

Journal of Universal Language 21-2. September 2020, 43-84
DOI 10.22425/jul.2020.21.2.43
eISSN 2508-5344

Revisiting Greenberg's "Generalizations about Numeral Systems" (1978)*

Bernard Comrie

University of California, Santa Barbara, USA

Abstract

This article examines seven of the 54 generalizations on numeral systems presented by Greenberg in 1978 either as absolute universals or as universal tendencies. Five of the generalizations deal with details of the arithmetic structure of numeral systems, with particular reference to the order of elements in the numeral expression and to the use of subtraction and related phenomena in numeral expressions. It is shown that the details of these generalizations need to be revised

§ Portions of the material included in this article have been presented at various fora. I am grateful to all those who contributed to the ensuing discussions. This article is dedicated to the memory of the late Joseph H. Greenberg, the giant on whose shoulders I stand.

Bernard Comrie
Professor, Department of Linguistics, University of California, Santa Barbara
Email: comrie@linguistics.ucsb.edu

Received 11 August, 2020; Revised 11 September, 2020; Accepted 15 September, 2020

Copyright © 2020 Language Research Institute, Sejong University
Journal of Universal Language is an Open Access Journal. All articles are distributed online under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

in light of new data, and argued that even the formulation of universal tendencies is problematic given the rarity of some of the phenomena at issue and the resulting difficulty of establishing statistically valid inferences. Two of the generalizations discuss fundamental properties of natural language numeral systems, namely whether they are necessarily finite, and whether they allow gaps. It is argued, contra Greenberg, that there are both natural number series that are infinite and those with gaps. Throughout, it is emphasized that only the careful formulation of the original generalizations by Greenberg has enabled us to progress in our understanding of these aspects of numeral systems in natural languages.

Keywords: numeral system, language universal, linguistic typology, recursion

1. Introduction

Greenberg (1978)¹ is a classic in the small but growing literature on the crosslinguistic study of numeral systems.² The article is a remarkable synthesis of universals and near-universals (54 “generalizations”) from what was known at the time about numeral systems in the world’s languages. It is all the more remarkable in that it predates the ready available of easily accessible crosslinguistic electronic databases such as Chan (2020), being based rather on Greenberg’s assiduously working

¹ I will refer to this publication by its original publication date, 1978, so that it is correctly placed in its historical context. The 1990 reprint corrects some (but not all) typographical errors, including most importantly the numbering of the generalizations: The original edition omits 8 in the numbering and duplicates 21. I have used the corrected numbering from the reprint, with the original numbering in parentheses in the format “(1978: n)” where different.

² I will use the term “numeral” when speaking of linguistic expressions, and “number” when speaking of arithmetic values, e.g., in English the numeral *five* expresses the number 5. I use the same form *numeral* as an adjective, but note that in citations from Greenberg I retain his occasional use of the adjective *numerical*.

through conventionally published print material. As Greenberg would himself have predicted, four decades later we now have much more extensive knowledge of the world's numeral systems, yet remarkably, Greenberg's generalizations have by and large stood the test of time. It should, incidentally, be noted that for some of the generalizations Greenberg was already aware of occasional exceptions, which are explicitly mentioned in the article; this is sometimes overlooked when generalizations that he in fact stated as universal tendencies are quoted as if claimed to be exceptionless.

The present article is divided into two main parts. In §2, I examine the details of five of Greenberg's generalizations relating to the arithmetic structure of numeral expressions, in some cases introducing corrections on the basis of newly available data that were not taken into account by Greenberg, in other cases offering reinterpretations of Greenberg's analyses on the basis of more extensive work on particular languages. This section offers an expanded and more systematic treatment of some issues first discussed in Comrie (2005). In §3, I examine two of the most fundamental claims from the 1978 article and argue that they are in fact wrong, but nonetheless of great interest in allowing us to probe the boundaries of numeral systems crosslinguistically.

2. The Arithmetic Structure of Numeral Systems

In languages with extensive numeral systems that go up into the hundreds and beyond, there is usually an underlying arithmetic scaffold that makes use of a base to which are applied the arithmetic operations of addition, multiplication, and exponentiation. Thus, in an English numeral expression like (1), we can identify a base 10. The

square of the base is named *hundred* (exponentiation), which in (1) is multiplied by 5. To this is added the result of taking the base and multiplying it by 6, and also the unit 7, giving the numeral expression in the first line of (1), the corresponding number in Arabic digits in the second line, and the arithmetic analysis given in the third line.

- (1) five hundred (and) sixty-seven
 567
 $(5 \times 10^2) + (6 \times 10) + 7$

While individual languages may introduce numerous complications in detail, e.g., in English 60 is expressed as *six-ty*, with the 10 being expressed by a suffix rather than by the free lexical item *ten*, the underlying scaffold is usually nonetheless clearly discernable, with addition, multiplication, and exponentiation as the arithmetic processes applied to the base. For a detailed investigation of a decimal system, see the analysis of the Chadic language Po Tangle in Amaechi (2014). Sanusi & Yusuf (2018) provide an analysis of a more complex system in the Gur language Bātōnū, which uses multiple bases and, in counting money, subtraction (see §2.2).

In this section, we will be concerned with two issues relating to further details of expressions such as (1). §2.1 considers the order of the components that are added together (hereafter the “addends”) to give the composite numeral expression as in (1). Modern English consistently orders from higher numeral expression to lower numeral expression, i.e., in (1) the order is hundreds, then tens, then units. §2.2 considers another family of arithmetic operations that can be subsumed under the general heading of subtraction and related phenomena.³

³ There are of course numerous other variations on the basic pattern that will not be

2.1. Order of Components

In Modern English, as illustrated in (1), the order of the addends is consistently decreasing, and this is fact the most widely attested order crosslinguistically; it is illustrated in (2) from Mandarin Chinese to give an example from a completely different language family.

- (2) Mandarin
 sān-bǎi wǔ-shí èr
 three-hundred five-ten two
 352
 $(3 \times 10^2) + (5 \times 10) + 2$

Although much rarer, there are some languages that consistently use the inverse, increasing ordering, as in (3) from Standard Malagasy (Rajaonarimanana 2001: 67, slightly modified).

- (3) Malagasy (Standard)
 roa amby dimam-polo sy telon-jato
 two plus five-ten and three-hundred
 352
 $2 + (5 \times 10) + (3 \times 10^2)$

This order is also one of the possibilities available to Arabic, although it is not the usual ordering found in the modern language (for which see below); Ali (2013: 71-72) gives examples noting: “This style is frequently used in classical texts or by those imitating the classical style”.

considered here, e.g., French has a basically decimal system but nonetheless has some expressions that use base 20 (vigesimal), e.g., *quatre-vingts* 80, literally ‘four-twenties’.

However, some languages make use of a mixture of decreasing and increasing order, with German being perhaps the most famous example of a language that uses basically decreasing order but places the units before the tens, as in (4).

