The notion of the numeral base in linguistics

Russell Barlow
16 February 2024

Abstract:

This paper details how the notion of the numeral base has been used by linguists, both in the description of individual languages and as a broader concept for comparing different linguistic numeral systems. Although common linguistic understandings of "base" do not always (or ever) equal mathematical definitions of the term, they may nevertheless be useful for understanding how different languages encode numbers differently. In the hopes of supporting efforts in comparative linguistics as well as facilitating interdisciplinary studies of numeral systems, this paper presents a typology of the various phenomena that may all be considered to be a "base".

1 Introduction: The use of *base* in descriptive linguistic literature

Linguists have long used terms like *base*, *decimal*, *quinary*, and *vigesimal* in describing linguistically encoded numeral systems. These terms are rarely defined in the descriptive linguistic literature, but often reflect some shared understanding by linguists of what is important in the study of linguistic means of referring to exact quantities. Since the use of these terms in the linguistic literature does not always match that of other fields, it is valuable here to summarize what linguists generally understand by *base*, as well as some associated terms.

In short, linguists have often used the term *base* in a broader sense than is used in mathematics and other fields. Comrie's (2005b: 207) definition of *base* conveys the general linguistic understanding of the term: "that numerical value to which various arithmetic operations are applied". Indeed, the term is often used even more broadly than this, since not all authors require of a *base* that *various* arithmetic operations are applied: that is, while a prototypical base would be involved both in addition and in multiplication, base systems are sometimes defined on the basis of addition alone. Perhaps the broadest general linguistic understanding of *base* can be summarized as:

> *A numerical value that is used systematically in the formation of numerals.*

The term *numeral* is to be understood here in a linguistic context: a numeral is a word that refers to an exact quantity. This description specifies the *systematic* use of the base for two reasons. First, language users may have various means of referring to quantities, including ad hoc or idiosyncratic formulations. However, only expressions that are conventionalized should be considered relevant for describing linguistic numeral *systems* and thus the *bases* that they employ.[1] Second, in order to be considered a *base*, the numerical value in question should be used with some degree of systematicity (or regularity) within the system. For example, a linguist would probably not describe a system as "base-5" if the number 5 is used in the

[1] Note, however, that linguistic descriptions of "numeral systems" are often written with little consideration of conventionalization. Sometimes the only information available on the numerals in a given language is derived from a single elicitation session in which a linguist, anthropologist, or missionary has queried a single speaker. It is thus not always possible to know the extent to which any putative system is conventionalized within a language. On "conventionalization" in numeral systems, see von Mengden (2008: 291–292).

formation of only *one* numeral other than 5.[2] At the same time, however, complete and utter regularity is rare within the grammatical domains of natural language, so most linguists would probably accept a system as "base-5" even if there is one exception to the pattern.[3] Thus, terms like *systematic* and *regular* should be understood as scalar concepts. The degree of systematicity or regularity required for identifying a *base* may vary according to author or description. Hammarström (2008: 291–292; 2010: 15) takes a systematic approach to defining this systematicity:[4]

> The number *n* is a base iff
> 1. the next higher base (or the end of the normed expressions) is a multiple of *n*; and
> 2. a proper majority of the expressions for numbers between *n* and the next higher base are formed by (a single) addition or subtraction of *n* or a multiple of *n* with expressions for numbers smaller than *n*.

This definition is more restrictive than some commonly used interpretations of *base* in the linguistic literature in that it requires a base to have special designations occurring at one or more multiples of itself. This is also requried of mathematical definitions of *base*, in which it is further required that one or more powers/exponents of the base are specially defined. In linguistic terms, we would thus expect a base-5 system to have a word for 5^2 (i.e., 25) that is underived by any other number word, much like in base-10 systems, such as English, there is an underived word for 10^2 (i.e., 100; e.g., *hundred*). However, such mathematically defined base-5 systems are essentially unattested among natural languages. Thus, commonly described "quinary" systems – provided they have conventionalized expressions for sufficiently large numbers – tend to switch to using 10 or 20 (or both) as elements for building numerals higher than 10 or 20.[5]

Several linguists have acknowledged this incongruence between the mathematician's definition and that of many linguists. For example, Lynch (2009: 392:fn3) writes:

---

[2] E.g.: {1, 2, 3, 4, 5, 6, 7, **5+3**, 9, 10, 10+1, 10+2, 10+3, …}. Occasionally, however, the term "base" is indeed used for cases such as this. See §3 for terminological suggestions.

[3] E.g.: {1, 2, 3, 4, 5, 5+1, 5+2, **8**, 5+4, 2·5, 2·5+1, 2·5+2, 2·5+3 …}.

[4] Cf. Schapper & Hammarström's (2013: 424) definition of base-5 systems as those in which more than half (i.e., ≥ 3) of the expressions for 6 through 9 are formed with 5.

[5] Thus, in such a "quinary" system, the number 25 is more likely expressed as something like 2·10+5 or as (1·)20+5. McElvenny (2006: 30) reports that 26 in Mundukumo (Yuat family, Papua New Guinea) can be expressed as 5·5+1, thereby implying that 25 would be expressed as 5·5, but this may not match earlier reports for the language (see Hammarström 2010: 32–33). The neighboring but unrelated language Ulwa (Keram-Ramu family, Papua New Guinea) is similarly recorded as being able to express 25 as 5·5 as an alternative to 10·2+5 (Barlow 2023b: 247), but it is not known whether it is possible for the system to continue past 25 along the same "quinary" lines. For the language Gumatj (Pama-Nyungan family, Australia) Harris (1982: 169–173) records 25 as 5·5 (albeit composed of two different morphemes meaning 'five'), but he raises the question that the system recorded was a novel creation of the language consultant. Dixon & Kroeber (1907: 689; cf. Kroeber & Grace 1960: 120) report for Luiseño (Uto-Aztecan family, United States) that 30, for example, may be formulated as 'five-times five, five upon'. However, no Luiseño forms are given, and this claim, taken from Mr. P. S. Sparkman, has not been substantiated by other attestations of the language (Comrie 2005b: 213). Hammarström (2010: 38:fn14) mentions two other languages with possibly similar attestations. Still, we know of no language with a putative base-5 system attested as having a term for 25 that would be analogous to a base-10 system's treatment of 100 (e.g., as *hundred*, as opposed to *ten-ten*).

