bound inference (Horn 1972).

(7)   a.   You must take **five** cards.

       b.   John took **five** cards, if not more.

However, abstract-counting numerals do not give rise to scalar implicatures and always get bilaterally bounded interpretations (cf. Sadock 1984, Horn 1992, Bultinck 2005). For instance, (8a) cannot be interpreted in a way that in order to get ten you must multiply two by at least five. That seems strange, and false if this interpretation is forced. Similarly, (8b) cannot be felicitously continued by the *if not more* clause.

(8)   a.   You must multiply two by **five** to get ten.

       b.   Two multiplied by **five** equals ten, #if not more.

The distinction is further corroborated in languages with gender marking. Here abstract-counting cardinals often display different gender than object-counting forms (cf. Fassi Fehri 2018, Wągiel to appear). For instance, the Polish numeral *pięć* ('five') triggers neuter agreement on the demonstrative in its abstract-counting use, see (9a). When used as an object-counting form, it agrees in gender with the modified noun. Consequently, the demonstrative in (9b) has the same gender as the noun as well (rather than the neuter gender as in (9a)).[3]

(9)   a.   **To**   **pięć**   zapisano   cyfrą   arabską,   a   tamto   rzymską.   Polish

           this.N   five.N   was.written   cipher.INS   Arabic.INS   and   that.N   Roman.INS

           'This five was written with an Arabic numeral, whereas that one with a Roman one.'

       b.   **Te**   **pięć**   jabłek   leży   na   stole,   a   tamte   na   podłodze.

           these.NV   five.NV   apples.NV.GEN   lies   on   table.LOC   and   those.NV   on   floor.LOC

           'These five apples are lying on the table whereas those ones on the floor.'

---

[3] In the paper, we will use the following abbreviations: N – neuter gender, F – feminine, M – masculine, V – virile gender, NV – non-virile; NOM – nominative case, GEN – genitive, ACC – accusative, DAT – dative, INS – instrumental, LOC – locative, ABL – ablative; TAM – tense/aspect/mood, TOP – topic marker, COP – copula, AUX – auxiliary, DUR – durative marker, DIR – directional prefix; ART – article, DEM – demonstrative; RED – reduplication; LIG – ligature; CL – classifier and NBR – abstract number marker.

With the (semantic) distinction between the two types of cardinals in place, let us now focus on the question how the numeral itself reflects (or fails to reflect) this difference morphologically.

3. Abstract-counting and object-counting numerals across languages

It has been observed in the literature that abstract- and object-counting numerals may have a different form (Hurford 1998, 2001). On the basis of the morphological relation between the numerals, we distinguish four patterns. We introduce them in this section.

It is important to keep in mind that the classification into patterns pertains to individual numerals, and not to full languages. The reason for this is that it may be the case that various patterns co-exist within one and the same language. It is possible, for instance, to find a language where the numerals 2–4 display one pattern, while numerals 5 and above show a different pattern. Our goal here is to classify the patterns rather than to describe individual languages. A full account for any given language emerges only after each of its numerals is classified according to the patterns we describe below and each such numeral is given a different lexical entry along the lines explored in Section 5.

3.1. The asymmetric pattern

In the pattern we will call ASYMMETRIC, the object-counting numeral contains the abstract-counting numeral as a proper subpart. This pattern is typically found in obligatory classifier languages. For instance, in Japanese (Sudo 2016), the object-counting numeral 5 contains the morpheme *go* ('five'), and an additional morpheme, e.g., *rin*, *hiki* or *ko*, usually referred to as a classifier, see (10b).[4] Bare cardinals are ungrammatical as nominal modifiers, see (10a).

(10) a. ***go**-no   hana   Japanese
         five-GEN   flower

      b. **go-rin**-no   hana
         five-CL-GEN flower
         'five flowers'

---

[4] In Japanese, the morpheme *rin* is a classifier used for counting flowers, *hiki* is for small animals, whereas *ko* is a general classifier for counting inanimate entities.

However, in an abstract-counting environment, e.g., an arithmetical statement such as (11a), the numeral cannot occur with a classifier. Though *ko* is a general classifier and could be used to indicate any type of inanimate entity, the sentence in (11b) is odd.[5]

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

Importantly, the presence or absence of the classifier is not a function of the syntactic position of the numeral. For instance, one could propose that the numeral has the classifier only before a noun (10) but not in isolation (11). The reason why this cannot be so is the numeral's behavior in the predicate position, see (12) (Sudo 2016: 8). Here the numeral is not an attribute to a noun, yet it has to be accompanied by a classifier. The classifier is therefore not governed by the syntactic position of the numeral. It is there to mark the object counting function.

(12) katteiru   doobutsu-wa   **yon-*(hiki)**-da.   Japanese
       have.as.pets animal-TOP   four-(CL)-COP
       'I have four pets.' (lit. 'The pets I have are four.')

Among the languages which distinguish between object- and abstract-counting numerals, the asymmetric pattern is relatively frequent. For instance, the same asymmetry is also attested in Mandarin. Specifically, in an object-counting context such as (13), numerals require classifiers. In environments unambiguously calling for abstract counting, classifiers are odd, see (14).[6]

(13) a. ***wǔ** shū   Mandarin
           five   book
      b. **wǔ-běn** shū
           five-CL   book
           'five books'

---

(14) a. shí chúyǐ **wǔ** shì èr. Mandarin
     ten divide.by five COP two
     'Ten divided by five is two.'

    b. #shí-gè chúyǐ **wǔ-gè** shì èr-gè.
      ten-CL divide.by five-CL COP two-CL

Similar contrasts are attested in other obligatory classifier languages such as Vietnamese and Thai.[7] We interpret these facts as suggesting that the abstract-counting function is basic and the object-counting function is derived from it (both morphologically and semantically).

3.2. Symmetric numerals

In a number of languages, however, we observe no asymmetry between the numerals, i.e., the two forms are identical. For instance, in English both functions are expressed by the same formal exponent, see (15). We will call cardinals displaying such behavior SYMMETRIC numerals.

(15) a. **five** roses
    b. Ten divided by **five** is two.

We interpret such forms as ambiguous. In one use, they are semantically equivalent to the Japanese classifier construction *go-ko*, in another to the bare *go*. In the former use, the bare numeral itself incorporates a classifier semantics (Krifka 1995).

It is important to emphasize that the patterns are not properties of a language as a whole, but rather of a particular numeral. This is required by languages such as Chol and Mi'gmaq, where some numerals show the asymmetric pattern, while other numerals exhibit the symmetric one (Bale & Coon 2014).

3.3. Idiosyncratic numerals

A separate pattern reported in the literature is represented by what we call IDIOSYNCRATIC numerals. A numeral is idiosyncratic if the two functions in question are expressed by two different morphologically simplex forms, i.e., the relation between the numerals is suppletive. We acknowledge that uncontroversial instances of this pattern are not easy to provide since

---

[7] We thank Tue Trinh and Pittayawat Pittayaporn for their judgments on Vietnamese and Thai, respectively.

there might be alternative explanations for the two different numeral forms. Nevertheless, we illustrate here the general shape of the facts as reported in the literature and later in Section 5 we show how such numerals fit into our story should these descriptions be correct.

An example of the idiosyncratic pattern is provided by the Maltese numeral 2 (a case treated as such in Hurford 1998). This numeral has two distinct, morphologically unrelated forms, specifically *żewġ* and *tnejn* (both 'two'). The form *żewġ* serves as an object-counting expression, see (16a) (from A. Borg 1974: 293). On the other hand, *tnejn* cannot be used as a nominal modifier, see (16b).

(16) a. **żewġ** nisa   Maltese

         two_1 women

         'two women'

    b. ***tnejn** nisa

         two_2 women

Conversely, *tnejn* can be used as an abstract-counting numeral in mathematical statements, see (17a) (from A. J. Borg 1987: 62), whereas *żewġ* is incompatible with environments calling for numeric arguments such as (17b).[8]

(17) a. Tnejn u   **tnejn** jagħmlu   erbgħa.   Maltese

        two_1 and two_1 they-make four

        'Two and two make four.'

    b. *Żewġ u   **żewġ** jagħmlu   erbgħa.

        two_2 and two_2 they-make four

This distribution can be captured by analyzing *żewġ* as a portmanteau form that corresponds to the agglutinative classifier construction, e.g., *ni-ko* ('two-CL') in Japanese. The form *tnejn* then corresponds to the bare numeral root in an obligatory classifier language, e.g., *ni* ('two').[9]

---

[8] We would like to thank Albert J. Borg for the discussion of the Maltese data and his judgments on (16)–(17).

[9] Note, however, that A. Borg (1974) suggests that the distribution can be captured in syntactic terms, whereby *żewġ* is used before a noun and *tnejn* is used when no nominal follows. For reasons of space, we do not resolve the tension between Hurford's and A. Borg's views here. Another interesting fact to note is that some Maltese numerals apparently display the inverse pattern, e.g., 6 is *sitta* in the abstract-counting function and *sitt* in the object-counting function (A. Borg 1974: 291–292). This is perfectly consistent with our approach since the patterns apply to particular numerals, not to whole languages.

Though rare, the idiosyncratic pattern is not limited to Maltese. Additional examples include at least the Ojibwe (Algonquian) *ninkotw ~ pešikw* (both 'one') (Denney & Odjig 1973), Palaung (Austroasiatic) *ū ~ hlɛh* (both 'one') (Greenberg 1978) as well as the Ịḅanị (Eastern Ijaw) numerals *gbẹrẹ ~ ngịẹ* (both 'one') and *oyi ~ atie* (both 'ten') (Obikudo 2016).