- (4) German
drei-hundert zwei-und-fünf-zig
three-hundred two-and-five-ty
352
 $(3 \times 10^2) + 2 + (5 \times 10)$

This is also the usual order in Modern Arabic (Ali 2013: 66-71).

The question arises: Are there constraints on the possible combinations of descending and ascending orders? Greenberg proposed two such constraints:

Generalization 26: If in a language, in any sum the smaller addend precedes the larger, then the same order holds for all smaller numbers expressed by addition.

Generalization 27: If in a language, in any sum the larger addend precedes the smaller, then the same order holds for all larger numbers expressed by addition.

These two generalizations are in a sense mirror images of each other, and can be combined into a single generalization that if a language combines descending and ascending orders, then it will use ascending order for lower numbers and descending order for higher numbers, although the cut-off point between the two patterns may vary cross-linguistically, and there may be an intermediate area where

both are possible. More specifically, the generalizations allow languages like German with the order hundreds—units—tens, but would disallow the inverse order tens—units—hundreds.

As it happens, languages of Europe and nearby areas present a rich variety of variations on the theme of consistency with this universal, i.e., languages with basically descending order but ascending order for some set of lower numbers (with different cut-off points). German, as illustrated in (4) and again in (5), has the cut-off point at 100.

- (5) German
- a. neun-und-neun-zig
nine-and-nine-ty
99
 - b. hundert-eins
hundred-one
101

Although English-speakers are typically unaware consciously of the fact, English also combines the two orders, although the cut-off point here is at 20, as illustrated in (6).

- (6) a. nine-teen
19
- b. twenty-one
21

Romance languages typically have a cut-off point within the teens, either between 16 and 17, as in Italian example (7), or between 15 and 16, as in Spanish example (8).

- (7) Italian
 - a. se-dici
six-ten
16
 - b. dici-as-sette
ten-and-seven
17

- (8) Spanish
 - a. quin-ce
five-teen
15
 - b. diec-i-séis
ten-and-six
16

The Insular Celtic languages provide, in their traditional vigesimal (base 20, "score") systems, evidence for various cut-off points around 60, although different sources are not always entirely in agreement on the precise details. For Manx, Kewley Draskau (2008: 106-107) gives only the order Unit—Score through 59, but describes the order Score—Unit as occurring "usually" from 61 up, as in (9).

- (9) Manx
- a. nuy jeig as da-eed⁴
 nine teen and two-twenty
 59
 $9+10+(2\times 20)$
- b. t(h)ree feed as 'nane
 three twenty and one
 61
 $(3\times 20)+1$

For Scottish Gaelic, Calder (1923: 127), describing the standard language, gives only the order Unit—Score for 21-23, 31-33, 41-42, both orders for 50 (alongside the alternative ‘half-hundred’ construction), and only the order Score—Unit for 70 and the higher twenties 130, 150, 170 and 190. Oftedal (1956: 228-229), describing a spoken variety of Scottish Gaelic, gives the order Unit—Score for 21, 22, 30 and 32, but the order Score—Unit for 41, 51 (alongside the ‘half-hundred and one’ construction, with ‘half-hundred’ as the only form cited for 50), and 73 (in 4573), suggesting a cut-off point in this dialect at 40. For Welsh, Bowen & Rhys Jones (1960: 109) give only the order Unit—Score below 60 but both orders above 60 with the recommendation “use the bigger number first”, i.e., Score—Unit; Hurford’s extensive study of numerals in the Welsh translation of the Bible from 1620 finds only Unit—Score below 60, both orders above 60 (Hurford 1975: 154-163). So where a variety has both descending and ascending order, it is always the case that ascending order is used for lower numbers, descending order for higher numbers, though there

⁴ Initial consonants in Insular Celtic languages undergo morphophonological alternations traditionally called “mutations”, whereby in certain environments the initial consonant of the citation form may change or disappear; the citation form *feed* ‘twenty’ appears as *eed* in (9a).

may also be an "intermediate" zone (which may extend to the upper limits of the system) where both orders are possible.

In his discussion of Generalizations 26-27, Greenberg noted that he was aware of "just one" exception, namely Trumai, where 3 is expressed as (2+1) but 6 is expressed as (1+5),⁵ and likewise for 7-9 and 11-14 (e.g., 11 as (1+10)). A sample of relevant forms from Greenberg's source, Steinen (1894: 541-542), with the analysis implied by Greenberg's remarks, is given in (10); square brackets identify instances where Greenberg's implied analysis is questionable.

- (10) Trumai (data from Steinen, analysis from Greenberg)
- a. mihin
one
1
 - b. hurš
two
2
 - c. hurs-tame
two-[one]
3
 - d. nekatkelan
five
5
 - e. mihin-po(k)péskun
one-[five]
6

⁵ Greenberg (fn. 5) actually has "(5+1)" but this is clearly a typographical error (uncorrected in the revised edition) for (1+5), judging both from Greenberg's source (see (10)) and from the fact that if this were the correct analysis Trumai would simply have consistent descending order.

- f. yepun-pokpéškun
 ?-[five]
 10
- g. mihin-pitsa
 1-[10]
 11

These forms can be compared by those given by Guirardello (1999: 46-48), a modern linguistic study of the Trumai language.⁶ Note that some alternative forms have been omitted from (10) and (11), and that the present-day speakers did not form numerals above 10 in Trumai.

(11) Trumai (data and analysis from Guirardello)

- a. mihin
 one
 1
- b. huch
 two
 2
- c. huch-tahme
 two-?
 3
- d. ine k'ad kel-an
 3M hand finger-LOC
 5
- e. k'ad kel wakpechkun
 hand finger cross
 6

⁶ The Trumai forms as given by Steinen and those documented in the current language are discussed in further detail in Comrie & Guirardello-Damian (2004).

f.	k'ad	kel	wanle	kan
	hand	finger	finish	?
	10			

The numeral 3 shows agreement across the two sources, but while the first component is recognized as identical to 2 by present-day speakers, the second is not recognized as 1, or indeed as anything—it would thus be a cranberry morph. Since Trumai has no demonstrable linguistic relatives, there is no possibility of attempting an etymology based on comparative material. One can speculate on possible etymologies (?big-two, ?beside-two), but this is pure speculation, and at most one can say that there is no evidence for Greenberg's analysis (2+1). The modern form for 6 does not contain any explicit reference to 5, but is literally something like '[one] hand finger crosses', i.e., one crosses over to one finger of the second hand in finger-counting; reference to 5 is at best derived from the fact that the first hand has five fingers. The form given by Steinen is apparently a contracted variant of this. The second component of the form for 11 given by Steinen is related to the word for 'foot', *pits* in the modern language recorded by Guirardello, so the overall form is probably a contraction of something like 'one finger crossing to the foot', so again the concept 10 is at best derived, from the fact that one has ten fingers, and beyond ten has to start counting the toes. Steinen's form for 11 bears no resemblance to that given for 10, whose first component is the item recorded by Guirardello (1999: 48) as *yupun* 'all', i.e., one has crossed over to all five fingers of the second hand. In sum, further details of Trumai, which were not available to Greenberg in 1978, suggest that there is no motivation for analyzing Trumai in a way that makes it an exception to Generalizations 26-27.

