Macrolinguistics, Vol.10 No.1 (Serial No.16) 2022, pp.1-35
DOI: 10.26478/ja2022.10.16.1

The Arithmetic of Natural Language: Toward a typology of numeral systems *

Bernard Comrie **
(University of California, Santa Barbara, USA)

Abstract: Numeral systems in natural languages show astonishing variety, though with very strong unifying tendencies that are increasing as many indigenous numeral systems disappear through language contact and globalization. Most numeral systems make use of a base, typically 10, less commonly 20, followed by a wide range of other possibilities. Higher numerals are formed from primitive lower numerals by applying the processes of addition and multiplication, in many languages also exponentiation; sometimes, however, numerals are formed from a higher numeral, using subtraction or division. Numerous complexities and idiosyncrasies are discussed, as are numeral systems that fall outside this general characterization, such as restricted numeral systems with no internal arithmetic structure, and some New Guinea extended body-part counting systems.
Keywords: numeral system, base of numeral system, arithmetic operation in natural language, typology, constituent order, ambiguity

1. Introduction

The aim of this article is to provide a systematic account of the variation that is found in numeral expressions in natural languages, concentrating on the arithmetic principles that are used in order to construct natural language cardinal numeral expressions. This concentration on arithmetic principles, on the one hand, encompasses considerable cross-linguistic diversity, since different languages often use constructions that differ considerably from those found in more familiar languages such as English and Chinese, while on the other hand it also restricts the domain to manageable proportions by restricting attention to cardinal numerals (to the exclusion of ordinal numerals, fractions,

---

* Presentations based on the material in this article have been given at various fora over the past quarter of a century. I am grateful to all those who participated in the ensuing discussions or otherwise contributed comments. Where no source is given for an example, I have used my own knowledge of the language in question, supplemented where necessary by checking (e.g., spelling) in a standard dictionary.
** Dr. Bernard Comrie: Research Associate and former Distinguished Faculty Professor of Linguistics, University of California, Santa Barbara, USA; Director and Scientific Member Emeritus, Max Planck Institute for Evolutionary Anthropology, Leipzig; Honorary Professor Emeritus of Linguistics, University of Leipzig. E-mail: comrie@ucsb.edu.

non-numerical quantifiers like *many*, etc.) and by excluding morphosyntactic properties of numeral expressions that go beyond the formation of the numeral expression itself, for instance, how numeral expressions are embedded in larger constructions such as noun phrases.

Cardinal numerals, in informal terms, answer to the question *how many?* (though excluding answers involving fractions). In slightly more technical mathematical terms, they denote the cardinality of a set. One might wonder: Why not simply say that they are the numerals used in counting? There are two reasons for not adopting this definition. First, some languages have at least some special numeral forms used in counting, distinct from the cardinal numerals, e.g., the cardinal numeral 1 in Russian is *odin*, but it is replaced by *raz* (literally 'once') in counting. But more significantly, counting is not the only way of establishing the cardinality of a set, and there are languages with restricted numeral systems (see Sect. 2) whose speakers do not count. There are essentially two ways in which one can establish the cardinality of a set, as can be illustrated by a psychological experiment in which the subject is shown a number of dots scattered randomly on a screen. If the number of dots is larger than about 4, then the subject has no option but to count them. However, if the number of dots is 3 or less, then the subject will immediately recognize how many dots there are without the need to count them. This second ability is a very basic human cognitive ability referred to technically as subitizing, with the corresponding verb to subitize. The term denotes the rapid, accurate, and confident judgment of numbers performed for small numbers of items with a cut-off point around 4 (Dehaene, 2011:55-60).

A note on terminology: In this article, number is a mathematical, more specifically an arithmetic concept referring to the cardinality of a set. Numeral – used both as a noun and as an adjective (e.g., in *numeral expression*) – is a linguistic concept, referring to a linguistic expression that denotes or expresses a number; numerals are thus specific to particular languages. When necessary, figure is used to refer to the representation of a number in terms of digits. Thus, the number of fingers on one hand can be represented by the English numeral *five* or by the figure 5. Although figures are often used to indicate the meaning of numeral expressions, this article does not aim to discuss the structure of systems of notation using figures, for which see Chrisomalis (2010). Both numerals and figures do, however, face the same challenge: how to represent a potentially infinite, and certainly very large (in most cultures) range of numbers in an efficient way that neither uses too many primitive elements (which would place an unnecessary burden on the memory) nor involves expressions that are unduly long (which would lead to practical problems of production and comprehension); in other words, how to make infinite (or at least extensive) use of finite means.[①]

One factor to bear in mind as we look at different kinds of numeral systems is that not only are many languages endangered, including several of the languages whose numeral systems we will be looking at, but numeral systems, if anything, are even more endangered than the languages, i.e., many languages that are not themselves high on the list of language endangerment nonetheless have numeral systems that are under strong pressure from other languages, and indeed in some cases have succumbed completely to such pressure, the inherited numeral system being replaced by that of the encroaching language. For further discussion, see Comrie (2005).

Finally, in this introduction, a note should be taken of the fact that a given language may have more than one numeral system, so that strictly speaking, the objects we will be examining are not languages but rather numeral systems in languages. Reference to cases of languages with multiple numeral systems will be made where appropriate.

2. Restricted systems, with little or no internal structure

Many languages of the world, primarily in Australia and Amazonia, have restricted numeral systems, often going no higher than 3 or 4, i.e., the limit of subitizing; speakers of such languages typically do not traditionally count. But the most restricted numeral systems are those of anumeric societies – here, there is no numeral system, although the language will still have quantifiers of the 'few', 'many' type. A famous example of such a language is Pirahã[②] (Everett, 2005:623-627).

Typical of indigenous languages of Australia is a system going up to 3, as in Mangarrayi example (1).

(1) 1 *(ŋa)wumbawa*
2 *ŋabaranwa*
3 *ŋabaḷawa*
[Mangarrayi (Merlan, 1982:92)]

Another indigenous language of Australia, Yidiny, is shown in (2) as having a numeral system going up to 5, with the word for 5 identical to that for 'face of the hand', and the variation in the terms for 4 suggesting that they may not be fully conventionalized.

(2) 1 *guman*
2 *jambula*
3 *dagul*

① The question of whether natural language numeral systems can be infinite, or whether each such system has an upper limit, however high, is not uncontroversial; see Comrie (2020:69-79) for some discussion.
② A list of all languages cited in this article, with genetic affiliation and geographical location, can be found in the Appendix.

4 *yunggan.gunyjii* or *mugungabi*

5 *mala* 'face of the hand'

[Yidiny (Dixon, 1991:224)]

Hixkaryána, an indigenous language of Amazonia, is described in (3) as having terms for 1-5 and 10, i.e., with a gap in its numeral system, though the numerals for 4, 5, and 10 are described as "rarely used", "non-basic", and "also used without precision". They also bear a clear relation to the visualization of finger counting.

(3) 1 *towenyxa*

2 *asako*

3 *osorwawo*

4 *towtɨnke rye* 'its brother twice over'

5 *kamorɨ ɨrakay (o) me* 'half of our hands'

10 *kamothɨrɨ tkatxehkaxe ro* 'our hands completely'

[Hixkaryána (Derbyshire, 1979:154-155)]

The numeral system of Mundurukú, another language of Amazonia, has been examined in detail not only with regard to its internal structure but also in connection to its effect on performance of Mundurukú-speaking subjects in various cognitive tests relating to number, in comparison with subjects who speak languages with more extensive numeral systems. While the numeral systems illustrated in (1)-(3) do not make use of arithmetic processes, that of Mundurukú does, as shown in (4), although there are some idiosyncrasies, e.g., the word for 1 in the numeral for 4 is different from that in the numerals for 1 and 3, while the numeral 2 on its own is different from that found in 3-4. It might therefore have been treated in Sect. 3, but is mentioned here because of the work on the psycholinguistics of number, which is included in the cited reference.