3.4. Inverse numerals

The final pattern is one where the abstract-counting numeral properly contains the form of the object-counting numeral. This pattern is the inverse of the asymmetric pattern and we therefore call it the INVERSE pattern. For instance, in the case of the German *eins ~ ein* (both 'one'), abstract counting is expressed by a morphologically more complex form (*eins*) than object counting (Hurford 1998, 2001). As demonstrated by the contrasts in (18) and (19), *eins* cannot be used as a nominal modifier while *ein* is incompatible with arithmetical environments.[10]

(18) a. **ein** Mädchen   German
        one girl
        'one girl

    b. ***ein-s**   Mädchen
        one-NBR   girl

(19) a. Zehn geteilt   durch **ein-s**   ist gleich zehn.   German
        ten   divided by   one-NBR is   equal ten
        'Ten divided by one equals ten.'

    b. *Zehn geteilt   durch **ein** ist zehn.
        ten   divided by   one is   ten

As in Japanese, the inverse pattern cannot be reduced to the fact that the numeral in (19a) is not followed by a noun, while the one in (18a) is. To see that, consider first the fact that when the neuter head noun *Mädchen* is elided in a phrase such as (20b), the numeral indeed cannot appear bare and takes the shape *eines* (see, e.g., Murphy 2018 for a recent discussion and references).

---

[10] We would like to thank Nina Haslinger for her judgments and comments concerning German. Notice that the absence of case cannot be the only source of the ill-formedness of (19b), because using the accusative masculine form *einen* ('one') does not help, see (i). The neuter accusative *eines* cannot be used either, as we show in (21).

(i) *Zehn geteilt   durch **einen**   ist gleich zehn.   German
      ten   divided by   one.M.ACC is equal ten

However, this form is different from the abstract-counting numeral and it cannot be used as one, see (21). In sum, neither of the numerals in (20)–(21) is followed by a noun, and thus the contrast must be attributed to the distinction between abstract and object counting.

(20) a. CONTEXT: How many girls came to the party?

      b. Nur **ein-*(es)**.   German

         only one-N.SG

         'Only one.'

(21) *Zehn geteilt durch **ein-es** ist zehn.   German

      ten   divided by   one-N.SG is ten

The inverse pattern seemingly requires the opposite semantic derivation compared to the object-counting function. In particular, it seems that here the abstract-counting function has some extra meaning, corresponding to the additional morpheme (the suffix *-s* in German). However, admitting this would jeopardize a morpho-semantic explanation of the widespread asymmetry illustrated in (10)–(11) and (13)–(14) as well as any unified typology of cardinal numerals. We will therefore propose an alternative derivation of this pattern.

Other cases of inverse numerals reported in the literature include the Hungarian *kettő ~ két* (both 'two') and the Eastern Basque *biga ~ bi* (both 'two') (Hurford 1998, 2001). However, our inquiry reveals that most probably the distinction has not been described properly, because the marked and unmarked forms introduced above are not in complementary distribution. For instance, in Hungarian, both *kettő* and *két* can be used as nominal modifiers, see (22), but only *kettő* can appear in mathematical contexts, as demonstrated in (23).[11]

(22) **két / ket-tő**   macska   Hungarian

      two two-NBR cat

      'two cats'

(23) Egy meg egy az ***két / ket-tő**.   Hungarian

      one plus one is two two-NBR

'One plus one equals two.'

Though space prohibits to investigate cases such as Hungarian 2 in detail, we report that to the best of our knowledge uncontroversial instances of the inverse marking pattern are scarce. However, on the strength of the German evidence we need to make some theoretical space for the existence of pairs such as *eins ~ ein*.

3.5. Interaction with gender

The final piece of evidence in favor of distinguishing two separate functions of numerals comes from languages that mark gender on cardinal numerals. It has been observed in the literature that there is a non-trivial interaction between grammatical gender and quantification (Arsenijević 2016, Fassi Fehri 2018, Wągiel to appear) and the data to be discussed seem to further corroborate this claim. As already signaled in Section 2, gender marking on cardinals is often correlated with the abstract/object counting distinction. For instance, let us consider the non-virile/virile distinction in Bulgarian. In this language, numerals 2–10 agree with the noun in virility. We illustrate this in (24), where the unmarked non-virile form *pet* ('five') is incompatible with virile NPs (24a), whereas the marked virile form *petima* ('five') is degraded with non-virile NPs (Cinque & Krapova 2007). Interestingly, only the unmarked form *pet* can appear in a context calling for a number concept, see (25).[12]

(24) a. **pet-(*ima)** ženi   Bulgarian
             five-V   women.NV
             'five women'
    b. **pet-?(ima)** mâže
             five-V   men.V
             'five men'

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

---

[12] We would like to thank Marina Pantcheva for her judgments and comments concerning Bulgarian.

The contrast in (25) is an instance of the asymmetric pattern arising as a result of the interaction with grammatical gender. A similar distinction is also attested in other Slavic languages that mark gender on cardinals, e.g., Polish and Slovak (Wągiel to appear).

Interestingly, the inverse pattern can also emerge as a consequence of gender marking. As an example, consider Standard Arabic. Numerals in this language are known for the gender mismatch between cardinals 3–10 and modified nouns ('gender polarity'). As demonstrated in (26a), when the numeral combines with a feminine noun, it has the unmarked masculine form. But when combined with a masculine noun, the numeral requires the feminine suffix *-at*, see (26b). Crucially, only the feminine form, i.e., *t̠alaat̠-at* ('three') in (27a), can be used as an abstract-counting expression (Fassi Fehri 2018). As a consequence, the more marked form is used as an abstract-counting expression, similar to the pattern observed for German 1 in (19).

(26) a. **t̠alaat̠-(*at)-u**   banaat-in   Standard Arabic

         three-F-NOM   girls.F-GEN

         'three girls'

   b. **t̠alaat̠-*(at)-u**   ʔawlaad-in

         three-F-NOM   boys.M-GEN

         'three boys'

(27) a. **t̠alaat̠-at**-un t-usawii ʔit̠nayni zaʔid waaḥid.   Standard Arabic

         three-F-NOM F-equals two   plus one

         'Three equals two plus one.'

   a. ***t̠alaat̠**-un y-usawii ʔit̠nayni zaʔid waaḥid.

         three-NOM M-equals two   plus   one

Before we move on to proposing an analysis that attempts to explain how all the patterns are derived, let us summarize the data.

3.6. Data summary

In this section, we have examined four patterns of morphological marking regarding the abstract/object counting distinction summarized in Table 1.

TYPE	LANGUAGE	NUMBER	ABSTRACT	OBJECT
SYMMETRIC	English	5	*five*	*five*
IDIOSYNCRATIC	Maltese	2	*tnejn*	*żewġ*
ASYMMETRIC	Japanese	5	*go*	*go-ko*
INVERSE	German	1	*ein-s*	*ein*

Table 1. Morphological-marking patterns

Symmetric numerals do not distinguish between the abstract- and the object-counting form. Idiosyncratic numerals have a suppletive form for each of the two functions. The asymmetric pattern arises when object-counting cardinals contain the abstract-counting form. And finally, in inverse numerals the abstract-counting form contains the object-counting form. Importantly, the patterns are numeral-specific rather than language-specific, since in one language there can be more than one type of numerals. Finally, the asymmetric and inverse pattern can emerge as a result of gender marking on the cardinal.

With the patterns in place, let us now move on to propose a set of ingredients which will consequently allow us to derive all the marking patterns in Table 1.

4. Universal semantic features

In this section, we put forth a set of cross-linguistically invariant semantic components that the two types of cardinal numerals are made out of. This will allow us to compositionally derive the abstract/object counting distinction examined in Section 2. The interest of the proposal is that the very same components will feed into a unified morpho-semantic system to account for the typological variation described in Section 3.

4.1. Classifiers and nominal denotations

In the semantic literature, it is commonly assumed that the distinction between obligatory classifier languages such as Mandarin and non-classifier languages such as English is due to a difference in the semantics of nouns (e.g., Chierchia 1998, 2010, Borer 2005, Rothstein 2010, Li 2011, Scontras 2013). According to this standard approach, all nouns in classifier languages are mass-like in the sense that they have uncountable denotations. Thus, in order to combine the NP with a numeral, a classifier is required. That is because classifiers compensate a semantic deficit of nouns by turning uncountable denotations into countable ones.

Though the received view is appealing, there are reasons to believe that it is in fact incorrect. One of the main counterarguments is based on the fact that in obligatory classifier languages, certain counting modifiers, e.g., Japanese *tasuu* ('numerous') and *nan-zen-toiu* ('thousands'), appear without classifiers, and they only combine with NPs denoting countable entities, e.g., *hana* ('flower') as opposed to *ase* ('sweat') (Sudo 2016). Another problem for the standard approach concerns languages such as Chol and Mi'gmaq, where some cardinals require classifiers, whereas others do not (Bale & Coon 2014).

Therefore, we adopt an alternative explanation of the role of classifiers in obligatory classifier languages. According to this alternative, it is not the semantics of nouns what differentiates the two types of languages but rather it is the semantics of numerals. In particular, in obligatory classifier languages, classifiers are required to compensate semantic deficits of cardinals. According to the hypothesis, these lack the semantics that would enable them to function as counting devices, i.e., as modifiers equipped with an operation allowing for numeric quantification (Krifka 1995, Bale & Coon 2014, Sudo 2016). In this paper, we will embrace this alternative way of thinking about numerals and classifiers.

