

The morphology of complex numerals: A cross-linguistic study

Lukáš Žoha, Marcin Wągiel and Pavel Caha

Masaryk University, Brno, Czech Republic

Abstract

Complex numerals are numerals composed of two or more numeral roots, e.g., *three hundred five*. Complex numerals fall into two classes called additive (e.g., *twenty-three* = 20 + 3) and multiplicative (e.g., *three hundred* = 3 × 100). There are two possible approaches to capturing their structure. Analysis A (e.g., He 2015) says that complex numerals form a constituent that quantifies over entities denoted by the noun. Analysis B (e.g., Ionin and Matushansky 2018) says that each numeral independently combines with the expression denoting counted entities. This article investigates the morphology of complex numerals in a sample of 17 diverse languages to determine which of these analyses (if any) is more accurate. Our goal is to lay out the patterns and discuss how well they fit with these theories. Our preliminary conclusion is that both structures should be allowed based on the data in our sample, though structures adhering to Analysis A (the complex numeral is a constituent) seem to be more common than the other type.

Keywords: complex numerals; counting; morphology; syntax; linguistic typology

1. Introduction

From the perspective of their composition, numerals can be divided into simple numerals and complex numerals. In (1a), there is an example of a simple numeral containing just a single numeral root. In (1b) and (1c), there are examples of complex numerals of which the simple numeral *three* is a part.¹

- (1) English
- a. ***three***
- b. *one hundred three*
- c. ***three*** *hundred*

¹ The term complex numerals currently has different uses in the literature. In Wągiel and Caha (2021), the term refers to numerals that are morphologically complex, e.g., the German *ein-s* ‘one’ has two morphemes. Here we use it differently, namely as referring to numerals that contain more simple numerals, as in (1b-c). This latter use of the term ‘complex’ is also found in Ionin and Matushansky (2018).

Complex numerals can be further divided into two groups, based on how we arrive at their meaning. If addition is involved, as in (1b), the complex numeral is called additive ($100 + 3$). If multiplication is involved (3×100), the complex numeral is called multiplicative, see (1c). (Addition and multiplication can of course combine within a single numeral with more than two elements.)

A boarder-line case is represented by numerals such as *six-ty*, where clearly, we have the numeral *six* and an additional morpheme *-ty*. Even though *-ty* is not a free-standing numeral, it is a morpheme with a constant contribution across a range of cases (*seven-ty*, *nine-ty*, etc.). We consider these cases as relevant for our study because we mainly focus on the form of the multiplier (*six*, *seven*) rather than on the form of the irregular base (*-ty*).

There are two main approaches to the structure of complex numerals. Under Analysis A (e.g., He 2015, Cinque 2021, He and Her to appear), complex numerals form a constituent that combines with the noun and quantifies over the entities denoted by the noun. This is shown in (2a,b) for additive and multiplicative complex numerals, respectively. In both cases, the complex numeral is formed first and then it is added to the noun as a constituent.

- (2) Analysis A: Complex numerals as constituents
 a. [twenty + three] books
 b. [three \times hundred] books

Under Analysis B (e.g., Ionin and Matushansky 2006, 2018), each numeral independently combines with the expression denoting counted entities. For additive numerals, Analysis B leads to a structure like (3a). At the beginning, there are two phrases with two nouns: *twenty books* and *three books*. The complex numeral *twenty-three* is created from these two phrases by eliminating the noun from the phrase *twenty books* (either by right node raising or phonological deletion).

- (3) Analysis B: Complex numerals as non-constituents
 a. [twenty books] + [three books]
 b. [three \times [hundred books]]

For multiplicative numerals, Analysis B leads to a structure like (3b), where the numeral *three* counts the number of [hundred books]. In neither of these structures do the two numeral roots form a constituent in the underlying structure to the exclusion of the modified noun.

There are several empirical tests that allow one to distinguish between the two analyses (see He 2015, Ionin and Matushansky 2018, He and Her to appear for a recent discussion). In this paper, we try to differentiate the two hypothetical structures by a novel test that has not been used in the literature so far. The idea is that in at least some languages, the two structures can be distinguished on morphological grounds by looking at the morphology of simple numerals inside complex numerals.

This new diagnostic builds on the observation by Wągiel and Caha (2020, 2021) that languages may morphologically distinguish between two kinds of numerals, which we call here abstract- and object-counting numerals (following the terminology in Wągiel and Caha).

Abstract-counting numerals refer to an abstract number concept (Bultinck 2005, Rothstein 2017). They are therefore used in statements expressing mathematical properties, as in (4a), or arithmetical equations, as in (4b); the relevant numerals are highlighted. Often (though not always), the same numerals are also used for counting in a sequence, see (4c). (The counting sequence is grouped with the arithmetical uses, e.g., in Tatsumi 2021.) The second type of numerals that we need to recognize are object-counting numerals. These are forms used inside NPs for counting objects, as in (4d).

- (4) English
- a. **Three** is a prime number.
 - b. Two plus **three** equals five.
 - c. one, two, **three**, ...
 - d. There are **three** cats in the room.

In English there is no morphological distinction between abstract-counting and object-counting numerals. In some languages, however, these numerals have different forms. For instance, the Mandarin numeral ‘two’ is *èr* for abstract counting, and *liǎng* + classifier for object counting (He 2015). We shall return to this example in Section §3.3.