However, we do now know of one clear counterexample to these

generalizations, from Sakalava Malagasy, and more specifically its Nosy Be dialect.⁷ It will be recalled (see (3)) that Standard Malagasy has consistently ascending order. Sakalava Malagasy differs from this in placing the tens before the units, as noted explicitly by Dahl (1968: 14), who gives the example (12).

(12) Sakalava Malagasy

limam-polo	roe	amby,	amby	telon-jato
five-ten	two	plus	plus	three-hundred
352				

This is a clear counterexample, and therefore raises the question whether there is any validity to the pair of Generalizations 26-27. One might counter that this is only one exception, so that the generalizations would still stand as near-universals. But it must be borne in mind that this exception can only arise in a language that has basically ascending order, and that basically ascending order is much rarer than basically descending order in the world's languages. There are therefore far more opportunities for languages to combine ascending order at the bottom with basically descending order than for languages to combine descending order at the bottom with basically ascending order. And simple statistical testing will not work, since the different languages that combine basically descending order with ascending order at the bottom are not all independent of one another. Unless and until a valid statistical test can be carried out, there is a real possibility that Generalizations 26-27 are simply spurious.

Note, however, that a more general conclusion drawn by Greenberg (1978) in his discussion of Generalizations 26-27 does still seem to

⁷ I am grateful to Joachim Sabel for confirming the occurrence of this constituent order and for identifying the dialect of Nosy Be as its locus.

hold up, namely that there is a strong crosslinguistic preference for descending order, perhaps motivated by the fact that descending order facilitates identification of the ballpark value of the numerical expression, especially for higher numbers, by giving its numerically largest component first.

Another possible counterexample to Generalizations 26-27 is presented by the sixteenth-century Toba data in (13) (Bárcena n.d.: 226),⁸ where the expression for 9 has the larger addend before the smaller, but the expression for 10 has the smaller addend before the larger.

(13) Toba

a. nivoca nalotapegat natedac
two four one
9

$(2 \times 4) + 1$

b. cacayni nivoca nalotapegat
two two four
10
 $2 + (2 \times 4)$

However, generalizations drawn from data as in (13) need to be treated with some caution. The problem is that they derive from unique sources for limited attestation of numerals that are no longer current in the modern language (for Toba: Harriet Manelis Klein, p.c.). While there is no reason to doubt the genuineness of the data in (13), there is no reason to assume that the versions given in the source

⁸ The date of Bárcena's original manuscript is apparently unknown. Different sources give slightly different dates of birth/death for Bárcena (whose family name also shows spelling variants), suggesting that he was born in 1528 or 1530 and died in 1597 or 1598.

were the only versions possible in usage at the time, rather than for instance both orders Larger—Smaller and Smaller—Larger being possible for both numerals, in the absence of an explicit statement by the author that alternatives are excluded. In other words, even if we have complete confidence in what the author says, we cannot be sure how to interpret what they do not say.

A second caveat is that in such examples we are often dealing with isolated linguistic formations rather than general patterns that are either fully productive or at least define a substantial series of numeral expressions, such as the teens in English or the combinations of tens and units in English or German. If a given formation is used for only a very limited number of numerals—Trumai (10c) on Greenberg’s analysis, for instance, would be a single instance of this construction—then it may well be that speakers simply learn the whole for as a unit, especially if the formation is even partially opaque. Sakalava Malagasy is therefore particularly valuable in that it clearly establishes a productive pattern that goes against Generalizations 26-27.

2.2. Subtraction and Related Operations

While addition, multiplication, and exponentiation form the cornerstones of nearly all numeral systems that extend to relatively high numbers, there are some other arithmetic processes that are used occasionally, including subtraction, as illustrated in Latin example (14), where both 18 and 19 are expressed by subtraction from 20, e.g., $20-2=18$.

- (14) Latin
- a. duo-de-viginti
two-from-twenty
18
 - b. un-de-viginti
one-from-twenty
19

It will be useful to introduce some terminology to refer to the different elements in the equation $y-z=x$; y will be referred to as the minuend and z as the subtrahend; the result is usually referred to as the difference, but for our purposes it will be more helpful to refer to it as the target, since it is the target number that the linguistic expression denotes. In example (14a), the target is 18, the minuend is 20, and the subtrahend is 2.

Greenberg (1978: 15) proposes the following constraint on minuends:

Generalization 14: Every minuend is a base of the system or a multiple of the base.

Since Latin has a decimal system, with base 10, and the minuend in (14) is 20, a multiple of the base ($20=2\times 10$), Latin is consistent with this generalization, as are most instances of subtraction in numeral systems of the world's languages.

Greenberg discussed three counterexamples to Generalization 14. The first two deal with restricted sets of expressions rather than a productive process, but are still worth examining, as they do throw light on the precise nature of the arithmetic processes involved. The first is from Arikara, as presented by Wied (1841: 472). Relevant forms are given in (15), along with the forms for the current language from Parks (2009).

(15) Arikara

	<u>Wied</u>	<u>Parks</u>
1	áchku	áxku
2	pítcho	pítku
7	taui-scháhpis-uahn	tawišaapiswaána
8	tauisháhpis	tawišaápis
9	nochenéh-uahn	nooxiniwaana
10	nochén	nooxíni?
11	pitkóche-nëuahn	nooxíni? na áxku
12	pitchóchin	nooxíni? na pítku

Once one makes allowance for Wied's somewhat impressionistic transcription, there is close agreement between the two sets of data through 10; for the teens, however, the systems diverge. The forms of interest are those for 7 and 9 in both columns and for 11 as given by Wied. In each case, the expression contains a first component that is identical or almost identical to the expression for the next higher number, followed by an element that we can refer to as *-waana* on the basis of the current form. The formation is thus reasonably transparent: *-waana* is an instruction to subtract 1 from the number that is expressed by the first component, to arrive at the target value, i.e., $8-1=7$, $10-1=9$, and in Wied's documentation $12-1=11$. Now, Arikara overall has a vigesimal numeral system (Parks 2009), but since the teens are formed by adding to the expression for 10, 10 serves as a lower base within the system for the range 11-19. Even though Wied does not list other teens, the expression for 12, *pitchóchin*, is apparently composed of *pítcho* 'two' and a contraction of *nochén* 'ten', so 10 presumably served as a lower base in the system documented by Wied too. If we interpret the expressions for 7, 9 and 11 as subtractive, then that for 9 is unproblematic for Generalization

14, since the minuend is 10, but the expressions for 7 and 11 are problematic, since neither 8 nor 12 is a base of the system or a multiple of a base.