4.2. Semantic components

In order to account for the morphological patterns discussed in the previous section, we propose the ingredients in (28) as the universal inventory of components that numerals are built out of. In particular, we postulate three syntactic heads SCALE, NUM (for 'number') and CL (for 'classifier'), accompanied with the standard function application operation.[13]

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

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

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

As indicated in (28a), the meaning of SCALE$_m$ is a closed interval $[0, m]$ (type $\langle n, t\rangle$), i.e., a set of natural numbers containing 0, a maximum m and all integers in between. We assume that the value of an upper bound m is lexically encoded and depending on the numeral it could be, e.g., $[0, 4]$ or $[0, 5]$ etc. In other words, for each numeral, there is a distinct SCALE head encoding a distinct maximum, as designated by the subscript $_m$ in (28a). Moreover, we posit that SCALE is

---

[13] The chosen labels are intended to evoke the relevant associations, e.g., NUM – the notion of a number concept. Notice, however, that we do not claim that, e.g., the semantics of CL is all there is to the meaning of classifiers in classifier languages. What we focus on here is strictly restricted to the abstract/object counting distinction.

the core feature underlying the semantics of all number words including cardinal numerals as well as ordinals, multiplicatives, fractions and other complex numerical expressions. We will provide the motivation for postulating SCALE in Section 4.4 below.

The NUM head is a function from intervals to numbers, see (28b). It takes a set of integers and introduces the maximization operation MAX which yields the greatest number from that set. For instance, if applied to the interval [0, 4], it will return 4 whereas when applied to [0, 5], it will yield 5 etc. Hence, the role of NUM is to forge a proper name of an abstract mathematical entity, i.e., an expression of primitive type n.[14]

Finally, CL is a function from an integer to a counting device (type $\langle n, \langle\langle e, t\rangle, \langle e, t\rangle\rangle\rangle$), see (28c). Specifically, it takes a number and shifts it to a predicate modifier equipped with the pluralization operation * (Link 1983) and the measure function $\#(P)$ (Krifka 1989).[15] While * adds to the denotation of a predicate all the sums that can be formed from the atomic elements of a singular denotation, $\#(P)$ is an extensive measure function which maps a plurality of entities onto a numeric value corresponding to the number of entities making up that plurality. It is additive, monotonic and has the Archimedean property. The goal of the CL head is thus to form an expression that can be used for numeric quantification over actual objects.[16]
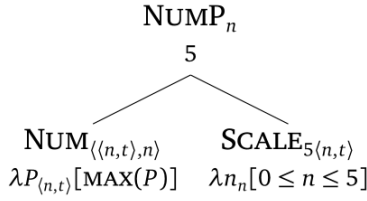
4.3. Composition

In order to combine the meanings of the heads introduced above, we assume standard function application. Combining the ingredients in (28) in a compositional fashion leads to the structures in (29)–(30), where SCALE$_5$ encodes semantic information associated with a particular numeral.

---

[14] The fact that abstract-counting numerals are proper names is quite likely reflected in their nominal nature (see Ionin & Matushansky 2018: 26–28).
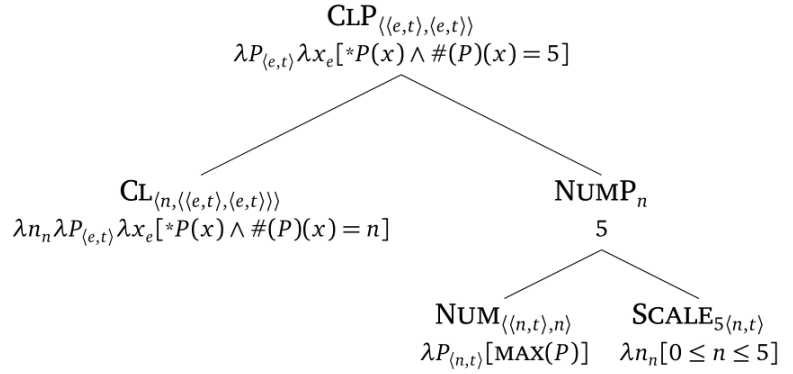
[15] Actually, Krifka uses the NU (for 'natural unit') operation. We depart from the original notation here since due to independent reasons, we do not postulate that numeric quantification resorts to the notion of a natural unit (for some discussion, see Wągiel 2018).

[16] Though we assume CL to be a function from number concepts to predicate modifiers, nothing prevents it to be a shift from integers to some other type, e.g., cardinal properties or determiners. In principle, the proposed system is compatible with other theories of numerals as long as it is possible to relate a type for the object-counting function with the primitive type n.

(29) Abstract counting   (30) Object counting

$$\text{NUMP}_n$$
$$5$$

$$\text{NUM}_{\langle\langle n,t\rangle,n\rangle} \qquad \text{SCALE}_{5\langle n,t\rangle}$$
$$\lambda P_{\langle n,t\rangle}[\text{MAX}(P)] \qquad \lambda n_n[0 \le n \le 5]$$

$$\text{CLP}_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle}$$
$$\lambda P_{\langle e,t\rangle}\lambda x_e[*P(x) \wedge \#(P)(x) = 5]$$

$$\text{CL}_{\langle n,\langle\langle e,t\rangle,\langle e,t\rangle\rangle\rangle} \qquad\qquad \text{NUMP}_n$$
$$\lambda n_n\lambda P_{\langle e,t\rangle}\lambda x_e[*P(x) \wedge \#(P)(x) = n] \qquad\qquad 5$$

$$\text{NUM}_{\langle\langle n,t\rangle,n\rangle} \qquad \text{SCALE}_{5\langle n,t\rangle}$$
$$\lambda P_{\langle n,t\rangle}[\text{MAX}(P)] \qquad \lambda n_n[0 \le n \le 5]$$

The tree in (29) represents the structure and meaning of an abstract-counting numeral, referring to the number 5. The tree shows that due to the application of MAX to SCALE₅, the interval [0, 5] is turned into the integer 5. The resulting expression is thus of the type *n*, designating a number concept. NUMP is thus a proper name of an arithmetical entity. Properties such as *be prime* or *be a Fibonacci number*, which we assume to be of type $\langle n, t\rangle$, can be predicated of this object.

The tree in (30) shows how the meaning of a corresponding object-counting numeral is derived. The CL phrase as a whole is interpreted as a predicate modifier. After the number slot of the CL head in (30) is saturated by 5, we obtain an expression that applies to a predicate and yields a set of pluralities of entities that have the relevant property and whose cardinality equals 5. Hence, when CLP is combined with a countable NP, we obtain a denotation such as the one in (31), i.e., a set of pluralities of roses such that each plurality in that set consists of 5 roses.

(31) 〚*five roses*〛 $= \lambda x_e[*\text{ROSE}(x) \wedge \#(\text{ROSE})(x) = 5]$

In sum, this proposal correctly derives different properties of abstract- and object-counting numerals. However, it apparently gives rise to redundancy. In particular, it may seem equally justified to postulate only NUM (denoting numbers) and CL (turning numbers to counting devices) and to do away with the SCALE part of the structure. The reasons why we need SCALE will become fully clear in Section 5 where we demonstrate how our system derives the morphological patterns investigated above and inverse numerals in particular. But before we get there, let us consider some independent evidence for decomposing the numeral's base into two components.

4.4. Motivating SCALE

16

The evidence comes from the observation that abstract-counting numerals, which are the basic numeral type on our account, can be morphologically complex. In what follows, we will present data from two languages with such complex abstract-counting numerals. Each of the two languages represents a different pattern identified in Section 3. Specifically, the first language contains numerals representing the symmetric pattern, i.e., the abstract- and object-counting numerals are identical. The second language represents the asymmetric pattern, i.e., the object-counting numeral is derived from the abstract-counting one. The only difference between the cardinals discussed in Section 3 and those examined here is that here the abstract-counting numeral is already morphologically complex.

The first language to be discussed is Shuhi, a Qiangic language of the Tibeto-Burman family. The data in (32) show the object-counting uses of the numeral 1 (Qi & He 2019: 65).[17] In each of the phrases in (32), the numeral root $dʑi^{33}$ ('one') is accompanied by a different classifier, whose choice depends on the noun. Of special interest is the so-called default classifier ko^{35} seen in (32a).

(32) a. rɔʔ³⁵   **dʑi³³-ko³⁵**   Shuhi

      horse   one-CL

      'one horse'

   b. nɯ⁵⁵gu³¹   **dʑi³³-ly⁵⁵**

      cloth   one-CL

      'one cloth'

   c. lɑ³³re⁵⁵   **dʑi³³-ʈshu⁵⁵**

      towel   one-CL

      'one towel'

The reason why this classifier is special is because it also shows up on numerals when speakers recite the counting sequence (*one*, *two*, *three* etc.) and in environments describing abstract arithmetical operations, as in (33) (from Qi & He 2019: 69).

(33) **dʑi³³-ko³⁵**-re³³   dʑi³³-ko³⁵-ɦo~³³   me³³-ba³³-le⁵⁵   ŋe³³-ko³⁵   le³³-ʐiʔ³³-dʑo~³³.   Shuhi

    one-CL-ABL   one-CL-LOC   DIR-add-AUX   two-CL   DIR-become-DUR

    'One plus one is two.'

---

[17] The superscript integers in the example lines in (32) denote tones.

The example (33) then shows that the numeral 'one' is symmetric and crucially, morphologically complex also in the abstract-counting function. In fact, the numeral root $d\!z\!i^{33}$ never occurs on its own. Under the hypothesis that morphemes express meaning, the data in (33) entail that Shuhi abstract-counting cardinals have at least two components of meaning. Specifically, we understand the classifier morpheme ko^{35} to be the exponent of NUM in abstract-counting contexts like (33). On the other hand, individual numeral roots such as $d\!z\!i^{33}$ simply lexicalize different SCALE heads.