Another language that distinguishes abstract- from object-counting numerals is Javanese. As we show in (5), in this language the numeral ‘five’ has three different forms, ordered in terms of morphological complexity.

- (5) Javanese ‘five’ (Robson 1992: 75–76)
- a. **ma**
 - b. *li-ma*
 - c. *li-ma-ng*

The numeral *ma* in (5a) can only be used when numerals are recited in the counting sequence, and we, therefore, consider it an abstract-counting numeral. The numeral *lima* in (5b) is ambiguous like English ‘five,’ and it can be used both in the counting sequence and as an object-counting numeral. This latter use is illustrated in (6a).

- (6) Javanese (Robson 1992: 75–77)
- a. *jeruk* **lima**
orange five
‘five oranges’
 - b. **limang** *rupiah*
five rupiah
‘five rupiahs’

The numeral *limang* can only be used for object counting. In (6b), it is used for counting currency, and thus we classify it as an object-counting numeral.

The idea that we follow in this paper is that in languages where we find a difference between object- and abstract-counting numerals, it is relevant to look at which of these numerals is used in object-counting complex numerals. To see why, consider again the two analyses, repeated for convenience in (7)–(8):

- (7) Analysis A: Complex numerals as constituents
 - a. [twenty + three] books
 - b. [three × hundred] books
- (8) Analysis B: Complex numerals as non-constituents
 - a. [twenty books] + [three books]
 - b. [three × [hundred books]]

Starting with Analysis B, we expect that each numeral inside the complex numeral is an object-counting numeral, because it is associated to a particular entity it quantifies over. If even one abstract-counting numeral is found inside a complex numeral, this would be unexpected.

On the other hand, Analysis A makes different predictions. For additive numerals, it predicts that at least one numeral inside the complex numeral is in the abstract-counting shape. The reason is the following. According to the simplest version of Analysis A, the numeral is created first through an abstract arithmetic operation (20+3), and this requires abstract-counting numerals. Only after the numeral is formed, it is used to quantify over objects. This hypothesis therefore predicts that inside additive complex numerals, abstract-counting numerals should be used.

However, it is not expected that both numerical elements have to be in the abstract-counting form. This is because morphemes indicating the object-counting function may attach to the complex numeral as a whole. In such cases, we expect that morphology conveying the object-counting function will affect the edges of the complex numeral (e.g., by a phrasal affix), but not its internal composition.

Turning now to multiplicative complex numerals, Analysis A also leads us to expect abstract-counting numerals inside the complex. However, it is not *a priori* clear whether the constituency in (7b) necessarily leads to the obligatory occurrence of abstract-counting numerals; it could be that the numeral *three* counts *hundreds* in the same way as it counts other objects. If so, the constituency could be [[*three hundred*] *books*] even if *three* were an object-counting numeral. However, clearly, if *three* were in the abstract-counting shape, this would support the constituency in (7b).

This paper presents relevant examples from a database of 17 languages that distinguish abstract-counting and object-counting numerals in at least a subpart of their inventory. Our goal is to lay out the patterns according to the expectations just described. We leave it open as to whether the respective analyses may be combined with some additional assumptions that would change the basic expectations emanating from the different structures.

2. Possible patterns of complex numerals

Based on the reasoning in the preceding section, we have constructed a sample of 17 languages that make a distinction between object- and abstract-counting numerals in at least a part of their inventory of numerals. The languages come from twelve different language families:

- (9) Bangla, Irish, Romanian (Indo-European), Javanese, Puyuma (Austronesian), Swahili, Zulu (Niger-Congo), Mandarin, Shuhi (Sino-Tibetan), Telugu (Dravidian), Thai (Kra-Dai), Huehuetla Tepehua (Totonacan), Maltese (Afroasiatic), Palikur (Arawakan), Vietnamese (Austroasiatic), Japanese and Korean.

In this section, we discuss the patterns we found. Recall that (nothing else said), Analysis B predicts that complex numerals contain object-counting numerals only. We discuss such languages in Section §2.1 (there is only one such language in our sample, namely Irish). In Section §2.2, we discuss languages where complex numerals always contain at least one abstract-counting numeral. There are seven languages of this type in our sample: Mandarin, Korean, Japanese, Thai, Vietnamese, Telugu, and Bangla. Recall that these languages point in the direction of Analysis A. In Section §2.3, we discuss languages which mix these two patterns, i.e., where some complex numerals contain only object-counting numerals, while others contain at least one abstract-counting numeral (Maltese, Puyuma, Romanian, Javanese). Finally, in Section §2.4, we discuss languages with patterns that cannot be subsumed under the previous types for various reasons (Shuhi, Huehuetla Tepehua, Palikur).

Based on these results, we shall conclude that both types of structures seem to be attested in natural languages (a conclusion suggested also in Wągiel and Caha 2021: fn 12). We note, however, that our approach is mainly descriptive, taking the data at face value without any attempts to reinterpret them *per force* one way or another. At the same time, we do not want to preclude the option that this may be possible (see, e.g., Ionin and Matushansky 2018: 126–139).