It should, however, be noted that the formative *-waana* is not the word for 'one', and one might therefore argue that these formations are not strictly speaking subtractive, but might be given some alternative interpretation, e.g., along the lines of 'the item before (sc. in counting)'. If Generalization 14 is interpreted as strictly referring to subtraction, then the Arikara forms in *-waana* do not fall under the generalization and therefore neither instantiate nor counterexemplify it. However, there is clearly an intuition that goes beyond subtraction in the literal sense and seems implied by Greenberg's use of the term "subtraction" for expressions that do not literally include any reference to subtraction of a particular subtrahend. What unites subtraction and examples like those from Arikara is that a numeral is identified by first identifying a higher numeral and then counting down from that higher numeral, in the case of Arikara counting down by just 1. We may then refer to the set of constructions that count down as "counting down" constructions. If we then generalize the term "minuend" to denote the number from which one counts down, whether this is expressed literally as subtraction or not, then we match Greenberg's intuition that there is something unusual about the Arikara expressions for 7 and (in Wied's data) 11, namely that they count down from a number value that is not a base of the system.

One might speculate on whether there are still constraints on possible minuends in this extended sense, and I will return below to some relevant empirical facts when I consider the Colonial Valley Zapotec numeral system. One possibility worth pursuing is that the minuend must be a "salient" number value in the culture, but validating this would of course involve checking whether 8 and 12 are

or were indeed salient number values in Arikara, as otherwise there is a danger of the claim becoming circular, declaring by fiat that any minuend is culturally salient. But there are clear instances of numbers that are salient in a given culture though not a base in the numeral system, e.g., 12 in English, which is not a base or a multiple of a base in the language's decimal numeral system, but nonetheless receives a special secondary lexical designation as *dozen*.

Before leaving Arikara, it should also be noted that apparently only a very limited set of numerals are formed with the formative *-waana*, in contrast to Latin where the subtractive formation is also used for 28, 29, etc., at least through 88 and 89. So the issue of lack of productivity, noted at the end of §2.1, does arise here, even though in this case the formations seem transparent.

Greenberg mentions another language as providing exceptions similar to Arikara, namely “Montagnais, an Athabaskan language”, i.e., not the Algonquian language Montagnais or Innu, but the Athabaskan language Chipewyan or Dine Suline, earlier called Montagnais by French missionaries. Greenberg's source is apparently Conant (1931: 53-54), which is listed in his References Cited and is itself based on Petitot (1876: lv). The examples in (16) show the forms given by Petitot (1876), in comparison with modern forms from Jaker (2014).

(16) Chipewyan

	<u>Petitot</u>	<u>Jaker</u>
a. 1	inl'apè	ʔɪ́ághe
b. 3	t'apè	taghe
c. 4	dinpi	dɪghɪ
d. 6	elkkè-t'apè	ʔelk'étághe
e. 7	t'a-yé-oyertan	totą
f. 7	inl'as-dinpi	láísɔdɪghɪ
g. 8	elkkè-dinpi	ʔelk'édɪghɪ
h. 9	inl'a-yé-oyert'an	ʔelɔ́ótą
i. 10	onernan	honéna

Most of the forms provided by Jaker are readily relatable to those from Petitot, though (16e) and (16h), if the forms in the two columns are indeed related, have undergone reduction in the current language that obscures their etymology.

Greenberg says that 9 is expressed as (10–1), that 8 is expressed as (4×2), and that 7 is expressed either as (10–3) or as (8–1), with the last being problematic for Generalization 14, since the system is overall decimal and does not use 8 as a base. The interpretation of (16g) seems straightforward, with *ʔelk'é* being a combination of the reciprocal prefix *ʔele-* and the postposition *k'é* 'on', i.e., 'on each other four', i.e., $2 \times 4 = 8$; it parallels the expression for 6 ('on each other three'). The interpretation of (16h) is also reasonably straightforward, with *lá* the stem for 'one'; Petitot translates it more literally as "il y en a encore un en bas" ('there is still one down'), referring to the fact that one finger has not been raised, i.e., $10 - 1 = 9$ fingers have been raised. The same is true of (16e), for which Petitot gives the more literal translation "il y en a encore trois (doigts) repliés" ('there are still three of them [fingers] bent'), so three fingers have not yet been raised, i.e.,

10–3=7 fingers have been raised. Both (16e) and (16h) use the implicit minuend 10, as expected in a decimal system. The problem is (16f), which Greenberg interprets as 8–1=7. Petitot explains first how the speaker holds their hands to indicate 7: “Il unit quatre doigts de la main gauche, rapproche du pouce gauche isolé le pouce et l’index de la main droite” (‘He joins the four fingers of the left hand, brings the thumb and index finger of the right hand towards the isolated left thumb’). He then explains the Chipewyan expression more literally as ‘on one side there are four of them’, which is also the English explanation given by Conant. While this may be a reasonable partial description of the configuration of the hands—there are four fingers joined together on one side—the expression seems to involve a fair amount of conventionalization (the left thumb is also extended, as are right thumb and index finger), and it does not lead to any interpretation along the lines of 8–1. This “exception” seems therefore to be spurious.

The third language discussed by Greenberg as an exception to Generalization 14 is Colonial Valley Zapotec (hereafter: CVZ), as described by Córdova (1578). For a fuller analysis of the system, see Sonnenschein et al. (Submitted); here I limit myself to those numbers that are relevant to Generalization 14, and to those numeral expressions that are relevant in the case of alternative expressions. CVZ has a vigesimal numeral system, although the linguistic expressions for the scores are often portmanteau, the intervening tens being expressed by adding 10 to the preceding score, as shown in (17), where the term for 30 is literally ‘twenty-and-ten’.

(17) Colonial Valley Zapotec

- | | | | |
|----|---|-------|-------------------|
| a. | | 20 | calle |
| b. | | 30 | calle-bi-chij |
| c. | 2 | topa | 40 toua |
| d. | 3 | chona | 60 cayona, quiona |
| e. | 4 | tapa | 80 taa |

While there are different possibilities for expressing a number that is 15 above the next lower twenty (or equivalently, 5 below the next higher twenty), our interest will center on a formation that Greenberg identifies as subtractive, as in the version of 55 in (18).

(18) Colonial Valley Zapotec

- | | |
|----------|--------|
| ce-caayo | quiona |
| DEF-five | sixty |
| 55 | |

This example clearly includes an element that means 5 and an element that means 60, so it not illogical to follow Greenberg in describing the construction as subtractive, i.e., $60-5=55$. However, as with some of the other examples discussed in this section, the structure of the CVZ expression does not seem to be literally subtractive, although it can be subsumed under counting down as a subtype that Sonnenschein et al. (Submitted) call a “deficit” construction. The use of the definite prefix before a numeral, as in *ce-caayo*, is attested independently in the language and means ‘another x ’ or ‘ x more’, so a more literal translation of (18) would be ‘another five sixty’, i.e., if you had another five, then the number would be 60—but since there is a deficit of 5, the target number is actually 55. In other words, the literal interpretation of (18) is given by the equation in (19a), not that

in (19b), even though both equations give the same result for x .