Let us now turn to a language with complex asymmetric numerals. To see what such cardinals look like, consider first the abstract-counting numerals 1–5 in the Oceanic language Vera'a (Vanuatu), as compiled in Table 2 (Schnell 2011: 73–74). Interestingly, they all include the prefix *vō-/ve-* attached to various numeral roots. The affixal status of *vō-/ve-* is evidenced by the structure of multiplicative numerals, which substitute *vō-* with the prefix *vag-*.[18]

NUMBER	CARDINAL	MULTIPLICATIVE
1	*vō-wal*	*vag-wal*
2	*vō-ru(ō)*	*vag-ru(ō)*
3	*vō-'ōl*	*vag-'ōl*
4	*vō-ve'*	*vag-ve'*
5	*ve-limē*	*vag-limē*

Table 2. Vera'a numerals

Such cardinals are not a Vera'a peculiarity. For instance, in the related language Vurës (Vanuatu) the numerals 1–10 are also morphologically complex. Table 3 shows that such cardinals include the invariant prefix *ni-* attached to various numeral roots (Malau 2016: 126). The reason why *ni-* must be acknowledged as a separate morpheme is the fact that ordinals and multiplicatives lack this prefix.[19] The added value of the Vurës forms is that they have a different prefix. In other words, what the languages share is an abstract pattern, rather than the specific pieces (even though the roots of the numerals seem obviously cognate).

---

[18] Numerals higher than 5 are apparently monomorphemic and lack the prefix *vō-/ve-*. This is in line with our general approach of looking at particular numerals rather than whole languages.
[19] The prefix *ni-* reappears in some higher ordinals and multiplicatives, e.g., *saṁul tiwial dēmē nirōne* ('twelfth').

NUMBER	CARDINAL	ORDINAL	MULTIPLICATIVE
2	*ni-rō*	*rō-ne*	*vagō-rō*
3	*ni-töl*	*töl-ne*	*vag-töl*
4	*ni-vet*	*vet-ne*	*vag-vet*
5	*ni-tevelēm*	*tevelēm-ne*	*vag-tevelēm*
6	*ni-levetē*	*levetē-ne*	*vag-levetē*

Table 3. Vurës numerals

Coming back to Vera'a, the new aspect of the cardinals 1–5 in this language relates to their object-counting use in NPs. Recall that in Shuhi, the bi-partite abstract-counting numerals such as *dʑi³³-ko³⁵* 'one' modify nouns directly, as in (32a). In Vera'a, the object-counting function has an extra morpheme, i.e., the so-called 'ligature' *ne*, see (34) (Schnell 2011: 73).[20]

(34) ēn   woqe'enge   **ne vō-ru**   Vera'a
      ART   tree   LIG NBR-two
      'two trees'

Importantly, the ligature *ne* appears only with numerals, i.e., it does not co-occur, e.g., with adjectives. Even more interestingly, *ne* appears only when numerals have the object-counting function. In the abstract-counting use, the numeral is preceded by an article and the ligature is missing, see (35) (from Schnell 2011: 83). This strongly suggests that Vera'a numerals 1–5 display the asymmetric pattern, similar to the majority of Japanese numerals, and the ligature is nothing else than a counterpart of the Japanese general classifier. Crucially, Vera'a differs from Japanese in that abstract-counting numerals are morphologically complex.

(35) vēvē-gi   ne   lukun ēn   naw, din   ēn   **vō-'ōl**…   Vera'a
      mother-3SG   TAM   count ART wave reach ART NBR-three
      'Then his mother counted the waves reaching (the number) three…'

---

[20] As Schnell (2011: 74), the ligature can sometimes be omitted. In such cases, "it seems that often the combination of noun and bare numeral is a fixed lexicalized expression". An example provided by Schnell is the combination of 'day' and 5, which refers to a ceremony taking place five days after a person's death. There are, however, syntactic contexts where the ligature is obligatory, as we discuss later.

The absence of the ligature in (35) cannot be attributed to the fact that the head noun is missing/elided. As pointed out by Schnell (2011: 82), in elliptical NPs "the ligature invariably precedes the numeral in all examples found in the corpus". An example is shown in (36), where the numeral is used in the object-counting function and the ligature is obligatory.

(36) di=m   ri-riv,   rivriv   qē   **ne vō-ruō** anē.   Vera'a
      3SG=TAM RED-plant RED:plant finish LIG NBR-two DEM
      'He planted, planted these two [i.e., yams].'

To sum up, (35)–(36) present a near minimal pair where neither numeral is followed by a noun, yet the numerals differ in that one has the obligatory ligature/classifier, while the other lacks it. The relevant difference between the two cases is therefore semantic and concerns the distinction between abstract and object counting. Importantly, in both Shuhi and Vera'a already the abstract-counting numeral is morphologically complex, which calls for postulating two underlying meaning components within this class of cardinals.

The facts discussed so far strongly motivate the presence of two ingredients in the denotation of abstract-counting numerals. However, they do not tell us directly what these ingredients are. The reason why we propose SCALE and NUM as the relevant semantic components is based on an intuition that numerals are at their core scalar entities and that each numeral is associated with a particular interval on the number scale.

A very similar assumption underlies the research on constructions in which cardinals are modified by spatial and directional prepositions such as *above five* and *up to five*, respectively (Corver & Zwarts 2006, Nouwen 2008). The key idea here is that numerals correspond to points on a scale with a non-arbitrary starting point and at the same time they are associated with other numbers via an ordering relation (Nouwen 2016). Consequently, the reason why we take SCALE heads to denote closed intervals of the $[0, m]$ type relates to the need for encoding a fixed lower bound and ordering which has been argued for independently.

Notice also that our proposal corresponds to the idea that numerical discrete infinity arises from the combinatorial mechanism of natural language (Chomsky 2008). In particular, the natural number system can be constructed syntactically by Merge from a single element, namely the empty set (Watanabe 2017). The fact that the empty set corresponds to the concept of 0 provides additional conceptual justification to assume that the intervals encoded by SCALE heads are of the $[0, m]$ type.

While different cardinals are associated with various SCALE components, we assume only one NUM head. This is supported by the fact that while Vera'a complex abstract-counting numerals contain various numeral roots, the morpheme *vō-/ve-* is invariant. This fact strongly suggests that *vō-/ve-* encodes a single operation applicable to different intervals which we take to be the maximization operator MAX, a shift well-motivated in compositional semantics.

To conclude, the proposal we are making is motivated by the fact that abstract counting-numerals may be morphologically complex. This would be hard to explain if abstract-counting numerals simply corresponded to a single number-denoting head. We have also argued that its semantic ingredients are conceptually justified. Though many questions remain open and much more research needs to be done, we leave the precise details for future work. In the next section, we will discuss how the morphological patterns examined in Section 3 relate to the postulated semantic components.

## 5.   Deriving the shape of cardinal numerals

We now proceed to show how the shapes of the various types of numerals examined so far are derived. In doing so, we adopt one of the basic insights of current theories of morphology, namely late insertion. The idea is that the actual exponents of meaning, e.g., */fʌɪv/*, are not present in the syntactic/semantic component at all. What is present in syntax are only the abstract (language-invariant) features, e.g., SCALE, NUM and CL, which the syntactic component assembles into structures like those in (29) and (30). The actual exponents are inserted only after syntax. Late insertion is the core of current approaches such as Distributed Morphology (Halle & Marantz 1993) and Nanosyntax (Starke 2009).

The rules that relate features to exponents are the lexical items of a language. Lexical items are language-specific objects that pair a well-formed syntactic structure *S*, e.g., [NUM SCALE], a well-formed phonological structure *P*, e.g., */fʌɪv/*, and a conceptual meaning *C* (if any). From the perspective of generating the phonological form, lexical items are read as instructions of the sort: if syntax constructs the structure *S*, pronounce/realize *S* by the exponent *P*.

In what follows, we will describe a model of spellout that uses lexical entries in order to translate the universal structures of numerals given in (29) and (30) onto language-particular outputs, yielding the full typology of numeral types introduced in Section 3. In doing so, we will rely on the framework of Nanosyntax (Starke 2009, Caha 2009), and in particular on the version of the theory that includes spellout driven movement (Starke 2018, Baunaz & Lander 2018, Caha et al. 2019). We think that a similar goal can be achieved using the framework of Distributed Morphology, but we will not show this for reasons of space.

5.1. Symmetric numerals

Let us start with the fact that a cardinal such as /fʌɪv/ can function as an object-counting numeral. In Nanosyntax, this means that this numeral is able to lexicalize all the three meaning components of the structure given in (30). The way this is technically achieved is by assigning this numeral the lexical entry as in (37). The lexical entry pairs a well-formed syntactic structure with phonology. When syntax builds a structure corresponding to the numeral, as in (30), the lexical item (37) can be used to pronounce this structure, because it is identical to it. This is indicated by the circle around the relevant structure in (38), i.e., the structure as a whole (specifically its top-most phrasal node) is pronounced by /fʌɪv/.