2.1. Patterns of complex numerals with object-counting numerals only

This type of pattern can only be found in one language of the sample, namely Irish. The language belongs to Indo-European language family (Celtic), and it is spoken in Ireland. In Irish, simple abstract-counting numerals from ‘one’ to ‘ten’ are always preceded by an unstressed *a* (Dylon and Ó Cróinín 1961: 64). We give an example in (10a), glossing *a* as NBR (number). In (10b), we illustrate the fact that object-counting numerals lack the initial *a*.

- (10) Irish ((a) is adapted from Stenson 2020: Ch. 20.1; Dylon and Ó Cróinín 1961: 63, 137)

- a. *A dó agus a dó sin a ceathair.*
 NBR two plus NBR two are NBR four
 ‘Two plus two are four.’
- b. *trí cinn de bhuaibh*
 three heads of cow
 ‘three cows’

Let us now turn to complex numerals. We start with a multiplicative numeral in (11). What the example shows is that regardless of whether the numeral is used in a counting sequence, as in (11a), or as an object-counting numeral, as in (11b), there is no *a* preceding the numeral ‘three’ or ‘hundred.’ This means that the numeral *trí* is in the object-counting shape.

(11) Irish (Dylon and Ó Cróinín 1961: 136–137, Stenson 2020: Ch. 20.1)

- a. *trí chéad*
three hundred
'three hundred'
- b. *trí chéad bó mhór*
three hundred large cows
'three hundred large cows'

For additive numerals, our sources show variation. In Dylon and Ó Cróinín (1961: 136), additive numerals such as 'thirteen' are given bare, without the preceding *a*. However, in Stenson (2020: Ch. 20.1), the counting sequence is given with *a* preceding the complex numeral. This is reflected in (12a) by placing the *a* in parentheses.

(12) Irish (Dylon and Ó Cróinín 1961: 136–137, Stenson 2020: Ch. 20.1)

- a. (a) *trí déag*
NBR three ten
'thirteen'
- b. *trí bhád déag*
three boat ten
'thirteen boats'

However, the crucial data concerns object-counting use, see (12b). The first relevant fact is that the modified noun comes in between the two numerals. This fact alone is hard to reconcile with the idea that the complex numeral forms a constituent to the exclusion of the noun, a point that has been raised for Scottish Gaelic and Biblical Welsh in Ionin and Matushansky (2018: 125). The second relevant fact is that there is not a single NBR *a* in (12b): both numerals are in the object-counting shape.

In sum, Irish is a language where each simple numeral inside the complex is in an object-counting form. Irish is the only language in our sample with this property. This exceptional property of Irish could, perhaps, be linked to another exceptional property, which is that Irish is the only language in our sample where abstract-counting numerals (which are marked by *a* in Irish) are morphologically more complex than object-counting numerals.²

2.2. Patterns with abstract-counting numerals

In Bangla, Japanese, Korean, Mandarin, Thai, Telugu and Vietnamese, we find complex numerals that contain at least one abstract-counting numeral. According to the reasoning in

² See Wągiel and Caha (2021) for an empirical discussion of the morphological patterns. The authors explain the frequent pattern where object-counting numerals are derived from abstract-counting numerals by proposing that object-counting numerals are semantically derived from abstract-counting numerals by the addition of a dedicated meaning component. The Irish situation apparently goes against this proposal. At the same time, it must be mentioned that current morphological theories contain various ideas how the realisation of a complex structure may contain fewer morphemes than the realisation of a complex structure, see, e.g., Blix (2022) as well as Wągiel and Caha's (2020) treatment of the distinction between *ein* and *eins* 'one' in German.

Section §1, this is only compatible with Analysis A. We present here Mandarin and Vietnamese as two examples of such languages.

Let us begin by showing the distinction between simple abstract- and object-counting numerals in Mandarin (building mainly on He 2015). In (13a), we can see that the abstract-counting numeral for ‘two’ is *èr*.

(13) Mandarin (Po-Ching and Rimmington 2015: 26; He 2015: 198)

- a. *yī jiā yī děngyú èr.*
one add one equals two
‘One plus one is two.’
- b. *liǎng gè xuéshēng*
two CLF student
‘two students’

The object-counting form of ‘two’ in (13b) shows two differences. First, the classifier *gè* appears. Second, the numeral has the suppletive form *liǎng*. It is impossible to use *liǎng* in (13a), and it is impossible to use *èr* in (13b).

Let us now turn to complex numerals, starting with multiplicative complex numerals in (14). In (14a), we show an abstract-counting numeral, where *èr* ‘two’ is the multiplier of ‘ten’.

(14) Mandarin (He 2015: 190, 195)

- a. *Èr shí sān shì sùshù.*
two ten three be prime.number
‘Twenty three is a prime number.’
- b. *èr shí sān wàn gè xuéshēng*
two ten three 10,000 CLF student
‘230,000 students’

The same form (*èr*) also appears in the object-counting numeral in (14b). We can see that the object-counting status of the whole complex numeral is signaled by the classifier that always appears in between the numeral and the counted noun. However, the presence of the classifier does not influence the shape of the numeral ‘two’, which is internal to the complex numeral and remains in the abstract-counting shape *èr*. Further, there is no classifier in between ‘two’ (*èr*) and ‘ten’ (*shí*). This also suggests that ‘two’ is not counting ‘tens’ in the same way in which it is counting ‘students’. Mandarin multiplicative numerals thus provide (at least) two kinds of evidence in favor of a structure where the numeral first forms a complex constituent consisting of abstract-counting numerals. The two types of evidence are (i) suppletion of ‘two’ and (ii) the absence of classifiers inside multiplicative numerals.