> Lincoln (in press) notes that so-called quinary systems are not really quinary or base–5 in the mathematical sense, since the 'milestones' are not 5, $5^2 = 25$, $5^3 = 125$, etc. However, this term has been so widely used in Oceanic studies that I will retain it here.[6]

Thus, although terms like *base-5* and *quinary* are commonly used in describing linguistic numeral systems, it seems to be the case – by the mathematical definition of *base* – that no natural human language makes regular and exclusive use of a conventionalized base-5 system. Putative cases of base-5 systems generally rather reflect what might be considered *sub-bases* of 5 (e.g., a base-10 system with a sub-base of 5, or a base-20 system with a sub-base of 5).

Similarly, linguists sometimes use the term *binary* to refer to systems that use the number 2 as an element for constructing higher numerals, despite not treating 2^2 (i.e., 4) as the next higher base. In New Guinea, for example, where such systems are relatively common, it is unusual for people to use these "binary" means for counting higher than 4 or 5. Often there are no conventionalized terms for such higher numbers. Sometimes reference is made to a word meaning 'hand' for numbers ≥ 5. For example, Holzknecht (1989: 127) refers to the Markham languages (Austronesian family, Papua New Guinea) as employing "binary number systems", in reference to which Ross (2023: 523) notes:

> However, it is not strictly correct to call this system "binary", as a binary system requires that a new base intervenes at 4. The concept of a "base" requires that the next higher base (or the highest conventional numeral) be a multiple of the lower base, and 5, the next higher base, is not a multiple of 2 (but 4 would be). Thus 2 is not a base, but simply an element from which 3 and 4 are built in each quinary round (Hammarström 2008:291–292).

2 Discussions of *base* in typological linguistic literature

Thus there is some dissonance between how linguistic typologists tend to use the term *base* when discussing numeral systems generally and how descriptive linguists often use terms like *base*-n or n-*ary* when describing the systems of individual languages.

In the typological literature, a commonly cited definition for *base* is that of Comrie (2005a: 530; 2013): "the value n such that numeral expressions are constructed according to the pattern $xn + y$, i.e. some numeral x multiplied by the base plus some other numeral". Hanke (2010: 69), citing Comrie (2005a), refers to this sort of numeral as an *additive-multiplicative base*. Other mentions of this definition can be found in Plank (2009: 341), Bowern & Zentz (2012: 140), Epps et al. (2012: 58), and Barlow (2023a: 290).

On the other hand, the descriptive uses of terms like *binary*, *quinary*, *decimal*, and so on tend to be more in keeping with the notion of *cycles*, as put forward by Salzmann (1950: 81):

> The *cyclic* pattern is a succession of morphemes or groups of morphemes according to which the numerical system is analyzable in terms of one or more similar or regular sets

[6] Cf. Lincoln (2010: 231): "Most languages have counting words from 1 to 10 and start to paraphrase at 11, which is typically some compound of 1 and 10. Some other languages start at 6, typically some compound of 1 and 5. In the following sections I will avoid referring to the former as a 'base ten' or 'decimal' system and the latter as a 'base five' or 'quinary' system because these mathematical terms have different implications. In the mathematical sense, a system starts over again at multiples of the base. A decimal system begins using two digits at 10, three digits at the base times the base 100 (10^2), four digits at the base times the base times the base 1000 (10^3) and so on. A base five system changes to two digit mode at 5, to three digits at what we call 'twenty-five' (5^2), and changes to four digits at what we call 'one hundred and twenty-five' (5^3)."

> of recurring morphemes or groups of morphemes. This pattern covers all systems that are referred to as binary, ternary, ... decimal, etc. By no means do we intend to abandon the use of these terms; however, it is suggested that they be used in structural analysis with care, since it often happens that a cycle does not function consistently thruout a system, being either modified or changed. Also, as has been shown above, the comparative value of a cycle is considerably limited.

Other references to Salzmann's (1950: 81) use of *cycle* can be found in Greenberg (1978: 260), Owens & Lean (2018: 20), Comrie (2022b: 150:fn3), Ross (2023: 518), and Barlow (2023a: 290).

Plank (2009: 341) makes a useful distinction between what he calls a *construction-base* and what he calls a *cycle-base*:

> Numerals are frequently referred to as "bases" when they are an atomic expression (or at any rate not transparently compositional, synchronically speaking) and when expressions for other numerals are formally based on them, with higher or lower numerals constructed by arithmetic operations (addition, subtraction, multiplication) with their help. The requirement of such a construction-base being itself atomic is sometimes waived … A cycle-base is the narrower concept: a numeral is a cycle-base if it is a construction-base and cyclically recurs in linguistic designations of multiples of the respective number … and/or in exponentiation with that base …

3 A typology of "bases"

Plank's terminological distinction does a nice job reflecting the two broad uses of the term *base* in the linguistic literature, although each term is capable of being more or less strictly defined. These two terms seem valuable, and I propose that we adopt them here as well. First, I would define *construction base* as follows:

Construction base: A numerical value to which an arithmetic operation is applied so as to form a numeral.[7]

e.g., 5 in 5+3 (= 8)
e.g., 10 in 10−2 (= 8)
e.g., 4 in 2·4 (= 8)
etc.

[7] The following question may be raised: when there are two numerical values represented in an expression, how do we know which is the value "to which" the arithmetic operation is applied? For example, why would we claim that 4, rather than 2, is the construction base in 2·4? Often the decision can be made according to system-internal evidence. Imagine, for example, that this 2·4 construction is part of the following system: {1, 2, 3, 1·4, 4+1, 4+2, 4+3, 2·4, 2·4+1, 2·4+2, 2·4+3, 3·4, …}. In this case we would have two clues that 4 is a construction base. First, the operation of multiplication occurs not simply one time with 4, but repeatedly, for all multiples of 4. Although 2 is used in multiplication, it is done so only once in this series, as are 2 and 3. Second, 4 occurs as a construction base not when multiplication occurs but also when addition occurs, again regularly. Furthermore, a possible heuristic may be that whenever exactly two numerals occur in a complex numeral expression, the large is probably the construction base. Finally, not every combination of multiple numerals in a complex numeral expression necessarily contains a construction base. So-called "pairing" methods do not seem to be to be profitably analyzable as containing a construction base. Consider, for example, the following system: {1, 2, 3, 4, 5, 3+3, 1+3+3, 4+4, 4+4+1, 5[+]5, …}. In this system, 6, 8, and 10 are formed by pairing (or doubling) like numeral elements. There is no system-internal evidence for choosing either of the two 3s that comprise 6 as a construction base. Perhaps they should both be considered construction bases.