(37) CLP ⟺ /fʌɪv/   (38)   (39)

CL NUMP

NUM SCALE$_5$
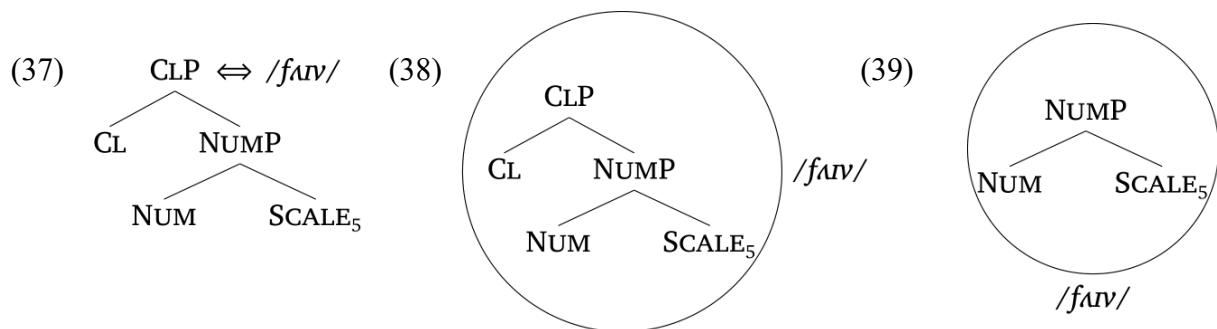
Suppose now that syntax builds only the structure for the abstract-counting numeral as in (29). In Nanosyntax, this structure can be also lexicalized by the lexical entry in (37), because it is contained inside the lexical entry. This is indicated in (39), where the relevant structure is circled, indicating that /fʌɪv/ can be inserted here. As a result, when a numeral has a lexical entry such as (37), it is ambiguous between the two different uses.
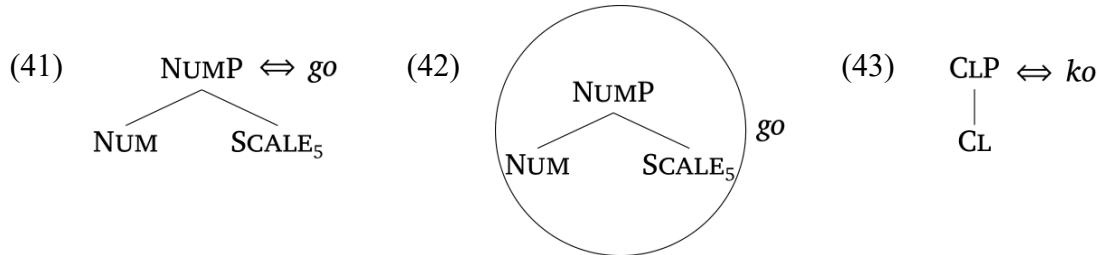
This type of ambiguity arises as a consequence of one of the core properties of late insertion models, which is that lexical entries are not tailor-made for one specific use. For instance, in Distributed Morphology, this is encoded by one of the core principles, the so-called Subset Principle. In Nanosyntax, the relevant principle works slightly differently, and it is called the Superset Principle, see (40). This principle achieves that a lexical entry can be used to pronounce a particular structure iff it contains that structure as a sub-part (proper or not). Since both trees in (38) and (39) are contained in (37), the numeral /fʌɪv/ is ambiguous.

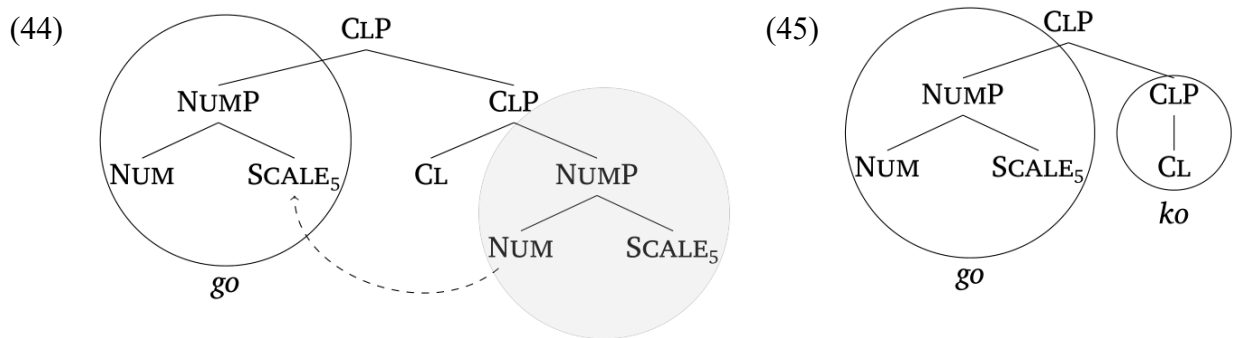(40) *The Superset Principle* (Starke 2009):

A lexically stored tree matches a syntactic node iff the lexically stored tree contains the syntactic node.

5.2. Asymmetric numerals

Let us now turn to asymmetric numerals. These arise when the lexical entry of the numeral only contains the NUM and SCALE components, see (41). Such an exponent can lexicalize the structure of an abstract-counting numeral, see (42). However, it cannot lexicalize the object counting structure, because it does not contain it (it lacks the feature CL).

(41) $\text{NUMP} \Leftrightarrow go$

      NUM SCALE_5

(42)

    NUMP

  NUM SCALE_5 *go*

(43) $\text{CLP} \Leftrightarrow ko$

      CL

In such numerals, CL needs to be spelled out by a separate entry, like the one in (43). In order for this entry to apply at the CLP node, NUMP is displaced from its base-position due to movement, see (44). The lower copy of the moved element, i.e., the trace, is shaded. In Nanosyntax, the movement in (44) is driven by the need to lexicalize the phrasal CLP node by the lexical entry (43). In particular, before the movement, the lexical item (43) does not match the lower CLP in (44), because it does not contain this phrase.

(44)

            CLP

  NUMP   CLP

NUM SCALE_5 CL   NUMP

                  NUM SCALE_5

   go

(45)

            CLP

  NUMP   CLP

NUM SCALE_5 CL

              ko

   go

After the movement (and ignoring the trace), it does so, as shown in (45), where the classifier *ko* is inserted at the relevant node, as indicated by the circle. We leave it unexplored here as to how precisely this movement is triggered, but see, e.g., Baunaz & Lander (2018), Caha et al. (2019), Vanden Wyngaerd et al. (2020) for detailed explanations.

To sum up, we have shown that when the numeral root is specified only for SCALE and NUM, it will appear bare only in the abstract-counting function given in (42). In the object-counting function in (45), CL must be spelled out by an additional element, i.e., a classifier. Importantly, there is no requirement that all numerals in a language must have the same lexical entries. As a consequence, a system of this type allows for the simultaneous presence of

multiple kinds of numerals within a single grammar. This is required to capture the fact reported in Section 3 that in some languages cardinals can display more than one pattern, e.g., in Chol and Mi'gmaq both symmetric and asymmetric numerals are attested (Bale & Coon 2014).

5.3 Idiosyncratic numerals

Let us now consider idiosyncratic numerals. Recall that such numerals have a dedicated non-decomposable form for each of the two functions. An example discussed in Section 3.3 was the Maltese numeral 2. Such a behavior can be modeled by postulating two suppletive lexical entries for a numeral. One entry introduces the root found in the abstract-counting function, see (46). The second entry introduces the root for the object-counting numeral, as in (47).

(46) NUMP ⟺ *tnejn*
       NUM   SCALE$_2$

(47) CLP ⟺ *żewġ*
     CL   NUMP
          NUM   SCALE$_2$
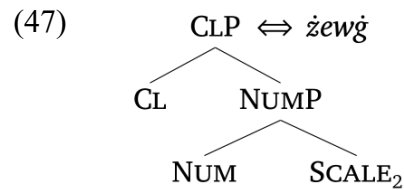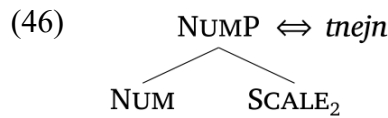
When syntax produces the object-counting structure, only *żewġ* ('two') can be inserted, because *tnejn* ('two') does not contain this structure, see (48). However, in the abstract-counting structure, both numerals are candidates for insertion, because they both contain the relevant structure. As a result, a competition arises between the two items, see (49). The competition is resolved by the so-called Elsewhere Condition (Kiparsky 1973), which says that when two rules compete, the more specific one applies. The more specific of the two items in (46) and (47) is *tnejn*, because it applies in a proper subset of cases compared to *żewġ*. Therefore, *tnejn* wins in competition and it is inserted in (49). The numeral *żewġ* loses and it is not inserted, as indicated by the strike-through below the circle in (49).

(48)
        CLP
żewġ | CL   NUMP
         NUM   SCALE$_2$

(49)
      NUMP
   NUM   SCALE$_2$
   żewġ / *tnejn*
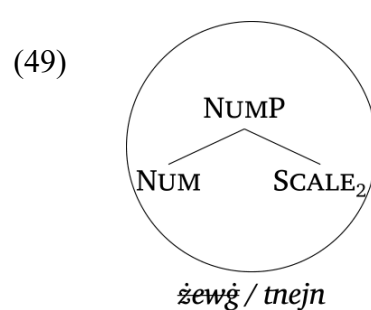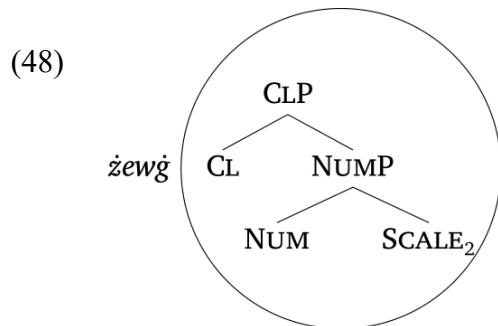
5.4. Inverse numerals

Let us now turn to the treatment of inverse numerals such as German 1. We start with the fact that the abstract-counting numeral *eins* is morphologically complex. This entails that the root *ein* cannot, on its own, spell out the full abstract-counting structure. Therefore, we assign to it the lexical entry in (50), which says that this numeral can only spell out the SCALE node. As a consequence, the numeral needs to combine with an additional morpheme in order to express all the components of the abstract-counting structure, in particular NUM. Hence, we specify the morpheme *-s* for exactly this function, as provided in (51). Thus, the abstract-counting structure looks as in (52), where the SCALE feature has moved from its base position across NUM, similarly to (44).