- (19) a. $x+5=60$
 b. $x=60-5$

However, since I have suggested extending Greenberg's Generalization 14 on subtraction to counting down constructions more generally, we can still ask the further question whether the minuend (in this extended sense) has to be a base of the numeral system in CVZ. Relevant examples are those between 15 above the next lower twenty and the next higher twenty, i.e., the range 16-19 above the next lower twenty, illustrated in (20) with the relevant CVZ expression for 56.

- (20) Colonial Valley Zapotec
 ce-caayo quiona-bi-tobi
 DEF-five sixty-and-one
 56

Greenberg interprets (20) as in (21a), which means that the minuend is 61, which is not only not a base of the system but also unlikely to be a salient number, given that the same pattern is found for 57-59, which would mean that 62, 63, 64 would also have to be salient numbers, which would deprive the notion of salience of any measurable import. Sonnenschein et al. (Submitted) argue that the correct interpretation of (20) is as in (21b), i.e., the value 56 is arrived at by first counting down 5 from 60 to get an intermediate result of 55, and then adding 1 to that intermediate result; the minuend would remain 60, a multiple of the base 20.

- (21) a. five from (sixty plus one) $(60+1)-5$ (Greenberg)
 b. (five from sixty) plus one $(60-5)+1$ (Sonnenschein et al.)

In effect, CVZ arrives at the value 56 by successive approximation, starting with 60, then counting down by 5 to give 55, then counting up by 1 to give the intended value of 56. A similar procedure is attested in Ket, as shown in (22) (Georg 2007: 181). In (21a), the value 80 is arrived at by subtracting 20 from 100; in (21b), the value 82 is arrived at by adding 2 to this interim result.

- (22) Ket
- | | | | | | |
|----|--------|-----------|---------|---------|---------|
| a. | éks | bánsaŋ | kiʔ | | |
| | twenty | without | hundred | | |
| | 80 | | | | |
| b. | ínam | ákam | éks | bánsaŋ | kiʔ |
| | two | left.over | twenty | without | hundred |
| | 82 | | | | |

In other words, one has to know the rules that native speakers follow intuitively on how to “bracket” a complex numeral expression: in Ket, $((100-20)+2)$, and not $(100-(20+2))$, which latter would give the incorrect output 78. One might object that the bracketing proposed for CVZ examples like (20) does not follow the word division. However, it is unclear what exactly Córdova intended by the placement of orthographic word divisions, and much more likely that this had something to do with the language’s phonological structure than its grammar. The CVZ examples thus consistently show subtraction from a multiple of 20, followed where necessary by a process of addition to arrive at the final target value.

Greenberg (1978: 14) proposes a further universal on subtraction, which we can also extend to other instances of counting down:

Generalization 13: If a number n is expressed by subtraction as $y-$

x , then every number z ($y > z > n$)⁹ is also expressed subtractively and with y as the minuend.

This generalization can be illustrated using the Latin examples in (14). Since 18 is expressed as (20–2), 19 must also be expressible by means of subtraction with 20 as the minuend, which is indeed the case since it is expressed as (20–1). Greenberg notes that Chipewyan is an exception since 7 is expressed as (10–3) but 8 is not expressed as (10–2), even though 9 is expressed as (10–1). CVZ is also a counterexample on Greenberg’s analysis, since 56 is expressed as (61–5), but 57 is not expressed as (61–4), but rather, on Greenberg’s analysis, as (62–5). On the reanalysis proposed by Sonnenschein et al. (Submitted), the minuend is consistently 60 for numbers in the range 56–59 that use a combination of subtraction and addition, so this counterexample is removed. But going further, if 55 is expressed as (60–5), then according to Generalization 13 it should also be possible to express, for instance, 57 as (60–3), and so on. In fact, this is possible, since Córdova gives an alternative expression for 57 as in (23) (and likewise 56, 58, 59), which is clearly interpreted as ‘three more will walk to sixty’—the irrealis in CVZ can express future time reference— i.e., the interpretation is that there is a deficit of 3 with respect to 60.

- (23) Colonial Valley Zapotec
- | | | | |
|------------------|----------|------|--------|
| ce-chona-c-aca | qui-zaha | chaa | cayona |
| DEF-three-IRR-be | IRR-walk | ? | sixty |
| 57 | | | |

⁹ The original has “($z > y > n$)”, which is inconsistent with the examples and with Greenberg’s clear intention—expressing a target number by subtraction from a minuend that is smaller than the target number would not even be possible using only positive natural numbers—another instance of a typographical error that went uncorrected in the reprinting.

So CVZ is not an exception. However, Ket, as illustrated in (22), is another exception, since although 80 can be expressed as (100–20) it is not possible, for instance, to express 81 as (100–19). The relevant factor seems to be that the subtrahend is rather large in relation to the minuend, which means that there is a large number of intervening numbers, some of which would themselves be quite complex. Indeed, this would violate another of Greenberg (1978: 13)’s generalizations.

Generalization 12: A subtrahend is always a simple lexical expression.

However, Ket does violate Generalization 12, as in (24), since the numerals 8 and 9 are themselves complex, being formed by means of subtraction: (10–2) and (10–1) respectively (Georg 2007: 180-181).

- (24) Ket
- | | | | | |
|------|---------|-----|---------|---------|
| ínam | bánsaŋ | qōs | bánsaŋ | kiʔ |
| two | without | ten | without | hundred |
- 92
100–(10–2)

Clearly, we need to know more about subtraction before we will be in a position to reliably formulate meaningful universals of subtraction (and related counting down processes) that stand some chance of being valid. Part of the problem is the relative rarity of subtractive numeral formations, within which there is a preponderance of simple cases with a subtrahend of 1, sometimes joined by 2, while only larger subtrahends enable serious testing of the limits on the operation.

3. The Limits of Numeral Systems

In addition to fairly specific generalizations of the kind discussed in §2, Greenberg also formulates two very general constraints that impose significant limitations on the extent of numeral systems, and these will be the subject of this section. They concern whether numeral systems in natural languages necessarily have an upper limit (§3.1), and whether or not there can be gaps in a numeral system (§3.2).

3.1. The (In)Finiteness of Natural Language Numeral Systems

According to Greenberg, numeral systems in natural languages are always finite, i.e., for each natural language numeral system there is a highest number that can be expressed. This is formulated in Generalization 1:

Generalization 1: Every language has a numeral system of finite scope.

Note that it is clear that many languages have numeral systems of finite scope, and this is not at issue. Indeed, some languages have very restricted systems, such as Mangarayi with only 1, 2 and 3 (Merlan 1982: 92). Greenberg claims that the American English numeral system is finite, and his discussion suggests that this is true at least in his usage. But before addressing this important question, it will be necessary to discuss a few preliminaries.

At issue are numeral expressions in the region that starts with the term *million* and its cognates in other languages. There are basically two different systems in operation, the short scale and the long scale,

the latter with some internal variation, as shown in Table 1.