Finally, let us discuss additive numerals. In (15a), we present the object-counting form of the numeral ‘fifty-two’. Interestingly, the numeral ‘two’ is expressed by *èr*, which is the abstract-counting shape, recall (13a). The object-counting *liǎng* cannot be used here, see (15b).

(15) Mandarin (He 2015: 198)

- a. *wǔ shí èr gè xuéshēng*
five ten two CLF student
‘fifty-two students’

- b. *wǔ shí **liǎng** gè xuéshēng
 five ten two CLF student
 Intended: 'fifty-two students'
- c. *èr gè xuéshēng
 two CLF student
 Intended: 'two students'

Importantly, example like (15a) cannot originate from a structure like [fifty students and two students]. If (15a) were to be derived from such a structure, the second member of such a hypothetical coordination would have to be the sequence *èr gè xuéshēng, but such sequence is ungrammatical, see (15c).

The remaining languages in our sample are classifier languages that show similar distribution of classifiers as Mandarin (but lacking suppletion). We group them together with Mandarin based on the fact that the classifier never appears inside the complex numeral.

As an example of such a language, we provide Vietnamese (Austroasiatic). Vietnamese is a classifier language with only a single classifier for the whole complex numeral. To illustrate the pattern, let us look at bare numerals such as *ba* 'three'. The bare form is used in the counting sequence (16), in arithmetic statements, see (17), and in complex numerals, see (18).

- (16) Vietnamese (Ngo 2020: 5)

ba
 'three'

- (17) Vietnamese (Wągiel and Caha 2021: 477)

Bốn nhân **hai** bằng **tám**
 four times two equal.to eight
 'Four times two is eight.'

- (18) Vietnamese (Ngo 2020: 6, 101)

- a. **mười ba**
 ten three
 'thirteen'
- b. **ba** **mười**
 three ten
 'thirty'
- c. **ba** **trăm**
 three hundred
 'three hundred'

Bare complex numerals are also used in arithmetic statements such as (19), something that we have also observed for Mandarin, recall (14a).

- (19) Vietnamese (Ngo 2020: 6, 101)

Mười lăm chia cho **ba** **mười** bằng không phẩy năm.
 ten five divided by three ten equal zero point five
 'Fifteen divided by thirty is zero point five.'

When numerals are used as object-counting numerals, a classifier appears after the numeral, see (20a).³ The example would be ungrammatical without the classifier.

(20) Vietnamese object-counting numerals (Ngo 2020: 21, Trang Phan, p.c.)

- a. **ba** *bức thư*
three CLF letter
'three letters'
- b. **mười ba** *bức thư*
ten three CLF letter
'thirteen letters'

Importantly, complex numerals require only one classifier, placed between the whole complex numeral and the noun, see (20b). Wągiel and Caha (2021: 480–482) interpret this to mean that the internal structure of such complex numerals contains abstract-counting numerals. The main indication is the absence of classifiers internally to the complex numeral.⁴

Patterns similar to Vietnamese are found also in Bangla (Biswas 2016: 94–95), Japanese (Wągiel and Caha 2020: 204), Telugu (Lisker 1963: 112), Korean (Lee and Ramsey 2000: 95–96), and Thai (Smyth 2002: 172–173, Wągiel and Caha 2020: 205).⁵

2.3. Non-uniform patterns

Taking the patterns seen in Sections §2.1 and §2.2 at face value, it seems that the structures proposed by Analysis A and B must both be available in individual languages. This conclusion is supported by the fact that sometimes we observe the two different structures even within a single language. We discuss such languages in this section.

As the first case, consider Zulu (Stuart 1940: 41–43). The relevant data set is in (21).

(21) Zulu (Stuart 1940: 41–3, 109)

- a. **ama-tunga** **ama-bili**
class.6-bucket CLASS.6-two
'two buckets'

³ The classifier is obligatory for some nouns, but may be optional with others, see Simpson and Ngo (2018), Phan (2019).

⁴ We assume that the numeral and the classifier form a constituent to the exclusion of the noun, following the ideas explored, e.g., in Krifka (1995), Bale and Coon (2014), He (2015), Sudo (2016), Cinque (2021), and Wągiel and Caha (2021). Under this view, the classifier serves as a device that turns the abstract-counting numeral into an object-counting one. Under this view, the fact that we do not find classifiers internally to the complex numeral leads us to conclude that the complex numeral contains abstract-counting numerals.

However, it must be mentioned that there is an alternative analysis of classifiers in the literature, according to which the classifier forms a constituent with the noun (excluding the numeral), as proposed, e.g., in Borer (2005), Chierchia (1998, 2010), Rothstein (2010), Li (2011), Scontras (2013). If that is so, one could hypothesize that the absence of the classifier is due to some other factor. For instance, for additive numerals, one could propose that ellipsis (or right rode raising) eliminates both the noun and the classifier in the first conjunct. As far as we can tell, this approach still cannot handle the data in (15), where the issue is not only the presence/absence of the classifier, but the fact that the numeral 'two' has a dedicated abstract-counting form in the complex numeral.

⁵ We are grateful to Priyanka Biswas and Inkie Chung for help with Bangla and Korean, respectively.