(50) SCALE $\Longleftrightarrow$ *ein*

(51) NUMP $\Longleftrightarrow$ *s*
    |
  NUM

(52)
NUMP
[SCALE$_1$ NUMP] *s*
  ein   NUM
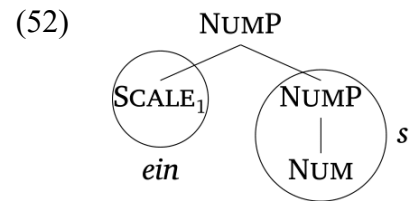
With this analysis in place, what can we say about the object-counting use of *ein* in German? Starting from the observation that the root of *ein*, recall (50), cannot pronounce the full structure of the object-counting numeral, we must draw the conclusion that the remaining features are spelled out by a different morpheme. Since there is no overt marker visible in the object-counting *ein*, we postulate a zero morpheme, i.e., *ein-Ø*. The lexical entry of the zero marker is given in (53) and the full structure of the object-counting numeral in (54). In this structure, the SCALE component moves out of the NUMP, which is spelled out by the relevant Ø marker.

(53)
CLP $\Longleftrightarrow$ Ø
CL   NUMP
      |
     NUM

(54)
NUMP
[SCALE$_1$ CLP] Ø
 ein CL   NUMP
            |
          NUM
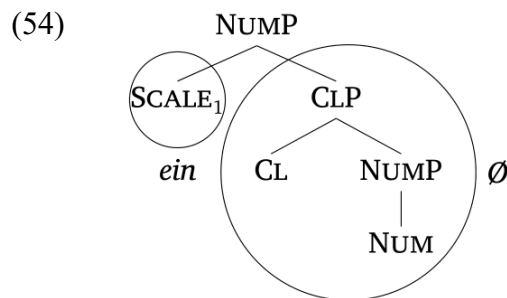
This analysis is supported on paradigmatic grounds. Specifically, the German object-counting numeral *ein* has a full case/number paradigm illustrated in Table 4. Here, only some cells of the paradigm show the problematic inverse pattern, while other cells have an overt affix. The idea of a zero marker is thus supported on analogy with the other paradigm cells.

	MASCULINE	NEUTER	FEMININE
NOM	*ein-Ø*	*ein-Ø*	*ein-e*
ACC	*ein-en*	*ein-Ø*	*ein-e*
GEN	*ein-es*	*ein-es*	*ein-er*
DAT	*ein-em*	*ein-em*	*ein-er*

Table 4. The inflection of the German indefinite article/numeral 1

An interesting advantage of the proposed analysis is that it explains why the inverse pattern is so scarce. The reason is that inverse numerals can arise only in a very particular configuration. Specifically, the numeral needs to be stored simply as SCALE, NUM needs to have an overt exponent and a portmanteau for NUM and CL needs to be silent. As a result, in the very rare cases were all those conditions are fulfilled abstract- and object-counting numerals are spelled out according to the inverse pattern.

5.5. Complex abstract-counting numerals

Finally, let us turn to the cardinals in Shuhi and Vera'a discussed in Section 4.4. Recall that these languages contain symmetric and asymmetric numerals, respectively. However, the crucial difference is that in Shuhi and Vera'a, the relevant abstract-counting numerals are morphologically complex.

Let us first demonstrate how Vera'a asymmetric numerals are derived. Recall that in this language, the abstract-counting numerals 1–5 decompose into the prefix *vō-/ve-* and a root, as depicted in Table 2. Object-counting numerals add another marker, i.e., *ne*, so that they routinely contain a sequence of three independent morphemes; e.g., *ne vō-ruō* ('two') in (34) corresponds to the object-counting use of the English monomorphemic numeral *two*.

In (55)–(57), we present simplified lexical entries for Vera'a 2. For simplicity, we assume that the morphemes target the relevant terminal nodes (in the more current versions of Nanosyntax, the prefixes would correspond to the so-called projecting specifiers as in Starke 2018). The full structure of the Vera'a numeral 2 is given in (58).

(55) SCALE ⟺ *ruō*

(56) NUM ⟺ *vō*

(57) CL ⟺ *ne*

(58)
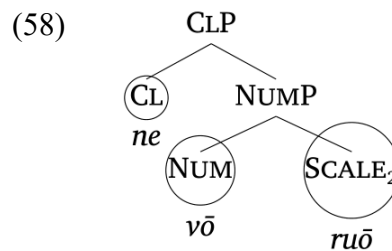
The lexical entries for Shuhi 1 would be modelled as shown in (59) and (60). Since the abstract counting numeral is complex, we attribute to dzi^{33} ('one') a lexical entry such that it spells out SCALE only, see (59). This means that already in the abstract-counting use, something else has to spell out NUM. This role is taken up by the classifier ko^{35}, with an entry as in (60). Note that the syntactic side of this entry is identical to the German $-\emptyset$ marker in (53).

(59) SCALE $\Longleftrightarrow dzi^{33}$

(60) CLP $\Longleftrightarrow ko^{35}$

(61)

(62)
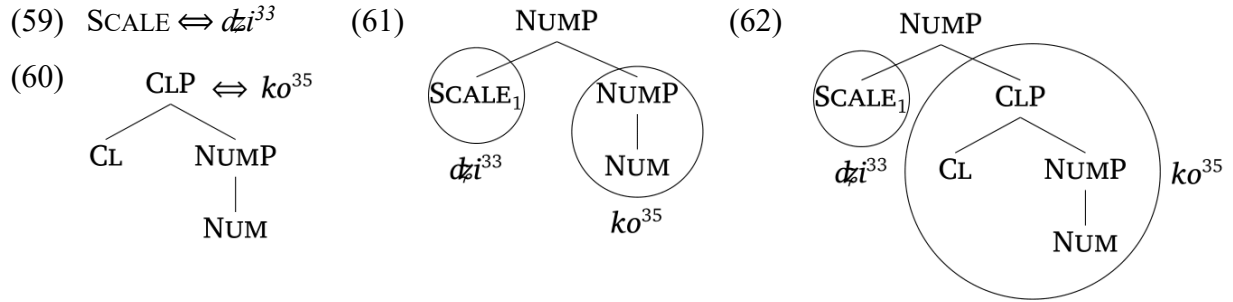
The derivations then produce (61) as the structure of the abstract-counting numeral, with SCALE moving across NUM, as in the German case in (52). In German, we had a dedicated NUMP marker $-s$. Shuhi lacks such a marker, and lexicalizes the NUMP by the default classifier ko^{35}. This is again allowed by the Superset Principle, because the lexical entry for ko^{35} in (60) contains NUMP. The structure of the object-counting numeral is shown in (62), with SCALE moving cyclically out of CLP. This structure is identical in the relevant respects to the German (54). CLP is lexicalized by ko^{35} and a symmetric numeral emerges.

As one can see, the approach developed offers a sufficient explanatory power to capture the patterns. Before we move on to a more detailed discussion of the role of gender in abstract/object counting, let us summarize the meaning/form correspondences derived by the proposed system.

5.6. Summary

Our account of the morphological patterns described in Section 3 is summarized in the first four lines of Table 5. The left-hand column of the table highlights our proposal for abstract-counting numerals. The column is split into two compartments, each containing one of the two relevant components of meaning, i.e., SCALE and NUM. The lines below indicate how the meaning components are pronounced by individual numerals. The same logic is followed in the object-counting column on the very right. The comparison of the two columns yields the symmetric pattern in the first line, followed by the idiosyncratic pattern, the asymmetric pattern and the inverse pattern.

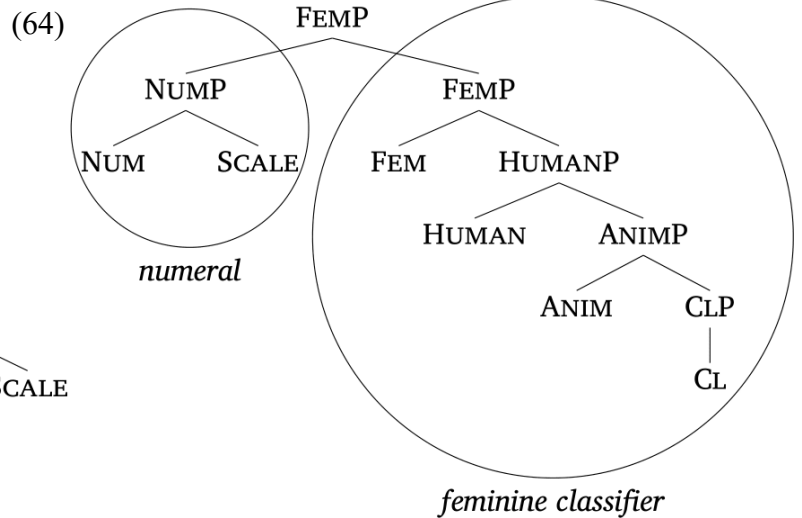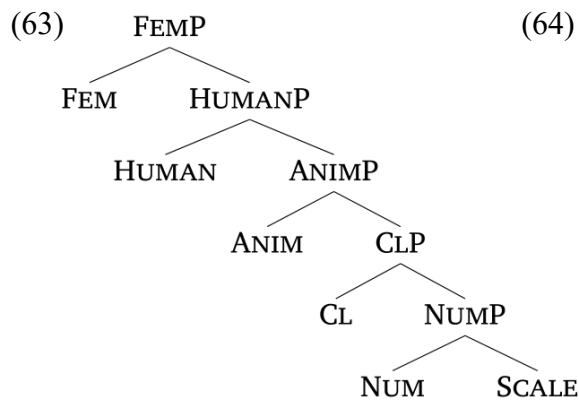| ABSTRACT COUNTING			OBJECT COUNTING		
SCALE	NUM		SCALE	NUM	CL
five		English 5	*five*		
tnejn		Maltese 2	*żewġ*		
go		Japanese 5	*go*		*ko*
ein	*s*	German 1	*ein*	*Ø*	
dʑi^{33}	*ko^{35}*	Shuhi 1	*dʑi^{33}*	*ko^{35}*	
ruō	*vō*	Vera'a 2	*ruō*	*vō*	*ne*

Table 5. Meaning/form correspondences

The two bottom lines show how the meaning components are lexicalized in languages with complex abstract-counting numerals. The line corresponding to Shuhi 1 is actually the same as the one above it, i.e., the one depicting the inverse pattern. The only difference is that the German -Ø is realized as an overt marker. Finally, the last line depicts the facts as found in Vera'a, where each component is pronounced by a separate marker. We end this section by restating the conclusion that the proposed system successfully derives all the attested patterns of marking cardinal numerals in their abstract- and object-counting uses.