It may be useful to identify and define several subcategories of *construction base*, such as the following:

Regular construction base: A numerical value n that is regularly used in the formation of numerals, with "regularly" being defined as $\geq n \div 2$ numerals other than n (within a single cycle).

e.g., 2 in {1, 2, 3, **2+2**, 5, …} (2 occurs in 1 other numeral)
e.g., 3 in {1, 2, 3, **3+1**, **3+2**, 5+1, 5+2, …} (3 occurs in 2 other numerals)
e.g., 4 in {1, 2, 3, 4, 5, **4+2**, **4+3**, 8, …} (4 occurs in 2 other numerals)
e.g., 5 in {1, 2, 3, 4, 5, 6, 7, **5+3**, **5+4**, **2·5**} (5 occurs in 3 other numerals)
e.g., 6 in {1, 2, 3, **6−2**, **6−1**, 6, **6+1**, 2·4, 2·4+1, 2·4+2, 2·4+3, …} (6 occurs in 3 other numerals)

Canonical construction base: A numerical value n that is used in the formation of all integer numerals between n and $2 \cdot n$, inclusive. A *canonical construction base* is a special type of *regular construction base*.[8]

e.g., 2 in {1, 2, 2+1, 2+2, …}
e.g., 3 in {1, 2, 3, 3+1, 3+2, 2·3, …}
e.g., 4 in {1, 2, 3, 4, 4+1, 4+2, 4+3, 4+4, …}
e.g., 5 in {1, 2, 3, 4, 5, 1+5, 2+5, 3+5, 4+5, 5·2, …}
e.g., 6 in {1, 2, 3, 4, 6−1, 6, 6+1, 6+2, 6+3, 6+4, 6+(6−1), 6+6, …}

Sporadic construction base: A numerical value n used in the formation of at least 1 numeral other than n, but less than $n \div 2$ numerals other than n (within a single cycle).

e.g., 5 in {1, 2, 3, 4, 5, 6, 7, 5+3, 5+4, 10, …}
e.g., 6 in {1, 2, 3, 4, 5, 6, 6+1, 6+2, 9, 10, …}

Hapax construction base: A numerical value n used in the formation of only one numeral other than n (within a single cycle). A *hapax construction-base* is, generally, a special type of *sporadic construction-base*.[9]

e.g., 3 in {1, 2, 3, **3+1**, 4, 5, 5+1, 5+2, …}
e.g., 5 in {1, 2, 3, 4, 5, 6, **5+2**, 7, 8, 9, 10, 10+1, 10+2, …}

Thus, all construction bases are either regular or sporadic. A subset of regular construction bases are canonical construction bases. A subset of sporadic construction bases are hapax construction bases. (However, 2 as a construction base may be both hapax and regular.)

Terms like *augend*, *minuend*, *multiplicand*, and *dividend* can be expressed in terms of various types of *construction bases*:

[8] Note that although the formation of $2 \cdot n$ requires n as an element in its formation, it need not employ the same arithmetic operation as used in forming lower numerals. For example, n may function as an augend in the numerals $n+1$ through $2n-1$, but as a multiplicand in the numeral $2 \cdot n$.
[9] It is possible, however, to have a hapax construction base that is not sporadic, but is rather regular, as in the example already given of the regular construction base 2 in {1, 2, 3, **2+2**, 5, …}. A similar case of a hapax regular construction base is the construction base 2 in {1, 2, **2+1**, 4, 5, …}. In short, the categories "hapax" and "regular" can only overlap in the case of a construction base of 2.

Additive construction base = augend
Subtractive construction base = minuend
Multiplicative construction base = multiplicand
Divisive construction base = dividend
Exponential construction base = [exponential] base

Following Plank's (2009: 341) terminological distinction, a definition for *cycle base* may be formulated as follows.

Cycle base: A regular construction base to which an arithmetic operation is applied so as to form another regular construction base.

The *cycle base* is thus closer to the mathematical notion of a base. In practice, the arithmetic operation used in regular construction bases is almost always – if not always – *addition*; and the (secondary) arithmetic operation used in forming one or more other regular construction bases is almost always – if not always – *multiplication*. Thus, the notion of *cycle base* is, in practical terms, synonymous with the notion of *additive-multiplicative base* (cf. Hanke 2010: 69). However, a cycle base could in theory represent an *additive-additive base* (in other words, an *additive cycle base*). Subcategories of *cycle bases* could include the following:

Additive cycle base:

e.g., 5 in {1, 2, 3, 4, 5, 5+1, 5+2, 5+3, 5+4, 5+5, 5+5+1, 5+5+2, 5+5+3, 5+5+4, 5+5+5, …}

I question the degree to which such additive cycle bases are used in widespread conventionalized systems. Although recorded (to some degree) for some languages with a construction base of 5 (e.g., Southwest Tanna, Austronesian family, Vanuatu), these formulations might be ad hoc.[10]

Multiplicative cycle base:

e.g., 5 in {1, 2, 3, 4, 5, 5+1, 5+2, 5+3, 5+4, 2·5, 2·5+1, 2·5+2, 2·5+3, 2·5+4, 3·5 …}

This is the "canonical" cycle base. Linguistic numeral systems that employ a multiplicative cycle base are the systems most widely accepted – both by linguists and by researchers in other disciplines – as representing ("true") base systems.

Exponential cycle base:

[10] In Chan et al. (2019) – citing personal communications from John Lynch (1998) and Ken Nehrbass (2009) – the following forms are given for Southwest Tanna: *kɨlkɨlip* '5', *kɨlkɨlip kɨlkɨlip* '10', *kɨlkɨlip kɨlkɨlip kɨlkɨlip* '15', and *kɨlkɨlip kɨlkɨlip kɨlkɨlip kɨlkɨlip* '20'. The description of Southwest Tanna is, however, somewhat different in Lynch (1982: 89), in which a ligature *mi* is said to be used after the form *kɨlkɨlip* in forming longer numerals and the word for '20' is described as containing a word meaning 'person'. In addition to possible examples of 5 being used as an additive cycle base, there are some examples of systems with a construction base of 2 being recorded as having additive cycle bases. For example, Orop (or Arop-Sissano, Austronesian family, Papua New Guinea) is recorded as encoding 10 as 2+2+2+2+2 (Chan et al. 2019, citing a personal communication from John Nystrom and Velma Foreman, 1988). Although formulations such as these (and occasionally even higher iterations of addition by 2) are recorded in wordlists, I question the degree to which speakers use them in natural discourse (or as part of a conventionalized system) (cf. Barlow 2023: 328).

e.g., 5 in {1, 2, 3, 4, 5, 5+1, … 2·5 … 3·5 … 4·5 … 5^2, 5^2+1 …}

Exponentiation may not exist in any language as something distinct from multiplication. Exponentiation, when represented in any form, is generally opaque, as in English *hundred* for *ten-to-the-two*.