(4) 1 *püg (ma)*

2 *xep xep*

3 *eba-püg* 2 + 1

4 *eba-dip-dip* 2 + 1 + 1

5 *püg pögbi* 'one hand'

[Mundurukú (Pica et al., 2004)]

There are two important take-home messages from work that has been done on languages that have only restricted numeral systems. First, one can establish an evolutionary development in natural language from less to more expressive numeral systems (i.e., in terms of the range of numbers that can be expressed), making numerals one of the few areas in languages where this is clearly possible; see further Everett (2017). Second, the differences, including different performances under psycholinguistic testing, relate to the realization of numeracy in the individuals and societies concerned, not to any difference in innate numerical ability: When exposed to more extensive numeral systems, speakers of languages with only a restricted numeral system can learn them and develop responses in psycholinguistic tests identical to their speakers.

3. Restricted systems with addition only

A number of languages have a restricted numeral system, typically going up to 4, that is binary and additive, i.e., one can form higher numerals by adding to a base (see Sect. 4.1). Such a restricted system often exists alongside another system that allows for the construction of higher numerals. Restricted systems are found in particular in a number of indigenous languages of New Guinea, such as Haruai, which has the transparent system given in (5). The terms for 1 and 2 are primitives, 3 is 2 + 1, while 4 is 2 + 2, in both cases with no overt indication of addition.

(5)	1	*paŋ*	
	2	*mös*	
	3	*mös paŋ*	2 + 1
	4	*mös mös*	2 + 2

[Haruai (Comrie, own fieldwork)]

It should be noted that the Haruai system is completely conventionalized. The terms for 1-4 are used frequently and are volunteered by speakers when they are asked about numerals. While it might seem logical to continue the system with *mös mös paŋ*, etc., such formations have not been encountered spontaneously and are rejected by speakers. The neighboring, though probably genetically unrelated language Kobon has a less transparent system, as shown in (6).

(6)	1	*añi (nibö)*
	2	*mɨhöp*
	3	*mɨhau nɨgaŋ*
	4	*mɨhau mɨhau*

[Kobon (Davies, 1981:208, with minor transcription corrections provided by the author)]

Here, the term for 2 used in 3-4 is slightly different from that used for 2 in isolation, while the second element in the numeral for 3 bears no relation to that for 1.

Occasionally, this binary additive process is extended beyond 4, as in the Adzera material in (7), where it is continued through 9.

(7)	1	*bits*	
	2	*iruʔ*	
	3	*iruʔ da bits*	2 + 1
	4	*iruʔ da iruʔ*	2 + 2
	5	*iruʔ da iruʔ da bits*	2 + 2 + 1

6 *iruʔ da iruʔ da iruʔ* 2 + 2 + 2

etc.

10 *dzi bangi marafain da dzi bangi marafain sib*

'my hand half and my hand half completed'

[Adzera (Lean & Owens, 2018:299, partially checked against Smith, 1988:66 and Holzknecht, 1989:127-128)]

Restricted systems with addition only often form a closed system on their own, often alongside another system that allows for the expression of a wider range of numbers, also starting from 1, e.g., the extended body-part numeral systems of Haruai and Kobon (Sect. 4.2). Sometimes, however, they are integrated into a single overall numeral system, expressing lower numbers in combination with a different strategy for expressing higher numbers, as in the case of Adzera (see (7)), and also, though minimally, Mundurukú (see (4)), since the expression for the only higher number in the language, namely 5, falls outside the additive system.

4. More complex systems using multiplication and addition applied to a base

Addition on its own allows the expression of a larger range of numerals than a set of morphologically unanalyzed expressions, e.g., Haruai requires only two rather than four primitive expressions to express the range 1-4, as shown in (5). However, in order to extend the range even further while keeping the number of primitives limited, it is necessary to supplement addition with multiplication.

4.1 Clear instances of systems using multiplication and addition applied to a base

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, as illustrated in Mandarin Chinese example (9a).

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

The number m is nearly always smaller than b (but see (28) for a rare instance where $m = b$), while the number n is usually smaller than b, though there are occasional examples where $n = b$ (see (32)) or even $n>b$ (see (42)). Note that when m is 0, it is never expressed in natural language numeral expressions, as illustrated in (9b); while when n is 1, it is expressed with some numerals in some languages (e.g., with the word for 10 in Nauete as in (17), but not in Mandarin Chinese as in (9c)). Furthermore, in the formula in (8) the order of elements is not significant, although the bracketing must be retained. The Mandarin Chinese example (9) illustrates the order $(n \times b) + m$, while the Birom example

(13) illustrates ($b \times n$) + m, and the first three words of the Malagasy example (58) illustrate m + ($n \times b$). One can think of the succession of products of the base (e.g., 20, 30, 40, etc. in a decimal system) as higher additive bases, to which can be added a value in the range 1-9 in order to express numbers between the products, e.g., 25.

The operation of the formula in (8) will now be illustrated by examining different bases, i.e., values of b in (8).

The most widespread base cross-linguistically is 10, giving rise to the system called decimal, as illustrated in (9), where the value 54 is arrived at by multiplying the base 10 by 5 and adding 4 to the result.

(9) a. *wŭ-shí* *sì*
five-ten four
54 [(5 × 10) + 4]

b. *wŭ-shí*
five-ten
50 [5 × 10]

c. *shí-sì*
ten-four
14 [10 + 4]
[Mandarin Chinese]

The format of (9) is one that we will use regularly throughout the article: The first two lines are standard original and gloss lines. The third line shows the number value of the expression in figures, accompanied where appropriate by an arithmetic formula showing how this result is arrived at. Sometimes, as in Mandarin Chinese, there is no overt indication of the operations of addition and multiplication, and one simply has to know the language-specific rules of when to multiply and when to add, i.e., to know that the relevant formula is ($n \times b$) + m in that order. However, many numeral systems do include overt indicators (or links) of one or both arithmetic operations, as illustrated in various examples in the remainder of the article; one might compare current English *twenty-four*, with no overt marker of addition, with the archaic *four-and-twenty*, with the conjunction *and*. See Greenberg (1978[1990]:283-284) for classification of links used to express addition.

The second most widespread base cross-linguistically is 20 (giving a vigesimal system), as illustrated in (10).

(10) *kəlgən-qlekken* *məngətken* *ŋireq* *parol*
fifteen-twenty ten two left
312 [(15 × 20) + (10 + 2)]
[Chukchi (Skorik, 1961:390)]

Other bases are much rarer cross-linguistically, though (11)-(14) illustrate bases of 60, 32, 12, and 8 respectively.[①]

(11)	a.	*èna*	*ma*	*gàati*	*dàimita*	*mutò*
		one	and	ten	and	sixty
		71 [60 + (10 + 1)]				
	b.	*muto*	*wìi*			
		sixty	four			
		240 [4 × 60]				
		[Ekari (Drabbe, 1952:30)]				

(12)	*ɨ̀fɔ*	*wǎdhɨ̀*			
	four	thirty-two			
	128 [4 × 32]				
	[Ngiti (Lojenga, 1994:355-358)]				
(13)	*ba-kuru*	*ba-ba*	*ná*	*CL-ā*	*CL-bā* [②]
	PL-twelve	PL-two	plus	CL-this	CL-two
	26 [(2 × 12) + 2]				
	[Berom (sometimes spelt as Birom)(Bouquiaux, 1970:259)]				
(14)	*karnuʔ*	*tenhiuɲ*	*rnuʔ*		
	three	eight	three		
	27 [(3 × 8) + 3]				
	[Northern Pame (Avelino, 2006:45)]				

Finally in this selection of unusual bases, we present an example of base 6 from Komnzo in (15) (see further (37)-(38)). This numeral system is, incidentally, used only for counting yams; there is also a separate restricted binary additive system for 1-4 (cf. Sect. 3). C. Döhler has posted a video of Komnzo Yam counting at: https://vimeo.com/54887315.