- b. *n-komo* *m-bili*
 CLASS.10-cow CLASS.10-two
 ‘two cattle’
- c. *Ngi funa osheleni aba* [*ishumi*] *na* [*m-bili*]
 I want shillings CLASS.2 ten and two
 ‘I want twelve shillings.’
- d. *Ngi funa [osheleni aba ishumi] n-* [*osheleni aba bili*]
 I want shillings CLASS.2 ten and- shillings CLASS.2 two
 ‘I want twelve shillings.’

(21a,b) show that Zulu numerals such as ‘two’ agree with the modified noun in class. In (21a), the numeral has the class 6 prefix, in (21b), it has the class 10 prefix, reflecting the class of the head noun. Different numerals in Zulu differ slightly in what kind of agreement they take depending on their precise numerical value, but we shall skip these details in the interest of space. The important thing is that in additive numerals, Zulu uses two strategies. The first one is illustrated in (21c). It consists in joining the numerals *ishumi* ‘ten’ and *-bili* ‘two’ by the conjunction *na-*, which leads to the emergence of a nasal prefix on the numeral *bili* ‘two’. This may be interpreted as the emergence of a default class 10 marker, recall (21b), but we shall not dwell on this. Importantly, there is only a single agreement marker *aba* preceding the whole numeral ‘twelve’, suggesting that it is the numeral as a whole what modifies the noun.

Interestingly, there is an alternative strategy, shown in (21d), with the noun ‘shillings’ occurring twice, which also leads to the emergence of two agreement markers. We conclude that the existence of two different structures within a single language supports the idea that UG allows for both types of structures. This hypothesis, recall, is independently suggested by the fact that we register different types of languages cross-linguistically.

There are more languages in our sample where complex numerals constitute a non-uniform category, and therefore cannot be easily subsumed either under Analysis A or Analysis B. These are Javanese, Romanian, Maltese and Puyuma. We shall now discuss these in more detail.

Javanese belongs to the Austronesian language family. It is spoken in Java and across the whole Indonesia. We focus on the numeral ‘two’, which has three different shapes. We give them in (22).

(22) Javanese ‘two’ (Robson 1992: 75–76)

- a. *ro*
 b. *lo-ro*
 c. *ro-ng*

The first shape *ro* can only be used in the counting sequence, and we, therefore, classify it as an abstract-counting numeral. The form *loro* in (22b) is ambiguous, and it can be used both as an abstract- and object-counting numeral. Finally, *rong* is only used as an object-counting numeral.

In additive complex numerals, we find the shapes *loro* or *ro*, depending on the numeral, see (23). Each numeral in (23) can be used both as an abstract- or object-counting numeral.

(23) Javanese (Robson 1992: 75–76)

- a. **ro-las**
two-teen
'twelve'
- b. **sawidak loro**
sixty two
'sixty-two'

In the numeral 'twelve' in (23a), we see the shortest form of 'two,' namely *ro*, which can only be used as an abstract-counting numeral, recall (22a). Therefore, it is impossible to analyze *ro-las* as a case where each of the simple numerals combines syntactically with a noun (as in Analysis B), because *ro* never combines with nouns as such (only *loro* does).

The numeral 'sixty-two' in (23b) contains the shape *loro*. This shape is ambiguous between the object- and abstract-counting meaning, and we cannot, therefore, conclude anything here.

In multiplicative numerals, we only find the object-counting shape *rong*, see (24a,b). The shape *rong* is unambiguously object-counting, see (24c).

(24) Javanese (Robson 1992: 75–77)

- a. **rong puluh**
two ten
'twenty'
- b. **rong atus**
two hundred
'two hundred'
- c. **rong kilo**
two kilo
'two kilos'

We thus conclude that multiplicative numerals contain an object-counting shape of the numeral, unlike what we see with the additive numeral in (23a). Javanese is therefore a non-uniform language, where additive numerals conform to Analysis A, while multiplicative structures conform to Analysis B.

We also analyze Romanian in this manner. In (25a), we can see the abstract-counting shapes of the numerals 'one' and 'two'. We can see that for 'one', this form is different from the form used in the object-counting use, see (25b). In the complex numeral 'twenty-one', we see the abstract-counting shape of 'one'. In addition, the counted noun following 'one' is in the plural, and a linker *de* appears; all of these facts are characteristic for higher numerals.

(25) Romanian ((a) Carmen Savu, p.c., (b-c) Ionin and Matushansky 2018: 130–131)

- a. **Unu plus unu fac doi.**
one plus one make two
'One plus one is two.'
- b. **un/*unu baiat**
one.M boy
'one boy'

- c. *două-zeci și unu de băieți*
 two.F ten.PL and one.M of boy.PL
 ‘twenty one boys’

However, the numeral ‘two’ uses the feminine form before the numeral ‘ten’ in (25c). This shape is classified by us as an object-counting shape, and we therefore conclude that Romanian is a mixed language: like Javanese, it shows some evidence for abstract-counting forms in additive numerals, and object-counting forms in multiplicative numerals.⁶

Let us now turn to Maltese. Maltese is an Afroasiatic language spoken in Malta. Below, we discuss the pattern of the numeral ‘four’. In (26a), we present its abstract-counting shape. In (26b,c), we see two different forms of the object-counting form *erba(t)*. The presence/absence of *-t* apparently relates to the phonological shape of the following noun.