The numeral systems of some languages employ morphemes that transparently represent their additive (construction) bases or their multiplicative (cycle) bases. For example, in the Mandarin numeral *shí liù* '16' (literally '10 [+] 6'), the additive construction base is the element *shí* '10'. In the Mandarin numeral *sān shí* '30' (literally '3 [·] 10'), the multiplicative cycle base is (also) the element *shí* '10'. And in the Mandarin numeral *sān shí liù* '36' (literally '3 [·] 10 [+] 6'), the element *shí* '10' serves as a base simultaneously for addition and for multiplication. However, I know of no language in which exponentiation is overtly indicated. For example, in Mandarin, 10^2 is expressed as *bǎi* 'hundred', which is an opaque term – that is, it contains neither the exponential cycle base *shí* '10' nor the exponent (or power) *èr* '2'.[11]

4 A note on modalities

The focus of discussion in this paper has been on linguistic numeral systems – in particular, on spoken (or signed) linguistic systems. Written linguistic systems could also be covered by the same categories used here, provided they are truly linguistic systems (e.g., phonetic, phonemic, or morphemic representation of a spoken or signed system) and not simply *notational* systems (e.g., Arabic numerals used without carrying any linguistic information). The use of the human body as a modality could also be covered, again provided the body is being used as part of a linguistic system (i.e., in a sign language), and is not instead a form of (non-linguistic) gesturing or tallying. Finally, material-based systems, such as abacuses or tally sticks, are, as far as I am aware, never a part of linguistic systems and can simply be excluded from the present discussion entirely.[12]

However, the boundaries between "linguistic" and "non-linguistic" systems are not always clear, especially since speech (or signing) can be used to refer to non-linguistic behavior. Some cultures that do not employ conventionalized linguistic numeral systems for referring to higher numerals may nevertheless count by using non-linguistic modalities, such as digit-tallying. A speaker from such a culture could refer indirectly to an exact quantity by means of reference to such a digit-tallying practice (e.g., 'a hand and put two fingers atop it' to mean '7'). However, only when such a linguistic expression becomes fully conventionalized (i.e., there is little to no variation in how members of a speech community would refer to a given number) could it be considered part of a linguistic *system* (e.g., if a speaker from this community could just as well refer to '7' as 'one full hand and two from the other', then there is probably not a fully conventionalized expression for this number). In practice, given such a possible historical pathway from tally-based physical practices to a linguistic system, this would probably entail some degree of lexicalization – that is, a more descriptive phrase would over time reduce in length and fossilized as a single (perhaps multimorphemic) word.

One very interesting body-based numeration system is the "body-part tallying system", which is attested only in New Guinea and Australia. It is a method of counting in which people point to different body parts to represent different numbers, generally starting by pointing to a finger of one hand, then proceeding through the other four fingers, then pointing to points along

[11] In our Mandarin example, it seems difficult to construct 100 this way without introducing yet another morpheme. The string of morphemes *shí-èr* '10-2' would be interpreted as 12 (i.e., '10+2'), and the string of morphemes *èr-shí* '2-10' would be interpreted as 20 (i.e., '2·10').

[12] See Bender & Beller (2018: 300) for discussion of modality in numeration systems.

the arm, shoulder, head, and so on, usually continuing in symmetrical fashion, pointing to various points moving downward along the other side of the body. Insofar as the various points on the body have linguistic expressions associated with them, a member of a community that uses such a body-part tallying system could refer indirectly to a number by means of reference to a give body part (e.g., 'elbow' to mean '7'). Again, only if such references become conventionalized lexical expressions should they be considered part of a linguistic system. Although Comrie (1999: 83) describes speakers of Haruai (Piawi family, Papua New Guinea) as using body-part terms as integrated linguistic expressions, most descriptions of body-part tallying systems suggest that speakers do *not* use body-part terms for numerical references (e.g., by saying something like 'I saw elbow pigs' [= 'I saw seven pigs']). Laycock (1975: 219) notes: "There are no 'numerals' in a tally system, so that one may not receive a reply to the question 'how many?', or find the points of the tally-system qualifying nouns, as do true numerals." Hammarström (2010: 12–13) further points out that, where we have more detailed social information on communities that use body-part tallying systems, the systems are used only in special ceremonial circumstances and the tallying has to be conducted on a physically present body.

Nevertheless, "body-part tallying systems" have been included in (otherwise purely linguistic) typological studies and typological databases. For example, "extended body-part systems" are included in the *WALS* chapter on "numeral bases" (Comrie 2005a, 2013), where they are discussed as examples of linguistic numeral systems that lack arithmetic bases. Four languages are coded as having such systems. Similarly, *Grambank* (Skirgård et al. 2023) includes a feature "Is there a body-part tallying system?" (GB336). In version 1.0, there are 44 languages that are coded as having this feature. (A number of languages in the sample that have been incorrectly coded, included all 11 languages outside of New Guinea, none of which actually have attested body-part tallying systems.) Barlow (2023a: 331–333) found 100 languages of mainland New Guinea and one language of the Torres Strait Islands for which there was evidence of their speakers using body-part tallying systems. That study, however, did not treat body-part tallying as a linguistic system.[13]

5 Conclusion

xxx

xxx
Stuff on sporadic bases:

Several other lower-number bases are theoretically possible but are unattested. For example, no language is known to use base-1 (unary) counting (e.g., 2 is nowhere attested as being constructed as '1+1'), although certain notational practices, such as some tally marks, could be said to be unary. Other lower numbers, like 7 and 9, are likewise nowhere attested as being used as regular construction bases. However, complex numerals may contain such numerical values as sporadic elements in their construction. For example, although the numeral system in Makasar (Austronesian, Indonesia) is basically decimal, 8 is formed as '1 with 7' (and 9 derives from '1 taken [from 10]') (Jukes, 2020). Similarly, a sporadic "septenary" element can be seen in Tschamba-Daka (Atlantic-Congo, Nigeria), in which 8 is formed from 7 plus an element *kēròrō* (and, similarly to Makasar, 9 takes the form *wamiti kum* '1 [subtracted

[13] The study also did not investigate speaker communities in mainland Australia, for which there is also evidence of body-part tallying systems – namely, Kulin languages (Pama-Nyungan family, Australia); see Howitt (1889: 317–318;1904: 697– 703; Hammarström 2010: 13).

from] 10’) (Strümpell, 1910). The numerals 9 and 10 were apparently later reanalyzed such that the form for 10 *kum* came to signify 9 (*kúūm* ‘9’) and the element *kērōrō* came to be used as ‘+1’ in a new for 10 (*kúūm kə́rə́rə́* ‘9+1’) (Boyd, 1989), thereby exemplifying a sporadic “nonal” element. See also Hammarström (xxx) for discussion of putative attestations of highly unusual base values.