(15)	*eda*	*nibo*	*a*	*eda*
	two	six	and	two
	14 [(2 × 6) + 2]			
	[Komnzo (Döhler, 2018:94, extracted from longer example (38))]			

If in a numeral system the lowest multiplicative base is higher than about 12, and the system is not an extended body-part counting system (see Sect. 4.2), then the higher numerals below the lowest multiplicative base make use of an additive base, e.g., Georgian (lowest multiplicative base: 20; lower additive base: 10), illustrated in (16); some languages use this even for a smaller lowest multiplicative base, e.g., Nauete (Naueti) (lowest multiplicative base: 10; lower additive base: 5), as shown in (17).

---

① The examples of the bases 60 and 32 are strictly speaking higher bases (Sect. 5), though they still illustrate the general point.

② The last two elements *-ā* and *-bā* require a class prefix corresponding to the class of the head noun.

(16)	1	*ert-i*	*-i* nominative case (NOM)
	2	*or-i*	
	5	*xut-i*	
	6	*ekvs-i*	
	9	*cxra*	NOM drops after *a*
	10	*at-i*	
	11	*t-ert-met'-i*	teen-one-teen-NOM
	15	*t-xut-met'-i*	
	16	*t-ekvs-met'-i*	
	19	*[t-]cxra-met'-i*	(*t* + *c*>*c*)
	20	*oc-i*	
	21	*oc-da-ert-i*	twenty-and-one-NOM
	30	*oc-da-at-i*	twenty-and-ten-NOM
	31	*oc-da-t-ert-met'-i*	twenty-and-teen-three-teen-NOM
	40	*or-m-oc-i*	two-Ø-twenty-NOM

[Georgian (Hewitt, 1995:51-52)] [1]

(17)	2	*kai-rua*	CL-two
	3	*kai-tolu*	CL-three
	5	*kai-lima*	CL-five
	7	*kai-lima resi kai-rua*	CL-five and CL-two
	8	*kai-lima resi kai-tolu*	CL-five and CL-three
	10	*weli-see*	ten-one
	12	*weli-see resi kai-rua*	ten-one and CL-two
	17	*weli-see resi kai-lima resi kai-rua*	ten-one and CL-five and CL-two
	20	*weli-kai-rua*	ten-CL-two

[Nauete (Chan, 2022, https://lingweb.eva.mpg.de/channumerals/Naute.htm, with data provided by G. Hull in 1996; see also Schapper & Hammarström, 2013:429, and Greenhill et al., 2008, https://abvd.eva.mpg.de/austronesian/language.php?id=1365, the last with data provided by A. Veloso in 2016. There are differences in detail among the sources, not affecting the main point.)]

For an account of the geographic distribution of different numeral bases across a sample of the world's languages, see Comrie (2013).

4.2 New Guinea Highland body-part counting systems

Extended body-part numeral systems are found in New Guinea, attested in many (but by no means all, or even most) Papuan languages. Their characteristic is the use of body parts

[1] The circumfix *t-...-met'* expresses 'teen'.

going beyond just the ten fingers (or the twenty fingers and toes), typically counting the fingers of one hand, then going up that arm, then across the top of the body, with a center point such as the hole above the sternum (breastbone) or the ridge of the nose, then down the other arm, finally counting the fingers of that hand. A numerical value may be indicated by touching the appropriate body part and saying its name, or by simply touching the body part, or by simply saying its name. The last possibility is crucial, since it shows that such numerals are part of the language and not just a gestural system.

A distinction must be made between symmetric and asymmetric extended body-part systems. In a symmetric system, the same body parts are used on the second side of the body as on the first side of the body, all in the reverse order. Kobon, as shown in (18), has such a symmetric system.

(18)	*wañig nibö*	little finger	1	23	24	46
	igwo	ring finger	2	22	25	45
	igwo aŋ nibö	middle finger	3	21	26	44
	igwo milö	index finger	4	20	27	43
	mamid	thumb	5	19	28	42
	kagoł	wrist	6	18	29	41
	mudun	forearm	7	17	30	40
	raleb	inside of elbow	8	16	31	39
	ajip	biceps	9	15	32	38
	siduŋ	shoulder	10	14	33	37
	agip	collarbone	11	13	34	36
	migan	hole above sternum	12		35	

[Kobon (Davies, 1981:206-208, with minor transcription corrections provided by the author)]

It is usual to start with the little finger of the left hand, and we will assume this orientation for all of the following discussions. The count proceeds across the five fingers of the left hand, then to five positions on the left arm from wrist to shoulder, then the collar bone followed by the center point, the hole above the sternum. The count continues by going through the body parts in reverse order down the right side, starting from the collar bone and ending with the little finger of the right hand. This constitutes one complete pass across the body, in Kobon reaching the numerical value 23. The name of a body part is the same for both sides of the body, but the second half of the body can be identified by adding *böŋ* ‘one side of’. The count can then be continued by starting with the little finger of the right hand, moving up the right side of the body to the center point and then down the left side of the body; the complete second pass reaches 46. The second pass can be identified by adding *ñin juöl adog da* ‘pulling out the hand, give back!’, and the third and subsequent passes can be specified by adding a numeral to indicate which pass across the body is current. The system can overall be described as having an additive-multiplicative base 23. It also has the interesting arithmetic property that the sum of the identical body part terms on the two sides of the body gives a constant total of 24 on the first pass, 70 on the second pass, and so on.

Haruai (19) will illustrate the distinctive feature of an asymmetric system, clearly so in that it uses exactly the same body parts as Kobon.

(19)	*aglŋ*	little finger	1			19				37
	aglŋ rolyöbö	ring finger	2			20				38
	wölö ml	middle finger	3			21				39
	köñö ngb	index finger	4			22				40
	mömd	thumb	5			23				41
	wrap cgb	wrist	6		18		24		36	
	mj	forearm	7		17		25		35	
	amñaxb	inside of elbow	8		16		26		34	
	mac	biceps	9		15		27		33	
	möyb	shoulder	10		14		28		32	
	katlöy	collarbone	11		13		29		31	
	mgan	hole above sternum		12				30		

[Haruai (Comrie, own fieldwork)]

The count is exactly the same as in Kobon up to the wrist of the right hand, numerical value 18. The next body part is not, however, the thumb of the right hand, but rather the little finger of the right hand, with the count continuing to the thumb, then to the wrist, and so on. This means that at the transition between passes across the body, the fingers of a given hand are counted only once, from little finger to thumb, introducing an asymmetry relative to the Kobon system, which at this point counts the fingers backwards (from thumb to little finger) and then forwards (from little finger to thumb). In Haruai, the second side of the body can be identified explicitly by saying *adök*^{w}*ebö* 'of that side'. The second pass across the body can be identified by saying *höwöylp* or *höbkalp*, literally 'returning'. Arithmetically, one might argue that the Haruai system has an additive-multiplicative base of 18. However, native speakers do not view the system in this way, but view the first count as terminating at 23, the second count at 41, and so on, i.e., the numerals 19-23, 37-41, etc., belong to the preceding instead of the following pass. The termination of the first count can be identified by saying *padö=k*^{w}*ö paŋyöbö dwa* 'one house post over there went', with subsequent counts identified by replacing *paŋyöbö* 'one' (attributive form) with the appropriate number (and replacing the singular verb *dwa* with plural *dwm*^{y}*a*). This means that the first pass across the body reaches 23, while each subsequent pass adds 18 to the total, giving a system that does not lend itself easily to simple characterization in terms of a base (which would be something like [23 + (n × 18)], where *n* may be 0).