6. Gender in object-counting numerals

Before we conclude, let us briefly discuss how the proposed system approaches data where the shape of the marker on the object-counting numeral pronouncing the CL head, e.g., a classifier or gender affix, is dependent on the morphological and/or semantic class of the modified noun. Among the cases discussed so far, such patterns arise prominently in the German cardinal numeral 1, recall Table 4, in Bulgarian 2–10, recall (21)–(22), in Standard Arabic 3–10, recall (23), and in general in classifier constructions in obligatory classifier languages with multiple types of classifiers, recall footnote 4.

What we propose is that in cases like the ones mentioned above it is not the bare CLP that modifies the noun. In addition to the CLP, which we take to be absolutely essential for object counting, the numeral may also include additional features that add presuppositions concerning the nature of the referents of the modified noun (following the approach outlined in Sudo 2016). In (63), we show a rather rich structure of this kind, where several relevant features are added on top of the CLP, each introducing a presupposition of its own.

(63)

```
        FEMP
       /    \
     FEM   HUMANP
          /     \
      HUMAN    ANIMP
              /    \
           ANIM   CLP
                 /    \
                CL   NUMP
                    /    \
                  NUM   SCALE
```

(64)

```
              FEMP
             /    \
          NUMP    FEMP
         /   \   /    \
       NUM SCALE FEM  HUMANP
                     /     \
                 HUMAN    ANIMP
                         /    \
                      ANIM   CLP
                              |
                              CL
```

numeral

feminine classifier

The pronunciation of such a structure proceeds on analogy with the previously described cases. In (64), we show what happens in asymmetric numerals of the Japanese type. Recall that such numerals spell out only NUMP. In order to allow for the spellout of FEMP, the NUMP must move out of the constituent in (63), yielding a structure such as the one in (64). Here a 'human-feminine' classifier attaches to the numeral, as indicated by the large circle.[21]

In order to demonstrate how the proposed system would work on a concrete example, let us consider the Abkhaz cardinals 2–10. In the gender system of Abkhaz (Northwest Caucasian), nouns are grammatically classified as human vs. non-human, with the class of human nouns being further subdivided into masculine and feminine (Chirikba 2003: 24–25).[22] In addition, the cardinal numerals 2–10 occur in two forms, as shown in Table 6 (Chirikba 2003: 34, Hewitt 2010: 33).

| NUMBER	ABSTRACT/NON-HUMAN	OBJECT HUMAN
4	*pš´-ba*	*pš´-j°ə́(k´)*
5	*x°-ba*	*x°-j°ə́(k´)*
6	*f-ba*	*f-j°ə́(k´)*
7	*bəž´-ba*	*bəž´-j°ə́(k´)*
8	*aa-ba*	*aa-j°ə́(k´)*

---

[21] An anonymous reviewer asks whether we cannot say that classifiers only spell out gender heads, which would dispense with the need of the CL head. There are two reasons. First, there are classifiers that are necessary for object counting and yet do not contribute any presupposition, e.g., the Vera'a 'ligature' *ne* and general classifiers such as Mandarin *gè*. To accommodate such non-presuppositional classifiers, we need the CL head. The same type of non-presuppositional CL head is also often present in monomorphemic object-counting numerals. In general, the ability of a numeral to be used as a counting device cannot be conflated with presuppositions about the counted objects since one needs to distinguish between object counting and abstract counting anyway, see Section 2.

[22] In fact, grammars of Abkhaz typically talk about noun classes rather than genders. However, for the purpose of this paper we assume that the terms are synonymous.

Table 6. Abkhaz cardinal numerals

In the left-hand column, we see object-counting numerals for non-human nouns. They are clearly complex and consist of a numeral root and the suffix *-ba*. The right-hand column gives object-counting numerals for human nouns, which take the suffix *-jʷə́(kʼ)*. Unlike the Bulgarian case of *pet ~ petima* ('five'), recall (24)–(25), there is no formal asymmetry between the two shapes of gendered numerals in Abkhaz both of them being without any doubt morphologically complex. However, only non-human cardinals can be used for abstract counting (Hewitt 2010: 33–34). In other words, Abkhaz non-human cardinals are ambiguous between the abstract- and object-counting function whereas human cardinals can only be used to quantify over human entities and are infelicitous in arithmetical environments.[23]

Within the proposed system, the Abkhaz cardinals 2–10 can be modeled similarly to the German numeral 1. In both languages, the relevant numeral roots spell out SCALE. As a consequence, already the abstract-counting function is expressed by a morphologically complex form, i.e., some other element than the root must pronounce NUM. The only difference is that the suffixes *-ba* and *-jʷə́(kʼ)* lexicalize bigger chunks of structure than German *-s* and *-Ø*, compare (51) and (53), respectively. For instance, Abkhaz 8 can be analyzed as (65) and (66). While both *-ba* and *-jʷə́(kʼ)* include NUM, due to the Elsewhere Principle only non-human numerals can be used for abstract counting.
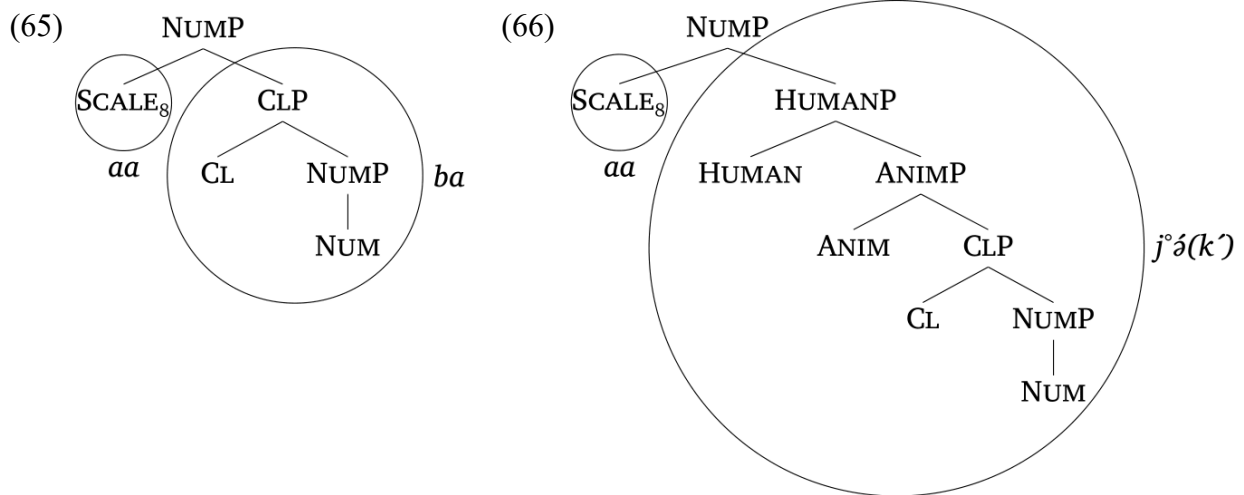
Importantly, the Abkhaz case illustrates that the abstract type of analysis that yields the inverse pattern of morphological marking in German 1, as developed in Section 5.4, is needed independently. Put simply, if *-jʷə́(kʼ)* was null, we would get an inverse system in Abkhaz.

The proposed algorithm can be also applied to other cases of different patterns arising as a result of the interaction with gender, as discussed in Section 3.5. Thus, our proposal provides a powerful tool to explain a lot of typological variation within numerals which, to the best of our knowledge, has not been accounted for so far.

7. Conclusion

In this paper, we have examined cross-linguistic meaning/form correspondences between what we refer to as abstract-counting and object-counting cardinal numerals. The former are numerals used to refer to number concepts, e.g., in arithmetical statements. The latter are used as nominal modifiers in order to quantify over individuals. We have argued that object-counting numerals both syntactically and semantically contain abstract-counting numerals.

More specifically, we have postulated three syntactic heads that correspond to the primitive semantic ingredients of numerals. SCALE defines a closed interval of natural numbers, NUM turns such an interval into a number concept corresponding to the greatest integer in that interval and CL turns a number concept into a counting device. These ingredients are assembled into the invariant structure [CL [NUM SCALE]] for object-counting numerals and [NUM SCALE] for abstract-counting numerals.

In order to derive the surface patterns, we have adopted a nanosyntactic model of morphology (while admitting that alternative approaches are conceivable). The theory of Nanosyntax allowed us to formulate an account where the universal numeral structures are the input to the lexicalization procedure, which produces variable outputs depending on what the (late-inserted) lexical entries look like. Under this approach, all the attested variation reduces to the shapes of the lexical entries available for a particular numeral.