Appendix: Some definitions of “base” in the linguistic literature:

Conant (1896: 102):

In the development and extension of any series of numbers into a systematic arrangement to which the term *system* may be applied, the first and most indispensable step is the selection of some number which is to serve as a base. …

In the selection of a base,—of a number from which he makes a fresh start, and to which he refers the next steps in his count,—the savage simply follows nature when he chooses 10, or perhaps 5 or 20. But it is a matter of the greatest interest to find that other numbers have, in exceptional cases, been used for this purpose.

i.e., a base is: “… a number from which [one] makes a fresh start, and to which [one] refers the next steps in [one’s] count …”

Stampe (1976: 601):

The *base* number of a number system, then, must be defined as that number from which counting starts over. In the vast majority of languages, at least originally, it is the highest of the simple numbers.

Greenberg (1978: 269–270):

A serialized multiplicand is a number whose successive multiplication by at least two other numbers results in serialized products which are either expressed as simple lexemes or as a product of the multiplicand and multiplier, and such that each serialized product is also a serialized augend or minuend. … Incidentally, in defining serialized multiplicand, we have also defined the notion of base which up to now has been the sole method of typologizing numeral systems. A serialized multiplicand is a base. Since both multiplication and addition are involved in this definition, a system without these operations cannot have a base. There can be, however, and commonly is, more than one base, e.g. 10; 100; 1000; 1,000,000 in ENGLISH. The smallest base will be called the fundamental base. If all the bases are powers of the fundamental base, the system will be called “perfect.” There are only four numbers which figure as fundamental bases in perfect numeral systems of the world in order of frequency: 10, 20, 4, and 12. Most systems with 20 as a fundamental base have 100 as the next highest base rather than $400 = 20^2$.

Greenberg (2000: 774):

Using concepts of serialized augend and power it is possible to define a “pure” system for a specific base. This has two requirements: every serialized augend must be a multiple of the base (including 1 as a multiple) and all higher bases must be powers of the fundamental base. In this sense Mandarin is a pure decimal system. Examples of decimal systems with non-decimal deviations include Welsh and French. In Welsh, 16, 17, 18 and 19 are lexically interpreted as '15+1', '15+2', '15+3' and '15+4'. Therefore 15 is a serialized augend which is not a whole multiple of the system's overall base, namely 10. This evidently involves a quinary principle. However, the term for 15 itself, *bymtheg*, consists of 'five' followed by 'ten', and 18 has an alternate form involving multiplication of 6 and 3. In French, *quatre-vingt* is 'four-twenty' and acts as an additive base for numbers 81–99. This clearly invokes a vigesimal principle, seen also in the forms for numbers 61 to 79.

There are pure decimal and pure duodecimal systems, but no pure quinary systems since 25, 125, etc. have not been documented as higher bases in any language. Although some vigesimal systems, e.g. Mayan, use powers of twenty for higher bases, expressions below 20 – used additively also in higher numerals – employ quinary and/or decimal principles.

The two basic lexical arithmetic operations are multiplication and addition; both are used in all systems which utilize a numeral base. The inverse of addition, subtraction, occurs widely but is highly marked, as shown by the restrictions to which it is subject.

…

The inverse of multiplication, division, is much rarer and subject to even greater restrictions than subtraction.

Seiler (1990: 192):

§4.2 'Bases'. In Greenberg (NS:270) a base is defined as a serialized multiplicand. In English 10, 100, 1000, 1000000 are bases. We should like to retain the term, but apply it to a wider range of phenomena, where the decisive definitory criteria would again be functional.

Bases are marks of hierarchical packing. Packs are classes of numerals, defined both extensionally, viz. by their correspondence to numerical value, and intensionally, viz. by the predominance of certain rule-types.

Heine (1997: 21):

The human hand provides the most important model for structuring the numeral system. Accordingly, the numeral '5' constitutes crosslinguistically the smallest recurrent base number, where "base number" is that number from which counting starts over (cf. Majewicz 1981).

Gvozdanović (1999: 2):

The notion of base thereby requires further clarification: as used in various discussions, it either refers to a building block explicitly involved in any mathematical operation, or a building block used in multiplication only. In view of language variation showing that grouping may be independent of multiplicative iteration, it seems advisable to distinguish between the two. If there is multiplication in a numeral system, it applies to bases (next to digits), thereby revealing what the bases are. If there is no multiplication, then derived grouping (of the type '(5+l)+l' for '7', as mentioned a.o. in Hurford's paper) may be taken to reveal the basic sets. In other words, we may distinguish between basic sets (such as '(5+1)' in the above example) and bases (such as '5', '10' etc., depending on the language): basic sets have a transparent inner structure, whereas bases are conceptualized as entities themselves.

Comrie (2005a: 530; 2013):

By the “**base**” of a numeral system we mean the value n such that numeral expressions are constructed according to the pattern xn + y, i.e. some numeral x multiplied by the base plus some other numeral.

Comrie (2005b: 207–209):

By the “base” of a numeral system is meant that numerical value to which various arithmetical operations are applied. …

The arithmetical operations that are most commonly applied to a base cross-linguistically are addition and multiplication, though subtraction is also found (as in Latin [lat] *un-de-viginti* ‘1-from-20’, i.e. 19; the descendant languages do not have this construction). (Putative instances of division all seem to involve multiplication by a fraction, usually ½, as in traditional Welsh [cym] *hanner cant* ‘half hundred’, i.e. 50.) Many languages also make use of exponentiation, i.e. the raising of the base to a power, but typically using new words that are not transparent to the arithmetical process involved. In English, for instance, hundred is 10^2 and thousand is 10^3, but there is nothing in the structure of these words that indicates this — they are portmanteau forms like Turkish *otuz* 30 discussed above.

In COMRIE (2005), where it was necessary for practical cartographic reasons to have a restricted number of types, I imposed the requirement on the identification of a numeral system base that it serve as a base for both addition and multiplication. For present purposes, however, this is probably too restrictive. One of the reasons for imposing the restriction was to avoid having to include totally idiosyncratic forms, such as traditional Welsh *deu-naw* ‘two-nine’ for 18 (i.e. 2 × 9), given that Welsh does not use 9 anywhere else as the base of an arithmetical operation. Such purely idiosyncratic formations will continue to be excluded. But I will include structures that are recurrent, even if used only in addition (I have no clear crucial examples involving only multiplication). For instance, in Sora [srb], the numerals 13-19 (and optionally 20) are formed by adding the appropriate number to 12, e.g. *miggal-yagi* ‘twelve-three’, i.e. 15 (BHATTACHARYA 1975: 196), but 12 is not used as a base for multiplication (the system above 20 is vigesimal). I prefer to leave open for now precisely how many instances are needed to justify counting a formation as “recurrent” and therefore establishing a base.