A video illustrating the extended body-part system of Foi (Foe) by G. Sebo is available at https://www.youtube.com/watch?v=H13Se4nBPDA. Foi has a symmetric system, with a total of 37 on each count – Kobon and Haruai are, incidentally, at the lower end for the number of body parts distinguished. The forms are set out in (20), based on Franklin (2001:152-153).[①]

(20)	*mena-gi*	little finger	1	37
	ha-gi	ring finger	2	36
	i-gi	middle finger	3	35
	tugu-bu	index finger	4	34
	kaba	thumb	5	33
	tama	palm	6	32
	bona-gi	wrist	7	31
	kwebo	forearm	8	30
	karo-habo	inside elbow	9	29
	ame-ni	upper middle arm	10	28
	ki	shoulder	11	27
	keno	collarbone area	12	26
	heno-go	lower neck	13	25
	fufu	upper neck	14	24
	kia	ear	15	23
	bobo	cheekbone	16	22
	i	eye	17	21
	to	side of nose	18	20
	kisi	ridge of nose	19	

[Foi (Foe) (Franklin, 2001:152-153)]

4.3 Idiosyncrasies relating to bases

Mandarin Chinese presents a very regular decimal system using addition and multiplication, and for this reason it was used for the first introduction of such a system. However, a number of languages have what can be clearly discerned as a system with a base, but nonetheless with certain irregularities. Several classes of such exceptions are now presented.

4.3.1 Portmanteau and other morphologically irregular forms

An extreme example is provided by so-called portmanteau morphs, i.e., forms that lack

①. A formatting error in the published source means that the numerals and body parts from 12/13 onwards are incorrectly aligned; this has been corrected and the material has been supplemented against the video.

the expected compositionality of regular numerals in the language, as with Russian *sorok* 40 in (21). The preceding and following tens can be decomposed into a second part conveying the sense 10 and a first part conveying the sense of the numeral by which 10 must be multiplied to give the target value.

(21)

units		tens	
2	*dva*	20	*dva-dcat′*
3	*tri*	30	*tri-dcat′*
4	*četyre*	40	*sorok*
5	*pjat′*	50	*pjat′-desjat*
6	*šest′*	60	*šest′-desjat*

[Russian]

Another example is provided by the English expression for 11, as in (22), which is not decomposable synchronically into elements meaning 1 and 10.[1]

(22) *eleven*
11 [expected 10 + 1]
[English]

A striking set of portmanteau morphs is provided by Balinese (with similar formations in Javanese and Madurese). The original Austronesian numeral system of the area was transparently decimal, but in Balinese some of the expected Austronesian reflexes were replaced by expressions related to ways of packaging Chinese coins, used in commercial transactions, as shown in (23). Note that the initial *s(e)-* is a bound form of the numeral 1.

(23)

25	*se-lae*	‘one thread (of Chinese coins)’
45	*se-timan*	‘one opium packet (costing 45 Chinese coins)’
50	*se-ket*	‘one tie (i.e., two threads of 25 Chinese coins)’
75	*telung benang*	‘three threads (of Chinese coins)’
200	*s-atak*	‘one bundle of 200 Chinese coins’
400	*s-aman*	‘one gold (coin worth 400 Chinese coins)’
900	*sanga*	[etymology unclear]

[Balinese (Eiseman, 1990:162-168)]

Sometimes the irregularity is less than complete suppletion, with the parts of the expression still recognizable although modified by changes to one or other component that go beyond regular phonological rules of the language in question, as can be seen in the English expressions in (24).

(24) *fif-teen* (**five-teen*) *twenty* *twelve*
five-ten

[1] Etymologically, the formation is ‘one left-over’, but this is no longer recoverable synchronically.

15
[English]

In the expression for 15, the expected first component *five* [faɪv] has been shortened to *fif-* [fɪf]. In the expression for 20, [ˈtwɛnti], the last part is recognizable as the usual second part of expressions for the tens in English (cf. *sixty* 60), and the first part bears some resemblance to *two* [ˈtu], but the intervening segments are unclear, although they do reappear in 12, [ˈtwɛlv], whose last part is, however, not recoverable synchronically; etymologically, it is related to the last part of 11.

In Hindi (25), arguably, all the numerals 1-99 are irregular, in the sense that it is never possible to predict unequivocally how a particular combination will be formed, although it is usually possible to identify a first part expressing the unit and a second part expressing the ten. Note that the expressions with 9 as unit value, from 19 through 79, are based on a subtractive expression (see Sect. 6.1), e.g., 29 is ‘1-less-than 30’.

(25)

	0	1	2	3	4	5	6	7	8	9
		ek	*do*	*tīn*	*cār*	*pāṁc*	*chah*	*sāt*	*āṭh*	*nau*
10	*das*	*gyārah*	*bārah*	*terah*	*caudah*	*pandrah*	*solah*	*satrah*	*aṭhārah*	*unnīs*
20	*bīs*	*ikkīs*	*bāīs*	*teīs*	*caubīs*	*paccīs*	*chabbīs*	*sattāīs*	*aṭṭhāīs*	*untīs*
30	*tīs*	*ikattīs*	*battīs*	*taiṁtīs*	*cauṁtīs*	*paiṁtīs*	*chattīs*	*saiṁtīs*	*aṛtīs*	*untālīs*
40	*cālīs*	*iktālīs*	*bayālīs*	*taiṁtālīs*	*cavālīs*	*paiṁtālīs*	*chiyālīs*	*saiṁtālīs*	*aṛtālīs*	*uncās*
50	*pacās*	*ikyāvan*	*bāvan*	*tirpan*	*cauvan*	*pacpan*	*chappan*	*sattāvan*	*aṭṭhāvan*	*unsaṭh*
60	*sāṭh*	*iksaṭh*	*bāsaṭh*	*tirsaṭh*	*cauṁsaṭh*	*paiṁsaṭh*	*chiyāsaṭh*	*sarsaṭh*	*aṛsaṭh*	*unhattar*
70	*sattar*	*ik’hattar*	*bahattar*	*tihattar*	*cauhattar*	*pac’hattar*	*chihattar*	*sat’hattar*	*aṭhhattar*	*unyāsī*
80	*assī*	*ikyāsī*	*bayāsī*	*tirāsī*	*caurāsī*	*pacāsī*	*chiyāsī*	*sattāsi*	*aṭṭhāsī*	*navāsī*
90	*navve*	*ikyānve*	*bānve*	*tirānve*	*caurānve*	*pacānve*	*chiyānve*	*sattānve*	*aṭṭhānve*	*ninyānve*

[Hindi (McGregor, 1972:61-62; note that some numerals have minor variants)]

4.3.2 Sporadic bases

Another way in which a language may depart from the ideal combination of addition and multiplication applied to a consistent base is by having one or more exceptional numerals constructed using a base that is different from the regular base. For example, French has basically a decimal system, but the expression for 80 is vigesimal, and the vigesimal pattern holds for the range 80-99, as in (26).