Specifically, object-counting cardinals, which are stored as the entire tree, can serve both the object- and abstract-counting function. On the other hand, numerals stored as [NUM SCALE] can be used only as abstract-counting expressions and additional morphology, e.g., a classifier, is required to pronounce CL. Very rare cases of fully suppletive forms arise when a language has both types of lexical entries. Finally, in scarce cases of inverse marking (where abstract-counting cardinals appear to be more marked than object-counting ones), the pattern is derived by a very small cardinal stored simply as SCALE, so that both the abstract-counting function and the object-counting function each require some extra morphology. When

the 'classifier-like' morpheme happens to be silent, the inverse pattern arises. When it is overt, we get a system where both the abstract- and the object-counting numeral are morphologically complex.

References

Arsenijević, Boban. 2016. *Gender as a grammaticalized classifier system: The case of the Serbo-Croatian neuter*. Unpublished manuscript, lingbuzz/002848.

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

Baunaz, Lena & Lander, Eric. 2018. Nanosyntax: The basics. In Baunaz, Lena, De Clercq, Karen, Haegeman, Liliane & Lander, Eric (eds.), *Exploring Nanosyntax*, 3–56. Oxford: Oxford University Press.

Borer, Hagit. 2005. *Structuring Sense I: In Name Only*. Oxford: Oxford University Press.

Borg, Albert J. 1987. To be or not to be a copula in Maltese? *Journal of Maltese Linguistics* 17(18): 54–71.

Borg, Alexander. 1974. Maltese numerals. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 124(2): 291–305.

Bultinck, Bert. 2005. *Numerous Meanings: The Meaning of English Cardinals and the Legacy of Paul Grice*. Amsterdam: Elsevier.

Bylinina, Lisa & Nouwen, Rick. 2018. On "zero" and semantic plurality. *Glossa: A Journal of General Linguistics* 3(1)/98: 1–23.

Caha, Pavel. 2009. *The Nanosyntax of Case.* Ph.D. thesis, University of Tromsø, Tromsø.

Caha, Pavel, De Clercq, Karen & Vanden Wyngaerd, Guido. 2019. The fine structure of the comparative. *Studia Linguistica* 73(3): 470–521.

Chierchia, Gennaro. 1998. Plurality of mass nouns and the notion of 'semantic parameter'. In Rothstein, Susan (ed.), *Events and Grammar*, 53–103. Dordrecht: Kluwer.

Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174(1): 99–149.

Chirikba, Viacheslav. 2003. *Abkhaz*. Muenchen: Lincom.

Chomsky, Noam. 2008. On phases. In Freidin, Robert, Otero, Carlos P. & Zubizarreta, Maria Luisa (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.

Cinque, Guglielmo & Krapova, Iliyana. 2007. A note on Bulgarian numeral classifiers. In Alboiu, Gabriela, Avram, Andrei A., Avram, Larisa & Isac, Daniela (eds.), *Pitar Moş: A*

Building with a View. Papers in Honour of Alexandra Cornilescu, 45–51. Bucharest: Editura Universitatii din Bucuresti.

Corver, Norbert & Zwarts, Joost. 2006. Prepositional numerals. *Lingua* 116(6): 811–836.

Dékány, Éva & Csirmaz, Anikó. 2017. Numerals and quantifiers. In Alberti, Gábor & Laczkó, Tibor (eds.), *Syntax of Hungarian: Nouns and Noun Phrases. Comprehensive Grammar Resources*, 1044–1150. Amsterdam: Amsterdam University Press.

Denney, J. Peter & Odjig, Lorraine. 1973. The meaning of *ninkotw* 'one' and *pešikw* 'one' in Ojibway. *International Journal of American Linguistics* 39(2): 95–97.

Fassi Fehri, Abdelkader. 2018. *Constructing Feminine to Mean: Gender, Number, Numeral, and Quantifier Extensions in Arabic*. Lanham: Rowman & Littlefield.

Greenberg, Joseph H. 1978. Generalizations about numeral systems. In Greenberg, Joseph H., Ferguson, Charles A. & Moravcsik, Edith A. (eds.), *Universals of Human Language*, vol. 3: *Word Structure*, 249–295. Stanford: Stanford University Press.

Haida, Andreas & Trinh, Tue. 2019. *A more inclusive theory of numerals*. A paper presented at the 12th conference on Syntax, Phonology and Language Analysis (SinFonIJA 12), Masaryk University in Brno.

Hewitt, George. 2010. *Abkhaz: A Comprehensive Self-Tutor*. Muenchen: Lincom.

Horn, Laurence R. 1972. *On the Semantic Properties of the Logical Operators in English*. Ph.D. thesis, University of California, Los Angeles.

Horn, Laurence R. 1992. The said and the unsaid. In Barker, Chris & Dowty, David (eds.), *Proceedings of Semantics and Linguistic Theory 2*, 163–192. Columbus: Ohio State University.

Hurford, James R. 1998. The interaction between numerals and nouns. In Plank, Frans (ed.), *Noun Phrase Structure in the Languages of Europe*, 561–620. Berlin: Mouton de Gruyter.

Hurford, James R. 2001. Languages treat 1–4 specially. *Mind & Language* 16(1): 69–75.

Ionin, Tania & Matushansky, Ora. 2018. *Cardinals: The Syntax and Semantics of Caridnal-Containing Expressions*. Cambrdige, MA: MIT Press.

Kiparsky, Paul. 1973. 'Elsewhere' in phonology. In Anderson, Stephen & Kiparsky, Paul (eds.), *A Festschritt for Morris Halle*, 93–106. New York: Holt, Rinehart & Winston.

Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, Renate van Benthem, Johan & von Emde Boas, Peter (eds.), *Semantics and Contextual Expression*, 75–115. Dordrecht: Foris.

Krifka, Manfred. 1995. Common nouns: A contrastive analysis of Chinese and English. In Carlson, Gregory N. & Pelletier, Francis J. (eds.), *The Generic Book*, 398–411. Chicago: University of Chicago Press.

Li, XuPing. 2011. *On the Semantics of Classifiers in Chinese*. Ph.D. thesis, Bar-Ilan University, Ramat Gan.

Link, Godehard. 1983. The logical analysis of plural and mass nouns: A lattice-theoretical approach. In Bäuerle, Rainer, Schwarze, Christoph & von Stechow, Arnim (eds.), *Meaning, Use, and Interpretation of Language*, 302–323. Berlin: Mouton de Gruyter.

Malau, Catriona. 2016. *A Grammar of Vurës, Vanuatu*. Boston: Walter de Gruyter.

Murphy, Andrew. 2018. Pronominal inflection and NP ellipsis in German. *Journal of Comparative German Linguistics* 21: 327–379.

Nouwen, Rick. 2008. Directionality in numeral quantifiers: The case of *up to*. In Friedman, Tova & Ito, Satoshi (eds.), *Proceedings of Semantics and Linguistic Theory 18*, 569–582. Ithaca: Cornell University.

Nouwen, Rick. 2010. Two kinds of modified numerals. *Semantics and Pragmatics* 3: 1–41.

Nouwen, Rick. 2016. *Making sense of the spatial metaphor for number in natural language*. Unpublished manuscript, lingbuzz/003100.

Obikudo, Ebitare F. 2016. Counting: The Ịḅanị way. In Ndimele, Ozo-mekuri & Chan, Eugene S. L. (eds.), *The Numeral Systems of Nigerian Languages*, 217–223. Port Harcourt: M&J Grand Orbit Communications.

Qi, Jianjun & He, Chuansheng. 2019. The morphosyntax of numerals dʑi^{33}/dʑĩ^{35} 'one' in Shuhi and implications for the semantics of numerals. *Lingua* 225: 63–80.

Rothstein, Susan. 2010. Counting and the mass/count distinction. *Journal of Semantics* 27(3): 343–397.

Rothstein, Susan. 2013. A Fregean semantics for number words. In Aloni, Maria, Franke, Michael & Roelofsen, Floris (eds.), *Proceedings of the 19th Amsterdam Colloquium*, 179–186.

Rothstein, Susan. 2017. *Semantics for Counting and Measuring*. Cambridge: Cambridge University Press.

Sadock, Jerrold. 1984. Whither radical pragmatics? In Schiffrin, Deborah (ed.), *Meaning, Form, and Use in Context: Linguistic Applications*, 139–49. Washington, DC: Georgetown University Press.

Schnell, Stefan. 2011. *A Grammar of Vera'a: An Oceanic Language of North Vanuatu*. Kiel: Kiel University.

Scontras, George. 2013. Accounting for counting: A unified semantics for measure terms and classifiers. In Snider, Todd (ed.), *Proceedings of Semantics and Linguistic Theory 23*, 549–569. Ithaca, NY: CLC Publications.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1): 1–6.

Starke, Michal. 2018. Complex left branches, spellout, and prefixes. In Baunaz, Lena, De Clercq, Karen, Haegeman, Liliane & Lander, Eric (eds.), *Exploring Nanosyntax*, 239–249. Oxford: Oxford University Press.

Sudo, Yasutada. 2016. The semantic role of classifiers in Japanese. *Baltic International Yearbook of Cognition, Logic and Communication* 11(1): 1–15.

Vanden Wyngaerd, Guido, Starke, Michal, De Clercq, Karen & Caha, Pavel. 2020. How to be positive. *Glossa: A Journal of General Linguistics* 5(1)/23: 1–34.

Watanabe, Akira. 2017. Natural language and set-theoretic conception of natural number. *Acta Linguistica Academica* 64(1): 125–151.

Wągiel, Marcin. 2018. *Subatomic Quantification*. Ph.D. thesis, Masaryk University, Brno.

Wągiel, Marcin. to appear. Grammatical gender meets classifier semantics: Evidence from Slavic numerals. In Gribanova, Vera, Grimberg, Sabrina, Petersen O Farrill, Erika, Portelance, Eva & Waldon, Brandon (eds.), *Proceedings of the 27th Meeting of Formal Approaches to Slavic Linguistics*. Ann Arbor: Michigan Slavic Publications.