Comrie (2022a: 6)

Most numeral systems across the world are characterized by the use of both addition and multiplication, making use of an arithmetic base, which is multiplied and to which (or to a product of which) is added in order to create higher numerals. The general pattern is illustrated in (8): To generate the full range of numerals in the system, the base *b* is multiplied by a number *n*, and a number *m* is then added to the result …

(8) For base b: $(n \times b) + m$

Comrie (2022b: 150:fn3):

My notion “base-n” corresponds to “n-cycle” in Owens & Lean (2018).

von Mengden (2008: 292–293):

A numeral system contains a set of mono-morphemic, arbitrarily shaped numeral forms (*simple numerals*). It also comprises a set of morphosyntactic rules which combine these simple numerals by means of underlying arithmetical operations to form more complex numerals.

There is thus a distinction between simple and complex cardinal numerals. Simple numerals, in turn, can be subdivided into atoms and *bases*. In English, for example, the atoms are the expressions *one*, *two*, *three*, etc., up to nine, whereas the bases are the expressions *ten*, *hundred*, *thousand*, etc. These examples should make the concept of atom and base intuitively quite clear and this, in turn, may help us now to discuss briefly an unambiguous and crosslinguistically valid definition of atoms and bases.

One of the main properties of atoms and bases is their mutual interaction in the numeral system. They play complementary roles in the formation of complex numeral expressions and in the arrangement (ordering) of the elements in the counting sequence. In the overall system, atoms constitute a sequence of continuously recurring elements within the counting sequence or, as Seiler (1990: 190, with my own addition in square brackets) puts it, atoms are "that particular set of numerals that has the highest potential of being recursively used in cycles or with [multiples of] bases." [fn3] In any combination of atoms and bases, atoms are the variables and bases are the constants. Defining bases in this way, the definition of base is still dependent on that of atom. Greenberg (1978: 269–270) simply describes bases as "serialised multiplicands," that is, as factors in multiplications which are applied to a whole series (sequence) of multipliers.

(footnote 3:)
For a valuable summary of the properties of atoms and bases, see Seiler (1990: 190–196). Note that I follow Seiler's terminology in this respect. Greenberg (1978), by contrast, uses atom for what I refer to as simple numeral, that is, as a cover term comprising both bases and our atoms.

von Mengden (2010: 33):

In any formation pattern of a numeral system, <u>bases</u> are those elements with which the smallest continuously recurring sequence of numerals is combined.

Hammarström (2008: 291–292; 2010: 15):

The number *n* is a base iff

1. the next higher base (or the end of the normed expressions) is a multiple of *n*; and
2. a proper majority of the expressions for numbers between *n* and the next higher base are formed by (a single) addition or subtraction of *n* or a multiple of *n* with expressions for numbers smaller than *n*.

Plank (2009: 340–341):

The question about a base universal and its validity, then, turns into one of how to define "bases" in numeral systems.

First, it should be clear that linguistic universals are about linguistic forms and constructions, about lexicons and grammars, and not about mathematics. Thus, while a good case can presumably be made for Sumerian and Babylonian mathematics being based on 60, 10, and 6, the Sumerian numerals were as follows (Thomsen 1987: 82):

1	*diš*, *dili*, *aš*
2	*min*
3	*eš*$_5$

4	*limmu*		
5	*iá*		
6	*àš*	= *iá* + *aš*	5 + 1
7	*imin*	= *iá-min*	5 + 2
8	*ussu*	= *iá-eš*$_5$	5 + 3
9	*ilimmu*	= *iá-limmu*	5 + 4
10	*u*		
20	*niš*		
30	*ušu*$_2$		
40	*nimin, nin*$_5$		
50	*ninnu*		
60	*g̃íš, g̃éš*		
3600	*šár*		

Evidently, the numerals for 6 through 9 were built on that for 5 (a quinary system in this sense) and there is no evidence for the numeral for 6 cyclically recurring in numerals for multiples of 6 or exponentiation with base 6.

Further, in discussions of numeral systems which are intended as dealing with linguistic expressions, two basic concepts of "base" can be distinguished: a CONSTRUCTION-base and a CYCLE-base. Numerals are frequently referred to as "bases" when they are an atomic expression (or at any rate not transparently compositional, synchronically speaking) and when expressions for other numerals are formally based on them, with higher or lower numerals constructed by arithmetic operations (addition, subtraction, multiplication) with their help. The requirement of such a construction-base being itself atomic is sometimes waived – in which case a numeral for 6, analysable into 5+1, can be called a base (as it is in Hurford 1999) when subsequent numerals include it, with 7 = [5+1]+1 etc. as in Miskito. If only a single other numeral is constructed from another numeral through a given arithmetic operation, like 18 = 3 × 6 in Breton or 12 = 2 × 6 in Mankanya (Atlantic, Niger-Congo; Zaslavsky 1999), this would almost seem too little to qualify as a construction-base. A cycle-base is the narrower concept: a numeral is a cycle-base if it is a construction-base and cyclically recurs in linguistic designations of multiples of the respective number (base-6: 6, 12, 18, 24, . . .) and/or in exponentiation with that base (base-6: 6, 36, 216, . . .). Comrie (2005: 530) is among those subscribing to a narrow definition of this kind: "By the 'base' of a numeral system we mean the value *n* such that numeral expressions are constructed according to the pattern . . . *xn* + *y*, i.e. some numeral *x* multiplied by the base plus some other numeral" – which explains why the WALS sample lacks relevant languages.

Hanke (2010: 68–69):

The most prominent additive pattern is the additive series, which occurs in the overwhelming majority of languages across the world. An example of an additive series is *twenty-one* to *twenty-nine*. An additive series consists of an *additive base* (or *augend* in Greenberg's terms) to which a sequence of numbers is added. In the example above, 20 is the additive base, expressed by twenty. In nearly all cases, the sequence of numerals starts with '1'. In principle, the minimal range of a series is just two adjacent numerals, but in practice the smallest series consist of 5, 10 or multiples of these numbers.
The apparent reason for the ubiquity of those values is the readily available model of the five digits of human hands.