(26) *quatre-vingt-douze*
four-twenty-twelve
92 [(4 × 20) + 12]
[French]

In the traditional Welsh numeral system, which is vigesimal with 10 and 15 also serving as additive bases, the expression for 18 is literally 'two nines' (27), although 9 is nowhere else used as a base.[①]

(27) *deu-naw*
two-nine
18 [2 × 9]
[Welsh (King, 1993:113)]

4.3.3 Overrunning

"Overruning" is a term specially created to refer to cases where, given the base of a language, one would expect multiplication to be used, but instead, addition is continued. An imaginary example in English would be to continue the series *seventeen*, *eighteen*, *nineteen*, to give **tenteen*, rather than the correct form *twenty*. Precisely this formation is attested in Polabian (28).

(28) *disąt-nocti*
10-teen
20 [10 + 10]
[Polabian (Polański & Sehnert, 1967:52)]

Note that as one counts on from 20 in Polabian, the expression in (28) is used as a product of the base to express the tens and to which the units are attached, as in (29), i.e., one says '1 + 20' (more fully: [1 + (10 + 10)]), rather than '11 + 10' (more fully: [(1 + 10) + 10], i.e., the analog of English **eleventeen*).

(29) *janü* *disąt-nocti*
1 10-teen
1 + [10 + 10]
[Polabian (Polański & Sehnert, 1967:52, 73)]

The French numerals 70-79, as illustrated in (30) can be considered an instance of overrunning, since they would correspond to pseudo-English expressions of the type **sixty-ten*, **sixty-twelve*, **sixty-seventeen*.

(30) a. *soixante-dix*
sixty-ten
70 [60 + 10]
b. *soixante-douze*
sixty-twelve
72 [60 + 12]
c. *soixante-dix-sept*
sixty-ten-seven

① Moreover, Welsh does not otherwise make use of "pairing" (see Sect. 6.4).

77 [60 + (10 + 7)]
[French]

Traditionally, such expressions in French are considered vigesimal, on a par with those in (26), although there is nothing vigesimal about the formation *soixante* [swasɑ̃t] 60, which, if anything, is based decimally on *six* [sis] 6; however, they do pose similar processing problems in relating the French expression to its representation in Arabic digits: If one is writing down numbers in Arabic digits from dictation in English, once one hears *sixty*, one can write down 6; in French, by contrast, if one hears *soixante*, one must wait for the continuation before one knows whether to write 6 or 7.

5. Exponentiation and other higher bases

While addition and multiplication allow one to express readily a range of higher numbers, e.g., from 1-99 in a regular decimal system using only ten primitive expressions (and explicit or implicit conventions to express addition and multiplication), this is still quite limited if one wants to express the kinds of higher numbers that are frequently used in the modern world. One therefore needs to create expressions for higher numbers to which multiplication and addition can then be applied.

5.1 Exponentiation

The next – and final, in the case of natural languages – process that is brought into play is exponentiation, i.e., terms for higher powers of the base. The first few terms in this series in English are illustrated in (31).

(31) 10^1 10^2 10^3 10^6
ten *hundred* *thousand* *million*
[English]

Note that, unlike the case with regular expressions using addition and multiplication, the forms in (31) are effectively portmanteau forms, which cannot be decomposed into an element representing 10 and another element representing the power in question; see further (39) below. Exponentiation provides higher multiplicative bases, which can then be multiplied to give products in order to express values between the powers, and to which one can then in turn add in order to give values intermediate between the products of the powers.

Some languages do not use exponentiation, even though they use addition and multiplication, e.g., Chukchi, where the expression in (32) is identified by Skorik as the highest number expressible in the language.

(32) *qliq-qlikkin*
twenty-twenty
400 (20 × 20)

[Chukchi (Skorik, 1961:388, 391)]

Note that this Chukchi expression is a kind of overrunning: In a strict vigesimal system one would expect an exponential expression for 20^2, but in fact multiplication is used.

In English, while there are separate terms for 10^2, 10^3 and 10^6, there are no special terms for 10^4 or 10^5, which requires the use of multiplication (*ten thousand*, *a/one hundred thousand*). Some languages have a new term for each higher power of 10, e.g., Sanskrit as illustrated in (33), on which Whitney comments, "The series of decimal numbers may be carried still further, but there are great differences among the different authorities with regard to their names, and there is more or less discordance even from *ayúta* on."

(33)	10	*dáśa*
	100	*śatá*
	1,000	*sahásra*
	10,000	*ayúta*
	100,000	*lakṣá*
	1,000,000	*prayúta*
	10^7	*kóṭi*
	10^8	*arbudá*
	10^9	*mahārbuda*
	10^{10}	*kharvá*
	10^{11}	*nikharva*

[Sanskrit (Whitney, 1889:177-178)]

Two of these terms have, however, become entrenched in Indian English, namely *lakh* 10^5 and *crore* 10^7. Thus, the number written out in figures that would be divided as in (34a) in most English-speaking countries would be divided as in (34b) in India, i.e., *13 crore, 34 lakh, 35 thousand, 360*.

(34) a. 133,435,360

b. 13,34,35,360

In Chinese, and similarly in Japanese and Korean, though with occasional local variation, there are new terms for each fourth power of 10 starting at 10,000, as in (35); see also (76).

(35)	*wàn*	万	10^4	
	yì	亿	10^8	
	zhào	兆	10^{12}	
	jīng	京	10^{16}	etc.

[Mandarin Chinese/East Asian]

So far we have illustrated exponentiation in decimal systems, but it is perfectly possible and well attested with other bases, e.g., base 20 in Nahuatl (36) and base 6 in Komnzo (37)-(38).

(36)

cem-pōhual-li	one-twenty-ABS	20
cen-tzon-tli	one-four.hundred-ABS	400 (20^2)
cen-xiquipil-li	one-eight.thousand-ABS	8000 (20^3)

[Classical Nahuatl (Andrews, 1975:397-398, 464, 482, 484)]

(37)

6	*nibo*	
36	*fta*	6^2
216	*taruba*	6^3
1296	*damno*	6^4
7776	*wärämäkä*	6^5
46656	*wi*	6^6

[Komnzo (Döhler, 2018:93-94)]

(38)

näbi	*fta*	*a*	*eda*	*nibo*	*a*	*eda*
one	thirtysix	and	two	six	and	two

50 [$(1 \times 6^2) + (2 \times 6) + 2$]

[Komnzo (Döhler, 2018:94)]

Although exponentiation is usually expressed by means of portmanteau morphs, there is some limited productivity in the current international system illustrated here with English. Actually, there are two variants of this system, the so-called short scale (as in English-speaking countries) and the so-called long scale (as in most continental European countries, and until recently in Britain and other Commonwealth countries). The Latin prefixes from *bi-llion* '2-llion' onward indicate the place of each formation in the sequence, as illustrated in (39).

(39)

		long scale	short scale
million	first	10^6	10^6
billion	second	10^{12}	10^9
trillion	third	10^{18}	10^{12}
quadrillion	fourth	10^{24}	10^{15}
(general pattern)	*n*th	10^{6n}	$10^{3(n+1)}$

[English]

Of course, many native speakers of English have problems with the higher items, given unfamiliarity with Latin, and beyond a certain point traditional Latin prefixes peter out, though the series can be continued by using pseudo-Latin formations; for a proposal by L. C. Noll on how to continue indefinitely high, see http://www.isthe.com/chongo/tech/math/number/howhigh.html.

5.2 Other higher bases

In a number of languages, instead of using exponentiation, higher multiplicative bases that are not powers of the lowest multiplicative base are used. A type that is quite widespread in Europe and Western and Central Asia is for the lowest multiplicative base to be 20 (usually with 10 as a purely additive base), used in the formation of numerals 20-99, but with 100, i.e., 10^2, and not a power of 20, as the next highest multiplicative base, as illustrated in the Georgian example (40), with the list of bases using parentheses to indicate those that are purely additive.