(26) Maltese (Azzopardi-Alexander and Borg 1997: 221, 266, 270; Aquilina 1965: 119, 123)

- a. *tnejn għal tnejn erbgha*
 two times two four
 ‘Two times two is four.’
- b. *erbat irġiel*
 four men
 ‘four men’
- c. *erba’ soldi*
 four pence
 ‘four pence’

Note that the object-counting form in (26b,c) lacks *-gh-*: the presence/absence of this segment helps us identify the type of numeral. Let us now turn to complex numerals. The abstract-counting shape *erbgha* (truncated due to the loss of the final vowel) can be seen in the multiplicative numeral ‘forty’ in (27a). We show its object-counting use in (27b).

(27) Maltese (Azzopardi-Alexander and Borg 1997: 150, 266–267; Aquilina 1965: 119)

- a. *erbgh-in*
 four-ty
 ‘forty’
- b. *erbgh-in kilo*
 four-ty kilo
 ‘forty kilos’

⁶ See Ionin and Matushansky (2018: 129–131) for an analysis where the numerals in examples like (25c) are analyzed as ‘nominal’, rather than abstract-counting numerals. An anonymous reviewer further suggests that the morpheme *un* ‘one’ may be an indefinite article, rather than a numeral. However, note that *un* ‘one’ has uses suggesting a numeral status (at least as an option). For instance, it is possible to provide (i) as an answer to the question *How many boys did you see?*

(i) *Am văzut fix un băiat*
 have.1SG seen exactly one boy
 ‘I saw exactly one boy.’

Such modification would be impossible for an indefinite article (consider the impossibility of **I saw exactly a boy* as an answer to *How many boys did you see?*).

On the other hand, the complex numeral ‘fourteen’ (28a) contains the object-counting shape *erbat*. The object-counting shape of ‘fourteen’ adds *-il* to the abstract-counting numeral, see (28b). The important point is that in the object-counting use, both parts of the complex numeral are in the object-counting shape. This can be explained if the structure is as in (28c).

(28) Maltese (Azzopardi-Alexander and Borg 1997: 150, 266–267; Aquilina 1965: 119)

- a. ***erbat-ax***
four-teen
‘fourteen’ (abstract-counting)
- b. ***erbat-ax-il***
four-teen-il
‘fourteen’ (object-counting)
- c. [erbat N] [ax-il N]

The object-counting form *erba’* is also found in the multiplicative numeral ‘four hundred’. We present the abstract-counting shape in (29a) and the object-counting form in (29b). In (29b), *mitt* is the object-counting form of *mija* ‘hundred’.

(29) Maltese (Azzopardi-Alexander and Borg 1997: 150, 266–267; Aquilina 1965: 119)

- a. ***erba’*** *mija*
four hundred
‘four hundred’
- b. ***erba’*** *mitt* *liyra*
four hundred pounds
‘four hundred pounds sterling’

To summarize, the Maltese numeral ‘four’ provides us with mixed results: in some multiplicative numerals, recall (27), it is in the abstract-counting form, whereas in other numerals, see (28–29), the object-counting form is found. We found a similar system in Puyuma (Fang-Ching Teng 2007: 108–111).

Summarizing, in Sections §2.1 and §2.2, we saw that some languages use abstract-counting numerals in complex numerals, while other languages use object-counting numerals. In this section, we discussed languages where the two systems coexist. Sometimes they are apparently in free variation (Zulu), whereas in other cases they are fixed for a particular numeral type (Javanese, Romanian, Maltese, Puyuma).

2.4. An unexpected type

Finally, our sample also contains three languages where complex numerals contain component parts that are neither abstract-counting or object-counting numerals. One language like that is Palikur, an Arawakan language spoken in French Guyana and Brazil (Launey 2003). We start our discussion by presenting the object-counting forms of the numeral ‘one’ in (30).

(30) Palikur (Launey 2003: 121–122)

- a. ***paha-kti*** *pilatno*
one-CLF banana
'one banana tree'
- b. ***paha-t*** *pilatno*
one-CLF banana
'one banana'

The numeral root *paha* has a suffix that functions as a prototypical classifier and determines the type of object that is quantified over. When it is a 'banana tree', the numeral has the suffix *-kti* (used for arboreal entities). When it is a single banana, the suffix *-t* (for elongated objects and body parts) appears.

More classifiers are illustrated in (31), using the numeral 'two'. Note that with this numeral, classifiers are infixes. In (31a), we see a classifier for clothing and piles, *-rik*. In (31b), there is a classifier for flat objects, *-ka*.

(31) Palikur (Launey 2003: 121–122)

- a. ***pi-rik-na*** *kagta*
two-CLF-two book
'two piles of books'
- b. ***pi-ka-na*** *kagta*
two-CLF-two book
'two books'

Let us now turn to multiplicative numerals, see (32a) and (32b).

(32) Palikur (Launey 2001: 38, 45)

- a. ***pina*** *madikwa*
two ten
'twenty'
- b. ***pima-vut*** *sah*
two-times hundred
'two hundred'

The example in (32a) shows the bare numeral root *pina* 'two' (without any infix) as a multiplier of the numeral 'ten'. The absence of a classifier (which always accompanies object-counting uses) suggests that *pina* is not an object-counting numeral. Similarly, in (32b), the marker *-vut* is translated in the reference grammar as a multiplicative marker analogous to the English 'times'. We do not treat it as a classifier, since it does not infix into the numeral, and it attaches to what appears again as a bare root of the numeral. Therefore, the conclusion is that multiplicative numerals in Palikur do not contain object-counting numerals.