Table 1. Short and Long Scales for Higher Numerals

	Short Scale	Long Scale	
		Without Intermediaries	With Intermediaries
10^6	million	million	million
10^9	billion		milliard
10^{12}	trillion	billion	billion
10^{15}	quadrillion		billiard
10^{18}	quintillion	trillion	trillion

The short scale was, around the middle of the twentieth century, most associated with American English, but since then it has expanded to encompass most varieties of English, including in Britain. The long scale is used in most languages of continental Western and Central Europe, and was formerly prevalent in Britain. The intermediaries in *-iard* were not used in Britain, although at least some are used in continental Europe, with consistent use being found, for instance, in German. I return to some of these details below.

One relevant question is who actually uses or used the systems in question, in particular who made practical use of the systems, rather than just using the terms in academic discussion (as with linguists), i.e., methodologically actual usage is more reliable than intuition. Traditionally, such high numbers were the purview of astronomers, and it is no accident that such higher numbers are referred to as "astronomical", though they are also encountered in some of the other natural sciences. The increasing size of national and international budgets from the mid-twentieth century means that higher numbers

also came to be used regularly by economists. For languages spoken in countries that experienced hyperinflation, such as Germany in the 1920s, the practical use of such high numbers in economic contexts emerged earlier and even spread to the general populace. Studying linguistic practice under such circumstances can give us insight into how the numeral systems of languages are actually used.

But let us now return to the American English system as described by Greenberg, although some of the details have to be reconstructed from his discussion rather than being stated explicitly. My reconstruction of the relevant points is given in (25).

- (25) Greenberg's American English short-scale system
- a. a new single-word term is required each time one multiplies the previously highest single-word term by 1000, i.e., combinations like **thousand million*¹⁰ are excluded;
 - b. there is a highest single-word term (for Greenberg in 1978: *decillion* (10^{33}));
 - c. the series of intermediate terms (*million, billion, trillion, quadrillion, quintillion, sextillion, septillion, octillion, nonillion, decillion*) is complete, i.e., shows no gaps;
 - d. the highest number expressible is thus ($10^{36}-1$), and all numbers between 1 and $10^{36}-1$ are expressible.

I emphasize that what is at issue is not whether the finite system as described by Greenberg exists, at least for some speakers of American English; I have no reason to doubt Greenberg's characterization of his native-speaker intuitions. What is at issue is whether the finiteness

¹⁰ In citing such large numbers in English, I will consistently omit the leading *a/one*, i.e., *million*, for instance, would in relevant contexts actually be *a/one million*.

limitation is a language universal.

I will now describe the numeral system, as it applies to higher numerals, as I internalized it in Britain in the mid-twentieth century. Lest it be thought that my account is purely academic and viewed through the imperfect lens of decades-old memory, I should point out, first, that as a child I read a lot about astronomy and so did encounter how British writers expressed large numbers, and second, that I was challenged to give an explicit description of this system by a skeptical linguistics instructor in the mid-1960s, which served to reinforce the system in my memory at a time when it was still in regular use. The relevant points are given in (26), with examples in (27); note that (a)-(c) are the crucial points, with (d)-(e) added for completeness.

- (26) Traditional British long-scale system (my version)
- a. a new single-word term is available each time one multiplies the previously highest single-word term by 1,000,000, up to trillion, i.e., there are no terms **quadrillion* and beyond;
 - b. there are no intermediaries, e.g., **milliard*, so 10^9 can only be expressed as *thousand million*;
 - c. numerals from *million* up can be combined to express multiplication, e.g., *million trillion* ($[10^6 \times 10^{18}] = 10^{24}$);
 - d. there is a strong preference to implement the "Packing Strategy" (Hurford 1975: 67), i.e., informally speaking, to avoid combining terms with lower numerical value when there is an available appropriate term with higher numerical value; thus, although *million million* is possible, *billion* is preferred;
 - e. it is not possible for a higher-value term to precede a lower-value term, e.g., for 10^{30} one can say *billion trillion* but not **trillion billion*.

- (27) a. 10^6 million
 b. 10^9 thousand million
 c. 10^{12} billion
 d. 10^{15} thousand billion
 e. 10^{18} trillion
 f. 10^{24} million trillion
 g. 10^{30} billion trillion
 h. 10^{36} trillion trillion
 i. 10^{54} trillion trillion trillion
 j. 8×10^{67} eighty billion trillion trillion trillion¹¹

Crucially, the system is recursive, since some terms, in particular *trillion*, can be repeated as many times as is necessary to express the relevant cumulative multiplication. Incidentally, in evaluating (26c), it is important to distinguish between the numeral and the homophonous noun. In English, this is easily done since the noun will stand in the plural if governed by a numeral greater than 1, while the numeral always remains in the singular. *Trillion* is thus a numeral in (28a), but a noun in (28b). (28d) is possible, since *thousand* is a noun, but not (28c), since numerals below *million* cannot be iterated. Note that the criteria for distinguishing numerals from nouns vary from language to language, so the tests valid for one language cannot simply be transposed to another.

- (28) a. a million trillion
 b. a million trillions
 c. *a thousand thousand
 d. a thousand thousands

¹¹ This number is an approximation of the number of distinct ways of ordering the 52 cards in a standard deck, i.e., (52!).

The systems in (25) and (26) are very different from one another, even if there is overlap in some of the terms used (*million*, *billion*, *trillion*). The expression "two countries divided by a common language" seems particularly apt, even if in the intervening half century British usage has here moved in the direction of American.

One of the characteristics of Greenberg's American English system as reconstructed in (25) is the strict application of the Packing Strategy, e.g., 10^9 must be expressed as *billion* and cannot be expressed as *thousand million*; note that if the latter were possible, then 10^{36} would be expressible as *thousand decillion*. Interestingly, there is one exception to the Packing strategy with lower numerals, presumably also for Greenberg (since it is generally widespread in American English), namely the possibility of expressing numbers in the thousands by counting in hundreds, i.e., (29b) and (29c) are both possible linguistic expressions for (29a), where (29c) violates the Packing Strategy.

- (29) a. \$1,500
 b. one thousand five hundred dollars
 c. fifteen hundred dollars

However, searching on the Internet reveals that at least some speakers of American English do not share the strict constraints implied by Greenberg. Examples (30) and (31) are from the syndicated column of US astronomer Dennis Mammana (on whom further information is available on his web site <http://www.dennismammana.com>)¹², with (30) illustrating *million trillion* (where Greenberg would require *quintillion*) and (31) *thousand trillion* (where Greenberg would require *quadrillion*, in full *three quadrillion five hundred trillion*).

¹² This and all following cited URLs were verified on 2020 Aug 07.