(40)	*cxr-as*	*otxm-oc-da-cxra-met'i*
	nine-hundred	four-twenty-and-nine-teen

999 [(9 × 100) + (4 × 20) + (10 + 9)]
Bases: (10,) 20, 100
[Georgian (Hewitt, 1995:54)]

A particularly complex system can be seen in the Supyire Senoufo example (41), where the multiplicative bases are 20, 80 (which is not a power of 20), and 400 (which is 20^2).

(41)	*kàmpwòò*	*ná*	*ŋ̀kwuu*	*sicyɛɛré*	*'ná*	*béé-tàànrè*	*ná*	*kɛ́*	*'ná*
	fourhundred	and	eighty	four	and	twenty-three	and	ten	and

báár-ìcyɛ̀ɛ̀rè
five-four
799 [i.e., 400 + (4 × 80) + (3 × 20) + {10 + (5 + 4)}]
Bases: (5, 10,) 20, 80, 400
[Supyire Senoufo (Carlson, 1994:169)]

Such non-power higher multiplicative bases are nearly always a lower number than the power would be, e.g., in Georgian 100 is lower than $20^2 = 400$, in Supyire Senoufo 80 is likewise lower than $20^2 = 400$ (which is then used separately as an even higher multiplicative base). An exception is provided by a number of medieval Germanic languages, illustrated in (42) by Old Norse, although the "long hundred", i.e., the use of the term *hundred* or its cognate to express 120, is widely attested even in medieval England. Its counterpart is the "long thousand", with a value of 1200. The long hundred is thus [12 × 10], while the long thousand is [10 × long hundred], with a shift between decimal and duodecimal (base 12) systems.

(42)	10	*tíu*		
	30	*þrír tigir*	3 × 10	
	100	*tíu tigir*	10 × 10	
	110	*ellifu tigir*	11 × 10	
	120	*hundrað*		("long hundred")
	240	*tvau hundrað*	2 × 120	
	1200	*þúsund*		("long thousand")

[Old Norse (Gordon, 1957:292-293)]

The Georgian, Supyire Senoufo, and Old Norse examples do illustrate a near-universal property of higher multiplicative bases, namely that they are almost without exception products of the lowest multiplicative base: 100 is a product of 20 in Georgian, 80 is a product of 20 in Supyire, 120 is a product of 10 in Old Norse. An exception is Coahuiltecan, where the lower multiplicative base is 3 (expressed as [2 + 1]), and the higher multiplicative base is 20. Patterns are illustrated in (43), while (44) illustrates the base 3 system.

(43)	12	4 × 3
	13	4 × 3 +1
	14	4 × 3 + 2
	15	5 × 3
	18	6 × 3 (?)
	20	20
	30	20 + 10
	40	2 × 20

(44) *puwāntz'an* *axti-k-pil'* *ko* *pil'*
four two-and-one and one
13 [i.e., {4 × (2 + 1)} + 1]
[Coahuiltecan (Swanton, 1940:48)]

Finally, the Resia Slovenian data in (45) illustrate an extremely rare pattern of alternating bases: The odd tens are expressed decimally, as products of 10, and the even tens vigesimally, as products of 20.

(45)	10	*dẹ́sat*	10
	20	*dwísti*	2 × 10
	30	*trə́sti*	3 × 10
	40	*dwákrat dwísti*	2 × 20
	50	*patardú*	5 × 10
	60	*trị́krat dwísti*	3 × 20

[Resia Slovenian (Steenwijk, 1992:125)]

6. Other arithmetic (and non-arithmetic) processes

While addition, multiplication, and exponentiation (or use of other higher bases) are the usual ways of building up a number system from a small number of primitives to the possibility of expressing large numbers, there are some other arithmetic and even non-arithmetic processes that are found less frequently in the languages of the world.

6.1 Subtraction

Subtraction is found, for instance, in the Latin example (46).

(46) *un-de-viginti*
one-from-twenty
19 [20 – 1]
[Latin]

Compare also the discussion of the Hindi forms with unit 9 in (25).

It is also possible to combine addition and subtraction, sometimes in intricate ways, as in Ket. Example (47) shows a subtractive formation to arrive at a target value of 80 from 100. To this one can then add 2 to reach the target value of 82, as in (48).

(47) *éks bə́nsaŋ kiʔ*
twenty without hundred
80 [100 – 20]

(48) *ɨ́nam ə́kam éks bə́nsaŋ kiʔ*
two left.over twenty without hundred
82 [(100 – 20) + 2]
[Ket (Georg, 2007:179-181)]

In such examples, it is important to respect the language's rules for interpreting such combinations, e.g., in (48) the subtraction precedes the addition, so that the result is [(100 – 20) + 2] = 82, and not [100 – (20 + 2)], which would give 78 – this is not a possible interpretation of the Ket form in (47).

6.2 Division

It is also in principle possible to express a number by means of division, although in practice the expression is always multiplication by a fraction, as in the traditional Welsh form in (49), which is literally 'half of a hundred' rather than 'a hundred divided by two'.

(49) *hanner cant*
half hundred
50 [½ × 100]
[Welsh (King, 1993:113)]

6.3 Overcounting

A particularly intricate way of forming numerals is provided by overcounting, which counts by taking aim at a range that includes but goes higher than the actual target, such as expressing 22 as '2 of the third 10'. The third ten runs from 21 through 30, so '2 of the third 10' is 22. Example (50) from a now obsolete but historically well-attested Tagalog numeral system illustrates the expression of 209 as '9 of the third 100', since the third hundred runs from 201 to 300.

(50) *ma-ika-tló$_n$-'ng daá$_n$-'ng siyám*
PREF-ORD-three-LNK hundred-LNK nine
209 ('9 of the 3rd 100')

[Tagalog (obsolete system) (Potet, 1992:173-174)]

A slightly more complex type is illustrated by the Danish example (51) – note that this form is archaic, the usual modern form being the shortened *halvtreds*. The first two components are more literally 'half of the third', i.e., 2½, which is then multiplied by 20 to give 50. The expression 'half of the third' involves overcounting, since the 'third (unit)' runs from 2 through 3, and half of that third unit implies 2½. See further (79).

(51) *halv-tred-sinds-tyve*
half-third-times-twenty
50 [half of the third, times twenty]
[Danish (Allan et al., 1995:127)]

In the Oriya example (52) the first element *paüne* means '¾ of the *n*th', in this example '¾ of the 3rd 100'.

(52)	*paüne*	*tini*	*šata*
	less.quarter	three	hundred

275 [three quarters of the third hundred]
[Oriya (Odia) (Karpuškin, 1964:38)]

6.4 Pairing

By pairing, we understand numeral formations that are expressed as 'two times *n*'. One might wonder whether this constitutes the use of 2 as a multiplicative base, but this would go both against the structure of the numerals in some languages, and against the usual (but not exceptionless, cf. (41)) principle that in multiplicative expressions the multiplier is not greater than the multiplicand. In Yaqui, the numerals 8 and 10 are expressed as the pairs of 4 and 10, with the prefix *wóh-* related to the numeral *woói* 2, as in (53).

(53)	1	*seénu, wepul*	
	2	*woói*	
	3	*báhi*	
	4	*naíki*	
	5	*mámni*	
	6	*búsani*	
	7	*wóo-búsani* [①]	
	8	*wóh-naíki*	two-four (i.e., 2 × 4)
	9	*bátani*	
	10	*wóh-mámni*	two-five (i.e., 2 × 5)

[Yaqui (Dedrick & Casad, 1999:229)]

① This appears to consist of a different allomorph of the expression for 2 followed by that for 6, but the value is clearly not [2×6]; the first part here may mean something like 'second', i.e., 7 is the 'second 6', the successor of 6.