At the same time, they do not contain abstract-counting numerals either. The reason is that when numerals are recited in a sequence, they have a classifier, see (33), contrasting with (32).⁷

⁷ Palikur numerals *madikwa* 'ten' and *sah* 'hundred' take no classifiers and show, therefore, no difference between object-counting and abstract-counting forms. Numerals like '(one) hundred two' therefore have no

(33) Palikur (Launey 2003: 114)

- a. ***paha-t***
one-CLF
'one'
- b. ***pi-ta-na***
two CLF-two
'two'

Summarizing, we conclude that Palikur multiplicative numerals contain neither object- nor abstract-counting numerals.⁸

A similar situation is found in Shuhi (Qi and He 2019). In this language, numerals are always followed by a classifier both when they are used for abstract counting (34a) or for object counting (34b,c). The object-counting numerals take a variety of classifiers, compare (34b,c); for the abstract-counting (34a), the so-called default classifier must be used.

(34) Shuhi (Qi and He 2019: 66–69)

- a. ***dzi³⁵-ko³⁵-re³³*** ***dzi³⁵-ko³⁵-fi³³*** ***me³³-ba³³-le⁵⁵*** ***ne³³-ko³⁵ le³³-ziŋ³³-dzō³³***.
one-CLF-ABL one-CLF-LOC DIR-add-AUX two-CLF DIR-become-DUR
'One plus one is two.'
- b. ***jo³³ku⁵⁵*** ***dzi³³-pha³³***
stone one-CLF
'one lama'
- c. ***sa⁵⁵zə⁵⁵*** ***dzi³³-zə⁵⁵***
tree one-CLF
'one tree'

As a result, numerals never occur on their own; they are bound roots that always require a classifier or another morpheme. This is similar to Palikur, which also has a classifier both for object and abstract counting, recall (31) and (33). The only place where the classifier is not needed with numeral roots is in complex numerals. In these cases, numeral roots occur bare (35a) and only one classifier for the whole phrase is found (35b).

(35) Shuhi (Qi and He 2019: 66–69)

- a. ***dzi³³ ee⁵⁵***
one hundred
'one hundred'
- b. ***sa⁵⁵zə⁵⁵*** ***fi³⁵-qe³³*** ***ni⁵⁵*** ***se⁵⁵-zə⁵⁵***
tree five-ten CONJ three-CLF
'fifty-three trees'

An additional language that represents this type in our sample is Huehuetla Tepehua (Kung 2007: 480–482).

object-counting classifier on 'hundred,' and only have a classifier on 'two.' Thus, the only thing we only know is that the complex numeral contains one object-counting numeral. But since we cannot decide on the form of 'hundred' or 'ten,' we cannot assign such complex numerals to any type. The same is true for Swahili (Almasi et al. 2014: 183, 187–188).

⁸ There is an alternative form for 'two hundred,' the French-creole based form *de-sah*.

3. Conclusion

In the current literature, there are contrasting proposals concerning the structure of complex numerals. Under one analysis, the complex numeral is formed as a constituent first, and it is then used to quantify over objects denoted by the modified noun, see (36a) for an additive complex numeral. The alternative is that each numerical element within the complex numeral combines independently with the modified noun one copy of which is subsequently deleted, see (36b).

- (36) Additive complex numerals
 a. [20 + 3] cats → 23 cats
 b. [20 cats] + [3 cats] → 23 cats

The same two analyses can be advanced also in the domain of multiplicative complex numerals, see (37).

- (37) Multiplicative complex numerals
 a. [3 × 10] cats → 30 cats
 b. 3 × [10 cats] → 30 cats

This paper tried to shed some light on the question of which structures are used in languages by looking at the distinction between abstract-counting numerals, i.e., expressions referring to abstract arithmetical concepts, and object-counting numerals, i.e., forms that modify nouns in order to quantify over entities. Nothing else said, the analyses in (36b) and (37b) predict that complex numerals contain object-counting numerals. On the other hand, (36a) and (37a) predict that complex numerals contain at least one abstract-counting numeral.

Our investigation revealed that the structures in (36a) and (37a) are supported by data from Mandarin, Vietnamese and other classifier languages, where complex numerals contain abstract-counting numerals (§2.2). On the other hand, in Irish, complex numerals generally conform to the templates in (36b) and (37b), recall §2.1.

In yet other languages (we discussed Maltese, Javanese, and Zulu), the different types seem to be mixed within a single language (§2.3). For Shuhi, Palikur and Huehuetla Tepehua, we could not reach any definitive conclusion (§2.4).

Considered in their totality, the patterns found in our sample do not support the idea that languages use only one type of structure. Rather, it seems that both possibilities of forming complex numerals are employed, with languages choosing one or the other structure in ways that we do not fully understand. Quite likely, the investigation of other diagnostics (i.e., other than the abstract/object-counting distinction) may shed more light on this issue.

One thing we would like to emphasize again at the end is that our study revealed that the majority of complex numerals in our sample contain abstract-counting numerals. One could speculate that languages, therefore, prefer to use the abstract-counting structures whenever possible, and only if this strategy is not available for reasons to be understood, the more complex structure is used.