- (30) Even our Milky Way galaxy is spinning like a giant Ferris wheel at about 140 miles per second, and we require some 250 million years to complete the 2 million trillion-mile trek around the galactic center—a “galactic year,” it’s called.
(https://www.noozhawk.com/article/dennis_mammana_theres_a_lot_going_on_as_the_world_turns_20190908)
- (31) Mars currently lies about 189.4 million miles from us, while Antares lies at a distance of about 3,500 trillion miles away—so far that the light we see from it began its journey in our direction around the start of the 15th century!
(https://www.noozhawk.com/article/dennis_mammana_mars_and_its_rival_antares_dawn_on_us_this_week_20200119)

It is also useful to compare the situation in a selection of other languages, which is often very different from that in American English, certainly as described by Greenberg. French uses the long scale with intermediaries, but although terms from *billion* up are in principle available, in practice the highest monolexical term in general use is *milliard*, and higher numbers have to be expressed by using the components *milliard* and lower, as in (32), if necessary repeating *milliard*, as in (33). Both examples, incidentally, are from the prestigious newspaper *Le Monde*. Note that in French numerals from million up must stand in the plural (in *-s*) if preceded by a number higher than 1, and require the preposition *de* ‘of’ before a following numeral or noun, e.g., the relevant part of (32) is literally ‘the thousand milliards of dollars’.

(32) French

Les mille milliards de dollars de capitalisation d'Apple valent celle des dix premières entreprises françaises

'The thousand milliard ($10^3 \times 10^9 = 10^{12}$) dollars of Apple's capitalization is worth that of the first ten French companies' (https://www.lemonde.fr/les-decodeurs/article/2018/08/03/les-1-000-milliards-de-dollars-de-capitalisation-d-apple-valent-celle-des-dix-premieres-entreprises-francaises_5338994_4355770.html)

(33) French

il faudrait tout de même attendre 1 000 milliards de milliards de milliards de milliards de milliards de milliards de milliards de milliards de milliards de milliards de milliards de milliards d'années pour que celui-ci s'évapore complètement

'one would still have to wait 1000 milliard milliard milliard milliard milliard milliard milliard milliard milliard milliard ($10^3 \times 10^9 \times 10^9 \times 10^9 \times 10^9 \times 10^9 \times 10^9 \times 10^9 \times 10^9 = 10^{66}$) years for it to evaporate completely' (https://www.lemonde.fr/les-decodeurs/article/2018/03/14/ce-que-les-travaux-de-stephen-hawking-ont-apporte-a-la-physique_5271032_4355770.html)

German uses the long scale with intermediaries, indeed is often cited as the paradigm example of this type. However, while single lexical items above *Billion* are available, they tend to be avoided in practice; thus (34) avoids *eine Billiarde* 'a billiard', while (35) avoids *eine Quadrillion* 'a quadrillion'. The numerals from *Million* 'million' up stand in the plural when governed by a number above 1, and are written with an upper-case initial, but unlike the corresponding nouns they do not allow an explicit linker to a following numeral or noun, e.g., (34) is literally 'thousand billions friends'.

- (34) German
Tausend Billionen Freunde
'A thousand billion ($10^3 \times 10^{12} = 10^{15}$) friends'
(<https://www.spektrum.de/magazin/tausend-billionen-freunde/1165730>)
- (35) German
Eine Billion Billionen Rechenoperationen gleichzeitig
'A billion billion ($10^{12} \times 10^{12} = 10^{24}$) calculations simultaneously'
(<https://www.computerwoche.de/a/fit-for-future-eine-billion-billionen-rechenoperationen-gleichzeitig,536499>)
- (36) German
Ausgehend von den oben erwähnten Dualitäts-argumenten halte ich die Wahrscheinlichkeit, in die Vergangenheit zu reisen und dann seinen Großvater umbringen zu können, für kleiner als eins zu zehn—mit einer Billion Billionen Billionen Billionen Nullen dahinter.
'Based on the above mentioned duality arguments I consider the probability of traveling into the past and then being able to kill one's grandfather as less than one in ten—followed by a billion billion billion billion billion ($10^{12} \times 10^{12} \times 10^{12} \times 10^{12} \times 10^{12} = 10^{60}$) zeros.'¹³
(<https://musemblaetter.de/artikel.php?aid=16291&suche=>)

In German, I have encountered a number of examples where,

¹³ This example is intended to be humorous, part of an expression of a very low probability, but its structure is impeccable, indeed this is part of the humor. Some parenthetical comments in the original sentence have been omitted.

contrary to principle (26e) in the characterization of the traditional British system (or at least my version of it), larger numbers precede smaller numbers in the set from *Million* up; an example is given in (37). Note that (26e) is not one of the constitutive principles of (26), which are (26a-c). This may indicate greater freedom in German here, though I would not be surprised to find individual variation in acceptability judgments in both languages.

(37) German

Weltweit leben eine Milliarde Millionen Menschen mit Behinderungen.

'Worldwide there live a milliard million ($10^9 \times 10^6 = 10^{15}$) people with disabilities.'

(<http://bezev.fastnetworx.de/index.php?id=201>)

Modern Arabic uses the European-origin terms *milyōn*, etc., following the short scale, though with *bilyōn* and *milyār* competing for 10^9 (Ali 2013: 100-106). In Classical Arabic, however, the highest available numeral was *ʔalf* 'thousand', and "any number above 999,999 was referred to as 1000 multiplied by 1000 as many times as necessary" (Ali 2013: 108). Ali (2013: 108-109) gives textual examples, with (38) being the first part of a longer expression used by the tenth-century traveler, historian, and geographer Al-Mas'udi.

(38) Classical Arabic¹⁴

θamāniyata	šašara	ʔalfa	ʔalfi	ʔalfi
eight	teen	thousand	thousand	thousand
ʔalfi	ʔalfi	ʔalfin		
thousand	thousand	thousand		
18×10 ¹⁸				

These examples should suffice to show that it is possible to combine high numbers, from ‘million’ in some European languages, from ‘thousand’ in Classical Arabic, in order to express a potentially unlimited range of higher number values. While (25) may be the correct characterization of the system used in (some varieties of) American English, it would be a mistake to extend its notion of an upper limit to all natural language numeral systems.

3.2. Gaps in Numeral Systems

We may briefly consider one further generalization proposed by Greenberg (1978), namely Generalization 2.

Generalization 2: Every number n ($0 < n < L$) can be expressed as part of the numerical system in any language.

In Greenberg’s notation, L is the number one above the upper limit of the numeral system, so what this says effectively is that every number from 1 up to the limit of the system is expressible within the

¹⁴ The original is presented in unvocalized Arabic script. The transliteration and vocalization are mine. Note that the Arabic term for 18 requires a following accusative (in *-a*), each of the thousands requires a following genitive (in *-i*); in the sequence of thousands, only the last (the only one that does not itself precede anything) takes the indefinite suffix *-n*.

system. In §3.1 I questioned the universal validity of L, so we can consider a slightly modified version of Generalization 2 that would say that there can be no gaps in a numeral system, i.e., if n is expressible in the system, then so are all numbers below n .