In Japanese, although there is no explicit segment sequence that can be identified as an exponent of pairing, a vowel shift pattern *i* – *o* > *u* – *a*, with retention of the consonants, establishes three pairs of numerals, as shown in (54) – unless, of course, the pattern is accidental.

(54)	1	*hito*	2	*huta*
	3	*mi*	6	*mu*
	4	*yo*	8	*ya*
	[Japanese]			

6.5 Non-arithmetic structures

One numeral is sometimes constructed on the basis of another numeral, but uses an expression that does not have any specific arithmetic value. Typically, the added element is 'big', as in Sanskrit example (55).

(55) 10^8 *arbudá-*
10^9 *mahārbuda-* (*maha-* 'big')
[Sanskrit (Whitney, 1889:177)]

The English/International term *million* was originally coined in Italian, where *milione* is the augmentative of the word for 1000, as shown in (56).

(56) 10^3 *mille*
10^6 *milione*
[Italian]

7. Ordering of constituents

When the formula in (8) was introduced, it was noted that it should be interpreted irrespective of the linear order of the elements. It is now time to look in somewhat more detail at linear order, in particular the order of the various elements that need to be added together in order to arrive at the target number.

By far the majority pattern of the languages across the world is for these elements to be ordered consistently from higher to lower (descending order), e.g., one starts with the highest power of 10 and ends with the units. This can be seen in the Mandarin Chinese example (57).

(57) *sān-bǎi* *wǔ-shí* *sì*
three-hundred five-ten four
354 [i.e., 300 + 50 + 4]
[Mandarin Chinese]

The opposite order, from the smallest to the largest (ascending order), is also found, though rarely. Example (58) illustrates this possibility using data from Malagasy (Standard Malagasy).

(58)	*efatra*	*amby*	*dima-mpolo*	*sy*	*telo-njato*
	four	plus	five-ten	and	three-hundred

354 [i.e., 4 + 50 + 300]

[Malagasy (Standard) (Rajaonarimanana, 2001:67)]

A variant of these possibilities is where the language in question has a general order, from the largest to the smallest or from the smallest to the largest, but the order of the tens and units is inverted. This is illustrated in (59) for German, where the order is hundreds – units – tens.

(59) *drei-hundert-vier-und-fünf-zig*
three-hundred-four-and-five-ten
354 [i.e., 300 + 4 + 50]

This inversion occurs, incidentally, wherever the combination of tens and units appears, so that in longer numeral expressions like (60), within each set of three elements, the second and third are inverted relative to the general order from the largest to the smallest.

(60) *zwei-hundert-sechs-und-fünf-zig-tausend-drei-hundert-vier-und-sieb-zig*
two-hundred-six-and-five-ten-thousand-three-hundred-four-and-seven-ty
256 374 [i.e., (200 + 6 + 50) × 1000 + (300 + 4 + 70)]
[German]

In fact, if one takes into account the teens, then a number of European languages show some inversion of the tens and units, always for the lower numerals in the sequence. Example (61) illustrates some of the different cut-off points separating the order units – tens (for lower numerals) from tens – units (for higher numerals); all languages have the basic descending order of elements; the source for Modern Greek is Holton et al. (1997:103-104).

(61) 12	*ðó-ðeka*	13	*ðeka-trís*	Modern Greek
	two-ten		ten-three	
15	*quin-ce*	16	*diec-i-séis*	Spanish
	five-ten		ten-and-six	
16	*se-dici*	17	*dici-as-sette*	Italian
	six-ten		ten-and-seven	
19	*nine-teen*	21	*twenty-one*	English
99	*neun-und-neunzig*	101	*hundert-eins*	German
	nine-and-ninety		hundred-one	

Basically ascending order but with inversion of the tens and units is extremely rare, and was indeed hypothesized to be non-existent by Greenberg (1978[1990]:291), but is attested in Sakalava Malagasy, more specifically the dialect of Nosy Be,[①] as in (62).

(62) *limam-polo* *roe* *amby,* *amby* *telon-jato*
five-ten two plus plus three-hundred
'352 [i.e., 50 + 2 + 300]'
[Malagasy (Nosy Be) (Dahl, 1968:14)]

The prevalence of the descending order probably has a functional explanation: It gives an earlier recognition of the approximate quantity involved, e.g., in 354 'three hundred' gives an approximate indication of the target value, while 'four' does not.

8. Ambiguity

It would seem that an obvious property of a numeral expression is that it should be unambiguous, given that the primary purpose of a numeral is to express the exact cardinality of a set. It is, of course, useful to be able to express approximate quantities, but this can be done by adding extra lexical information (e.g., *about* in English) or having specifically approximate number expressions (like *centaine* in French, contrasting with the regular numeral *cent* 100), etc. Nonetheless, one does sometimes find numeral expressions in natural languages that are ambiguous, and this section aims to document the main factors underlying such ambiguity.

8.1 Parsing ambiguities

In English, there are apparently no ambiguous expressions involving the natural number set (1, 2, 3, etc.), but one can construct ambiguous expressions with fractions, as in (63).

(63) *a million and a half (apples)*
(i) 1½ million, i.e., 1,500,000
(ii) 1,000,000 ½
[English]

The two interpretations depend on whether one construes the fraction *a half* with the head noun *apples* (giving 'half an apple', to be added to the million apples), or with the numeral *million* (giving 'half a million' to be added to the already expressed million, i.e., 1,500,000 apples). This is thus a parsing ambiguity: On one parse, the fraction belongs with the head noun, on the other it belongs with the numeral.

Hurford (1975) discusses a number of such examples in Biblical Welsh, i.e., the variety of the traditional Welsh numeral system found in the authoritative sixteenth-century Bible translation. In (64), the 'three twenties', i.e., 60, which appears between the numeral 1000 and the head noun, can be construed either with the head noun 'asses', giving '(1000 + 60) asses', i.e., '1060 asses', or with the numeral 'thousand', giving '(60 + 1) thousands', i.e.,

[①]. I am grateful to Joachim Sabel for information on regional varieties within Malagasy.

'61,000 asses'. With different word orders, only one interpretation is possible, as in (65), where '12 + (3 × 20)' must go with 'thousand', and one can of course also avoid ambiguity by repeating the intended head noun or numeral, as in (66), which is clearly '7000 + 80,000', i.e., '87,000'.

(64)	*un*	*fil*	*a*	*thri*	*ugain*	*o*	*asynod*
	one	thousand	and	three	twenty	of	asses
	(i) 1060 asses (possible interpretation)						
	(ii) 61,000 asses (Numbers 31.34 – intended interpretation)						
(65)	*deuddeg*	*a*	*thri*	*ugain*	*mil*	*o*	*eidionau*
	twelve	and	three	twenty	thousand	of	cattle
	72,000 cattle (Numbers 31.33)						
(66)	*saith*	*mil*	*a*	*phedwar*	*ugain*	*mil*	
	seven	thousand	and	four	twenty	thousand	
	87,000 (I Chronicles 7.5)						
	[Welsh (Biblical) (Hurford, 1975:192, 184, 176)]						

8.2 Abbreviation

Some languages have conventionalized ways of abbreviating numeral expressions, and this can in turn lead to ambiguity. In Thai, for instance, a numeral in the range 1-9 after a numeral in the hundreds or higher can either be interpreted literally as indicating the addition of those two numbers, or with the apparently unit number interpreted as occupying the next slot to the right in Arabic figures. Thus, (67) can mean literally 1002, alternatively, the 2 can be interpreted as filling the next slot to the right, i.e., 1200. Note that the 2 can only be interpreted as occupying either its literal unit slot (interpretation (i)), or the slot immediately to the right of the preceding numeral (interpretation (ii)); an intermediate interpretation such as 1020 (interpretation (iii)) is excluded. One can make interpretation (ii) unequivocal by specifying 'two hundred' rather than just 'two', as in (68), and one can make interpretation (i) explicit by adding an overt linker, which excludes the abbreviated interpretation, as in (69).