Additive series are nearly always embedded in patterns with wider range. The second major type of numeral pattern – nearly as common as additive series – is the *multiplicative series*, in which a *multiplicative base* (or *multiplicand*) is multiplied by a sequence of numbers, as in English *twenty* to *ninety*. The two types of series are very often combined, as in the range from ten to ninety-nine. For this range, 10 is the *additive-multiplicative base* (often simply called *base*, cf. Comrie 2005), because 10 and its multiples are the augends of the succeeding additive series.

Calude & Verkerk (2016: 93):

The creation of a syntagm involves three components: (already existing) *atoms*, a *base* (which is typically an existing atom used serially to derive larger numerals), and a calculatory or arithmetic (Seiler 1990; Greenberg 1978, respectively) operation such as addition, subtraction, multiplication, division, or exponentiation.

Everett (2017: 14):

A base is a building block for other numbers.
(endnote 3 on p.266:)

For more formal definitions of bases, see, for example, Bernard Comrie, “The Search for the Perfect Numeral System, with Particular Reference to Southeast Asia,” *Linguistik Indonesia* 22 (2004): 137–145, or Harald Hammarström, “Rarities in Numeral Systems,” in *Rethinking Universals: How Rarities Affect Linguistic Theory*, ed. Jan Wohlgemuth and Michael Cysouw (Berlin: De Gruyter Mouton, 2010), 11–59, 15, or Frans Plank, “Senary Summary So Far,” *Linguistic Typology* 3 (2009): 337–345. Such formal definitions are avoided here as they differ from one another in minor ways that are not central to our story.

Everett (2021: 14):

Bases of verbal numbers are the key numbers around which larger numbers are structured, usually in a multiplicative fashion.

Owens & Lean (2018: 34):

In some of the older linguistic literature concerned with the description of natural language numeral systems, it was common to use the descriptive term “base” when discussing the cyclic nature of the system. Thus we find counting systems variously termed “binary” (base 2), “ternary” (base 3), “quinary” (base 5), “decimal” (base 10), and “vigesimal” (base 20). Using a single number to characterise a counting system is reasonably adequate when we are dealing with, say, the English counting system which, with some irregularities, is essentially a base 10 one. The cyclic structure of many of the counting systems found in Melanesia is often more complex than the English system, in that a single system may have elements of base 2, base 5, and base 20; others have a structure in which we can discern elements of base 5, base 10, and base 20. This was recognised in the older literature in which we find reference to “mixed base” systems and such terms as “incomplete decimal” systems (that is, one which had elements of both base 5 and base 10).

Schapper (2020: 490):

The chief variable here is the base that a language’s numeral system uses, that is, the value upon which higher numeral expressions are constructed. A language may have more than one base, for instance, using 5 as the base to form numerals 6–9, and 10 for higher numerals.

Holz (2021: 232–233):

A base is a numeric value to which arithmetic operations are applied to create number words of a higher value. Very common are quinary, decimal, and vigesimal systems, which feature a base-5, base-10, or base-20, respectively.

References:

Barlow, Russell. 2023a. Papuan-Austronesian contact and the spread of numeral systems in Melanesia. *Diachronica* 40(3). 287–340. https://doi.org/10.1075/dia.22005.bar

Barlow, Russell. 2023b. *A grammar of Ulwa (Papua New Guinea)* (Comprehensive Grammar Library 6). Berlin: Language Science Press. https://zenodo.org/record/8094859

Bhattacharya, Sudhibhushan. 1975. *Studies in comparative Munda linguistics*. Simla: Indian Institute of Advanced Study.

Bender, Andrea & Beller, Sieghard. 2018. Numeration systems as cultural tools for numerical cognition. In Daniel B. Berch & David C. Geary & Kathleen Mann Koepke. *Language and culture in mathematical cognition* (Mathematical Cognition and Learning 4), 297–320. Cambridge: Academic Press. https://doi.org/10.1016/B978-0-12-812574-8.00013-4

Bowern, Claire & Zentz, Jason. 2012. Diversity in the numeral systems of Australian languages. *Anthropological Linguistics* 54(2). 133–160. https://doi.org/10.1353/anl.2012.0008

Calude, Andreea S. & Verkerk, Annemarie. 2016. The typology and diachrony of higher numerals in Indo-European: A phylogenetic comparative study. *Journal of Language Evolution* 1(2): 91–108. https://doi.org/10.1093/jole/lzw003

Chan, Eugene, Hans-Jörg Bibiko, Christoph Rzymski, Simon J Greenhill, and Robert Forkel. 2019. *channumerals* (v1.0). doi: 10.5281/zenodo.3475912. Derived from Eugene Chan's "Numeral systems of the world's languages" (accessed 30 September 2019). Currently available as: https://lingweb.eva.mpg.de/channumerals.

Comrie, Bernard. 1999. Haruai numerals and their implications for the history and typology of numeral systems. In Jadranka Gvozdanović (ed.), *Numeral types and changes worldwide* (Trends in Linguistics, Studies and Monographs 118), 81–94. Berlin: Mouton de Gruyter.

Comrie, Bernard. 2005a. Numeral bases. In Martin Haspelmath & Matthew Dryer & David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures* (WALS), 530–533. Oxford: Oxford University Press.

Comrie, Bernard. 2005b. Endangered numeral systems. In Jan Wohlgemuth & Tyko Dirksmeyer (eds.), *Bedrohte Vielfalt: Aspekte des Sprach(en)tods / Aspects of language death*, 203–230. Berlin: Weißensee Verlag.

Comrie, Bernard. 2013. Numeral bases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online* (WALS). Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/131. Data available at: https://zenodo.org/records/7385533

Comrie, Bernard. 2022a. The arithmetic of natural language: Toward a typology of numeral systems. *Macrolinguistics* 10(1): 1–35. https://doi.org/10.26478/ja2022.10.16.1

Comrie, Bernard. 2022b. Beyond endangered: Some reflections on the future of indigenous numeral systems. *Language and Linguistics in Melanesia* 40. 149–159. https://www.langlxmelanesia.com/llm-vol--40-2022

Conant, Levi Leonard. 1896. *The number concept: Its origin and development*. New York:

Macmillan.