For Greenberg in 1978, the highest single lexical numeral in American English was *decillion* (10^{33}). In the meantime, higher monolexical numerals have come into circulation, e.g., *centillion* (10^{303}). While speakers I have asked have generally accepted that there is such a word and that it would be the hundredth term in the series, they are often at a loss when I ask them to name the twentieth term in the series—though some do come up with *vigintillion*, which is also widely listed in dictionaries—and almost invariably at a loss when I ask them to name the twenty-third term in the series. While there are proposals on how to denominate such terms (e.g., the twenty-third would be *tresvigintillion*—see for instance https://en.wikipedia.org/wiki/Names_of_large_numbers), they have not caught on in usage, nor are most speakers aware of them. For speakers who have constraint (25a) in their system against repetition of higher numerals to express recursive multiplication, this means that there are some numbers between *decillion* and *centillion* that have no natural language expression, i.e., Generalization 2 is counterexemplified. At first it might seem strange that there should be such gaps in “counting”, but it should not be forgotten that no one in their right mind is actually going to try and count up to or beyond a decillion—if you were to count one numeral each second for a year with no breaks, you would only make it up to a paltry 31,556,952—so there seems to be nothing inherently wrong with a system where there are names for certain round numbers but not for certain numbers that occur between the round numbers.

4. Conclusion and Prospect

Needless to say, in selecting among Greenberg's generalizations for discussion, I have chosen those that seem to me to be problematic, leaving aside those I have no reason to question, whether because I have investigated them and found no counterevidence or have not had reason to investigate them. The questions I have raised with respect to the generalizations discussed in this article are a small portion of the total. In all cases, I have been able to challenge the generalizations because they are well stated, i.e., it is clear what would constitute a counterexample, and typically also well illustrated.

Generalizations 1-2 define fundamental properties of numeral systems and are, I hope to have shown, wrong. I believe they stem from universalizing idiosyncratic properties of a specific numeral system, a procedure that is generally quite alien to Greenberg's work on language universals.

For Generalizations 13-14 and 26-27, Greenberg himself noted counterexamples, i.e., for him these were strong universal tendencies rather than literally universals, and the exception found to Generalization 12 would shift it from being an absolute universal to being a universal tendency rather than eliminating it altogether. But with all these generalizations, there has emerged a problem that the feature that would be needed in order to test the universal is rare crosslinguistically—a generally decreasing order of addends in numeral expressions, or widespread use of subtraction (counting down) with different subtrahends—which may mean that we will never have enough relevant cases to be able to test the statistical significance of the claim. But only further work will reveal if such pessimism is indeed well founded; let us hope that future work shows that it is not. Clearly, much more documentation and analysis of

numeral systems is needed.

References

- Ali, J. 2013. *Using Numbers in Arabic*. Washington DC: Georgetown University Press.
- Amaechi, M. 2014. The Syntax of Po Tangle Numerals. *Journal of Universal Language* 15.2, 35-54 (<https://doi.org/10.22425/jul.2014.15.2.35>).
- Bárcena, A. n.d. [1896]. Arte y Vocabulario de la Lengua Toba. *Revista del Museo de La Plata* 7, 189-227.
- Bowen, J. & T. Rhys Jones. 1960. *Teach Yourself Welsh*. London: English Universities Press.
- Calder, G. 1923. *A Gaelic Grammar, Including a Chapter on Personal and Place Names*. Glasgow: Gairm Publications.
- Chan, E. 2020. Numeral Systems of the World's Languages. Available at <<https://mpi-lingweb.shh.mpg.de/numeral>>.
- Comrie, B. 2005. Endangered Numeral Systems. In J. Wohlgemuth & T. Dirksmeyer (eds.), *Bedrohte Vielfalt: Aspekte des Sprach(en)tods* 203-230. Berlin: Weißensee Verlag.
- Comrie, B. & R. Guirardello-Damian. 2004. Trumai and the Internal Syntax of Numeral Expressions. Paper presented at the Conference Syntax of the World's Languages (SWL 1), Leipzig.
- Conant, L. 1931. *The Number Concept: Its Origin and Development*. New York: Macmillan.
- Córdova, J. 1578/1886. *Arte del Idioma Zapoteca*. Morelia: Imprenta del Gobierno.
- Dahl, O. 1968. *Contes Malgaches en Dialecte Sakalava: Textes, Traduction, Grammaire et Lexique*. Oslo: Universitetsforlaget.

- Georg, S. 2007. *A Descriptive Grammar of Ket, Part I: Introduction, Phonology and Morphology*. Folkestone: Global Oriental.
- Greenberg, J. 1978. Generalizations about Numeral Systems. In J. Greenberg et al. (eds.), *Universals of Human Language III: Word Structure* 249-295. Stanford, CA: Stanford University Press. (Reprinted with corrections in K. Denning & S. Kemmer (eds.), *On Language: Selected Writings of Joseph H. Greenberg* 271-309. Stanford, CA: Stanford University Press, 1990).
- Guirardello, R. 1999. A Reference Grammar of Trumai. Ph.D Dissertation, Rice University.
- Hurford, J. 1975. *The Linguistic Theory of Numerals*. Cambridge: CUP.
- Jaker, A. 2014. Chipewyan (Dëne Sùłhne). Available at <<https://mpi-lingweb.shh.mpg.de/numeral/Chipewyan.htm>>.
- Kewley Draskau, J. 2008. *Practical Manx*. Liverpool: Liverpool University Press.
- Merlan, F. 1982. *Mangarayi*. Amsterdam: North-Holland.
- Oftedal, M. 1956. *The Gaelic of Leurbost, Isle of Lewis*. Oslo: Aschehoug.
- Parks, D. 2009. Arikara. Available at <<https://mpi-lingweb.shh.mpg.de/numeral/Arikara.htm>>.
- Petitot, E. 1876. *Dictionnaire de la langue Dènè-Dindjié; dialectes Montagnais ou Chippewayan, Peaux de Lièvre et Loucheux renfermant en outre un grand nombre de termes propres à sept autres dialectes de la même langue; précédé d'une monographie des Dènè-Dindjié, d'une grammaire et de tableaux synoptiques des conjugaisons*. Paris: Ernest Leroux.
- Rajaonarimanana, N. 2001. *Grammaire moderne de la langue malgache*. Paris: L'Asiathèque.
- Sanusi, I. & L. Yusuf. 2018. The Mathematical Operations in Bâtònū

Numeral System. *Journal of Universal Language* 19.1, 57-84
(<https://doi.org/10.22425/jul.2018.19.1.57>).

Sonnenschein, A. et al. (Submitted) Four Zapotec Numeral Systems.

Steinen, K. 1894. *Unter den Naturvölkern Zentral-Brasiliens. Reiseschilderung und Ergebnisse der zweiten Schingú-Expedition, 1887-1888*. Berlin: Dietrich Reimer.

Wied, M. 1841. *Reise in das innere Nord-America in den Jahren 1832 bis 1834 II*. Koblenz: J. Hölscher.