References

- Almasi, O., Fallon, M. D., and N. P. Wared. 2014. *Swahili Grammar for Introductory and Intermediate Levels*. Lanham: University Press of America.
- Aquilina J. 1965. *Teach Yourself Maltese*. London: The English Universities Press LTD.
- Azzopardi-Alexander, M., and A. Borg. 1997. *Maltese*. Abingdon: Routledge.
- Bale, A., and J. Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry* 45(4): 695–707.
- Biswas, P. 2016. *Number marking and definiteness in Bangla*. Doctoral dissertation, University of Southern California.
- Blix, H. 2022. Interface Legibility and Nominal Classification: A Nanosyntactic Account of Kipsigis Singulatives. *Glossa: A journal of general linguistics* 7(1).
- Borer, H. 2005. *Structuring Sense I: In Name Only*. Oxford: Oxford University Press.
- Bultinck, B. 2005. *Numerous meanings: The meaning of English cardinals and the legacy of Paul Grice*. Amsterdam: Elsevier.
- Chierchia, G. 1998. Plurality of mass nouns and the notion of ‘semantic parameter’. In S. Rothstein (ed.), *Events and Grammar*, 53–103. Dordrecht: Kluwer.
- Cinque, G. 2021. *On Linearization. Towards a restrictive theory*. Ms., University of Venice.
- Dylon, M., and D. Ó Cróinín. 1961. *Teach Yourself Irish*. London: Hodder and Stoughton.
- Ionin, T., and O. Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23(4): 315–360.
- Ionin T., and O. Matushansky. 2018. *Cardinals: The Syntax and Semantics of Cardinal-Containing Expressions*. Cambridge, MA: MIT Press.
- He, C. 2015. Complex numerals in Mandarin Chinese are constituents. *Lingua* 164: 189–214.
- He, C., and O.-S. Her. to appear. Complex numerals in English: Constituents or not? *English Language and Linguistics*.
- Krifka, M. 1995. Common nouns: A contrastive analysis of Chinese and English. In G. N. Carlson, and F. J. Pelletier (eds.), *The Generic Book*, 398–411. Chicago: University of Chicago Press.
- Kung, S. S. 2007. *A Descriptive Grammar of Huehuetla Tepehua*. Austin. PhD thesis, University of Texas at Austin.
- Launey, M. 2003. *Awna parikwaki. Introduction à la langue palikur de Guyane et de l'Amapa*. Paris: Institut De Recherche Pour Le Développement.
- Lee, I., and R. Ramsey. 2000. *The Korean Language*. Albany: State University of New York.
- Li, X. 2011. *On the Semantics of Classifiers in Chinese*. Ph. D. thesis, Bar-Ilan University, Ramat Gan.
- Lisker, L. 1963. *Introduction to Spoken Telugu*. New York: American Council of Learned Societies.
- Ngo, B. 2020. *Vietnamese: An Essential Grammar*. Abingdon: Routledge.
- Po-Ching, Y., and D. Rimmington. 2015. *Chinese: A comprehensive grammar*. London and New York, NY: Routledge.
- Phan, T. 2019. The absence of classifiers in numeral classifier constructions in Vietnamese. In M. Bağrıaçık, A. Breitbarth, and K. De Clercq (eds.), *Mapping Linguistic Data*, 208–213. Online publication (<https://biblio.ugent.be/publication/8625919/file/8625920.pdf>).
- Qi, J., and C. He. 2019. The morphosyntax of numerals *dzi33/dzi35* ‘one’ in Shuhi and implications for the semantics of numerals. *Lingua* 225: 63–80.
- Robson, S. 1992. *Javanese Grammar for Students*. Melbourne: Monash University.
- Rothstein, S. 2010. Counting and the mass/count distinction. *Journal of Semantics* 27(3): 343–397.
- Rothstein, S. 2017. *Semantics for counting and measuring*. Cambridge: Cambridge University Press.
- Scontras, G. 2013. Accounting for counting: A unified semantics for measure terms and classifiers. In T. Snider (ed.), *Proceedings of Semantics and Linguistic Theory* 23, 549–569. Ithaca, NY: CLC Publications.
- Simpson, A., and B. Ngo. 2018. Classifier syntax in Vietnamese. *Journal of East Asian Linguistics* 27(3): 211–246.
- Smyth, D. 2002. *Thai. An Essential Grammar*. London: Routledge.
- Stenson, N. 2019. *Modern Irish: A comprehensive grammar*. Oxon: Routledge.

- Stuart, P. A. 1940. *A Zulu Grammar for Beginners*. Pietermaritzburg: Shuter & Shooter.
- Sudo, Y. 2016. The semantic role of classifiers in Japanese. *Baltic International Yearbook of Cognition, Logic and Communication* 11(1): 1–15.
- Tatsumi, Y. 2021. *Linguistic Realization of Measuring and Counting in the Nominal Domain: A Cross-linguistic Study of Syntactic and Semantic Variations*. PhD dissertation, University of Connecticut.
- Wągiel, M., and P. Caha. 2020. Universal semantic features and the typology of cardinal numerals. *Catalan Journal of Linguistics* 19: 199–229.
- Wągiel, M., and P. Caha. 2021. Complex simplex numerals. *Acta Linguistica Academica* 68(4): 470–515.