Dixon, Roland B. & Kroeber, A. L. 1907. Numeral systems of the languages of California. *American Anthropologist* 9(4). 663–690. https://www.jstor.org/stable/659483

Epps, Patience. 2013. Inheritance, calquing, or independent innovation? Reconstructing morphological complexity in Amazonian numerals. *Journal of Language Contact* 6(2). 329–357. https://doi.org/10.1515/lity-2012-0002

Everett, Caleb. 2017. *Numbers and the making of us: Counting and the course of human cultures*. Cambridge: Harvard University Press. https://doi.org/10.4159/9780674979185

Everett, Caleb. 2021. The diversity of linguistic references to quantities across the world's cultures. In Annemarie Fritz & Erkan Gürsoy & Moritz Herzog (eds.), *Diversity dimensions in mathematics and language learning: Perspectives on culture, education and multilingualism*, 3–20. Berlin: De Gruyter. https://doi.org/10.1515/9783110661941-001

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

Greenberg, Joseph H. 2000. Numerals. In Geert Booij & Christian Lehmann & Joachim Mugdan & Wolfgang Kesselheim & Stavros Skopeteas (eds.), *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung, 1. Halbband / Morphology: An international handbook on inflection and word-formation, volume 1*, 770–783. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110111286.1.10.770

Gvozdanović, Jadranka. 1999. Types of numeral changes. In Jadranka Gvozdanović (ed.), *Numeral types and changes worldwide*, 95–112. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110811193.95

Hammarström, Harald. 2008. Complexity in numeral systems with an investigation into pidgins and creoles. In Matti Miestamo & Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change* (Studies in Language Companion Series 94), 287–304. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.94.18ham

Hammarström, Harald. 2010. Rarities in numeral systems. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory* (Empirical Approaches to Language Typology 45), 11–60. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110220933.11

Hanke, Thomas. 2010. Additional rarities in the typology of numerals. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: How rarities affect linguistic theory* (Empirical Approaches to Language Typology 45), 61–90. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110220933.61

Harris, John. 1982. Facts and fallacies of Aboriginal number systems. In Susanne Hargrave (ed.), *Language and culture: Work Papers of SIL-AAB*, series B, volume 8, 153–181. Darwin: Summer Institute of Linguistics, Australian Aborigines Branch.

Holz, Christoph. 2021. Sequential number word formation in the Tungag-Nalik Languages (New Ireland). *Oceanic Linguistic* 60(1). 231–242. https://doi.org/10.1353/ol.2021.0007

Holzknecht, Susanne. 1989. *The Markham languages of Papua New Guinea* (PL-C115). Canberra: Pacific Linguistics. https://doi.org/10.15144/PL-C115

Hurford, James R. 1999. Artificially growing a numeral system. In Jadranka Gvozdanović (ed.), *Numeral types and changes worldwide*, 7–41. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110811193.7

Kroeber, A. L. & Grace, George William. 1960. *The Sparkman grammar of Luiseño* (University of California Publications in Linguistics, 16). Berkeley: University of California Press.

Laycock, Donald C. 1975. Observations on number systems and semantics. In Stephen A. Wurm (ed.), *New Guinea area languages and language study, volume 1: Papuan languages and the New Guinea linguistic scene* (PL-C38), 219–233. Canberra: Pacific Linguistics. http://hdl.handle.net/1885/145150

Lincoln, Peter C. 2010. Count not on substrata. In John Bowden & Nikolaus P. Himmelmann & Malcolm Ross (eds.), *A journey through Austronesian and Papuan linguistic and cultural space: Papers in honour of Andrew K. Pawley*, 225–244. Canberra: Research School of Pacific and Asian Studies, Australian National University. http://hdl.handle.net/1885/146763

Lynch, John. 1982. South-West Tanna grammar outline and vocabulary. In John Lynch (ed.), *Papers in linguistics of Melanesia no. 4* (PL-A64), 1–92. Canberra: Pacific Linguistics. http://hdl.handle.net/1885/145145

Lynch, John. 2009. At sixes and sevens: The development of numeral systems in Vanuatu and New Caledonia. In Bethwyn Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross* (PL-605), 391–411. Canberra: Pacific Linguistics. http://hdl.handle.net/1885/146753

Majewicz, Alfred F. 1981. Le rôle du doigt et de la main et leurs désignations dans la formation des systèmes particuliers de numération et de noms de nombres dans certaines langues. In Fanny de Sivers (ed.), *La main et les doigts dans l'expression linguistique* (LACITO - Documents Eurasie 6 2), 193–212. Paris: SELAF.

McElvenny, James. 2006. *Notes on the Mudukumo language of the Yuat River, East Sepik Province, PNG*. Sydney: University of Sydney. (Unpublished MS).

Owens, Kay & Glendon Lean. 2018. *History of number: Evidence from Papua New Guinea and Oceania* (History of Mathematics Education). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-45483-2

Plank, Frans. 2009. Senary summary so far. Linguistic *Typology* 13(2): 337–345. https://doi.org/10.1515/LITY.2009.016

Ross, Malcolm. 2023. Digit tallying. In Malcolm Ross, Andrew Pawley & Meredith Osmond (eds.), *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society, volume 6: People: Society*. Canberra: Asia-Pacific Linguistics. http://hdl.handle.net/1885/106908

Salzmann, Zdeněk. 1950. A method for analyzing numerical systems. *WORD* 6(1). 78–83. https://doi.org/10.1080/00437956.1950.11659369

Schapper, Antoinette. 2020. Linguistic Melanesia. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge handbook of language contact*, 480–502. London: Routledge. https://doi.org/10.4000/archipel.371

Schapper, Antoinette & Hammarström, Harald. 2013. Innovative numerals in Malayo-Polynesian languages outside of Oceania. *Oceanic Linguistics* 52(2). 423–456. https:// http://doi.org/10.4324/9781351109154-29

Seiler, Hansjakob. 1990. A dimensional view on numeral systems. In William A. Croft & Suzanne Kemmer & Keith Denning (eds.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday*, 187–208. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.20.12sei

Skirgård, Hedvig et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9: eadg6175. https://doi.org/10.1075/10.1126/sciadv.adg6175. Data available at: https://doi.org/10.5281/zenodo.7740140

Stampe, David. 1976. Cardinal number systems. In Salikoko S. Mufwene & Carol A. Walker

& Sanford B. Steever (eds.), *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, 594–609. Chicago: Chicago Linguistic Society.

Thomsen, Marie-Louise. 1987. *The Sumerian language: An introduction to its history and grammatical structure.* 2nd edition. København: Akademisk Forlag.

von Mengden, Ferdinand. 2008. The grammaticalization cline of cardinal numerals and numeral systems. In María José López-Couso & Elena Seoane (eds.), *Rethinking grammaticalization: New perspectives* (Typological Studies in Language 76), 289–308 Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.76.14men

von Mengden, Ferdinand. 2010. *Cardinal numerals: Old English from a cross-linguistic perspective* (Topic in English Literature 67). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110220353

Zaslavsky, Claudia. 1999. *Africa counts: Number and pattern in African cultures*. 3rd edition. Westport: Lawrence Hill & Co.