(67)	*(nìŋ)*	*phan*	*sɔ̌ɔŋ*	
	one	thousand	two	
	(i) 1002			
	(ii) 1200			
	(iii) *1020			
(68)	*(nìŋ)*	*phan*	*sɔ̌ɔŋ*	*rɔ́ɔy*
	one	thousand	two	hundred
	1200			

(69) *(nɨ̀ŋ)* *phan* *kàp* *sɔ̌ɔŋ*
one thousand with two
1002
[Thai (Smyth, 2002:173)]

Chinese, illustrated here with Mandarin Chinese forms, has a similar possibility for abbreviation, with (70b) being the abbreviated version of (70a).

(70) a. *sān-bǎi* *liù-shí*
three-hundred six-ten
b. *sān-bǎi* *liù*
three-hundred six
360

However, Chinese avoids ambiguity – (70b) can only mean 360 – by having an additional principle: If in Arabic figures there would be one or more zeroes in succession, then this is indicated by inserting *líng* in that position, as in (71), which is thus unequivocally 306. One occurrence of *líng* can correspond to a string of zeroes in Arabic figures, as in (72), and the interpretation of *líng* must always be maximal, i.e., as many zeroes as possible; thus, (72) can only mean 3006, with two zeroes separating the thousand slot from the unit slot, and not 3060; as shown by (73), the latter must be expressed by means of an explicit 60, though *líng* is still needed for the empty hundreds slot.

(71) *sān-bǎi* *líng* *liù*
three-hundred zero six
306

(72) *sān-qiān* *líng* *liù*
three-thousand zero six
3006

(73) *sān-qiān* *líng* *liù-shí*
three-thousand zero six-ten
3060
[Mandarin Chinese]

Chao (1968:575) also notes a then (or earlier) current alternative system in which the number of instances of *líng* must exactly match the number of zeroes in sequence in Arabic digits, i.e., (72) would be *sān-qiān líng líng liù*.

The Thai and especially the Chinese systems point to a cross-linguistically unusual reflection of the position system of Arabic digit representation in natural language.[1] The

[1] Reflections of positional notation do, however, seem to be more frequent in deaf sign languages. See, for instance, the American Sign Language numerals 31-40 as presented by W. G. Vicars at https://www.lifeprint.com/asl101/pages-signs/n/numbers31-40.htm.

more widespread pattern, as in English, is to rely on the individual component expressions to carry the intended meaning, as in *three hundred (and) sixty* versus *three hundred (and) six*.

8.3 Diachronic merger

Occasionally, a diachronic merger can accidentally lead to two numeral expressions becoming homophonous. While this has not quite happened with the English teens and tens, the phonetic difference between, for instance, *sixteen* 16 and *sixty* 60 is perilously small in many varieties of English ([ˌsɪksˈtiːn], [ˈsɪksti]), with the difference in stress tending to be reduced or even lost when the numerals are in attributive position (e.g., *sixteen/sixty books*).

An actual diachronic merger took place in Northern Mansi, with identical phonetic results for both 7 and 100. Differentiation was introduced, presumably as the two started becoming dangerously similar, by adding 'big' to the expression for 100, an instance of a non-arithmetic structure (see Sect. 6.5), as shown in (74).

(74) 7 *sāt*

100 *sāt* > *janiɣ sāt* (*janiɣ* 'big')

[Northern Mansi (Honti, 1993:125)]

8.4 Semantic change

Semantic change can also lead to an intermediate stage where numeral expressions are ambiguous. The shift in British English from the long scale for numerals *billion* and above to the short scale, as shown in (75) (cf. (39)), led to an intermediate period when speakers were unsure of the value of expressions including *billion*, *trillion*, etc., a situation mitigated only by the fact that in ordinary use such high numbers were rarely needed.

(75) *billion*

(i) older, "long scale"	10^{12}
(ii) newer, "short scale" = US English	10^{9}

(long scale: 10^{6n}; short scale: $10^{3(n+1)}$)

[British English]

The East Asian numeral system from *wàn* 10,000 up is attested historically in at least four interpretation variants, as shown in (76). Row (ii) reflects current usage for Chinese, Japanese, and Korean, though with occasional local variation. Rows (i), (iii), and (iv) are presented by Martzloff (1997:99), based on the Chinese mathematical classic *Shushu Jiyi* (数术记遗).

(76)

	万/萬	亿/億	兆	京
	wàn	*yì*	*zhào*	*jīng*
(i)	10^4	10^5	10^6	10^7
(ii)	10^4	10^8	10^{12}	10^{16}

(iii) 10^4 10^8 10^{16} 10^{24}

(iv) 10^4 10^8 10^{16} 10^{32}

[East Asian]

8.5 Specialized use

A given numeral may sometimes have an interpretation different from its regular meaning in particular contexts. As shown in the Yucatec Maya forms in (77), in Mayan the numeral that has the regular interpretation 400 takes on the special interpretation 360 in calendrical accounts, 360 being one of the units in the Mayan "long calendar".

(77) *bak*
usually 400
but 360 days (long calendar)
[Mayan (Tozzer, 1921:97, and more generally 97-103)]

One might compare the use of the international prefix *kilo-* 1000 to mean 1024 (2^{10}) in *kilobyte* (1024 bytes).

8.6 Extended body-part systems

Finally, in extended body-part systems (Sect. 4.2), a given body part numeral without any indication of the side of the body or the number of the pass across the body will be multiply ambiguous, as shown in (78).

(78) *siduŋ* 'shoulder' = 10, 14, 33, 37, 56, 60, etc.
[Kobon (Davies, 1981:206-208, with a minor transcription correction provided by the author)]

9. Internal structure and psychological reality

The question of the psychological processing of numeral expressions by speakers of a language is a huge issue that goes well beyond the scope of this overview article, but nonetheless a few remarks are in order, if only to avoid reading too much into formal structures, especially of low productivity, as inherent features of speakers' internalization of the numeral system as opposed to features of linguists' analyses. In English, for instance, the numeral *sixteen* 16 has a transparent internal structure, in which the unit precedes the ten, and one might expect this to cause problems when an English speaker has to write the number down in Arabic figures from dictation, given that the order is then that the ten precedes the unit. My own intuitive experience is that no such problem arises, which might suggest that *sixteen* has simply been internalized as a whole. By contrast, if someone asks me, as a native speaker of English, to write down in figures the archaic but perfectly familiar *four-and-twenty* 24, then one of three things is likely to happen: (i) I hesitate while thinking, translate the expression into Modern English *twenty-four*, then write 2 followed by 4; (ii) I leave a space, write the 4, and then go back to write the 2, i.e., I break the normal direction of writing to match the order of the parts of the numeral expression; (iii) I write 42, getting it wrong. In this section, we examine a number of numeral expressions where the internal structure proposed by linguists may well not correspond to how native speakers have stored the expression.

As noted in Sect. 6.3, Danish has an unusual way of constructing the tens, more specifically in the range 50-90. In more detail, as set out in (79), the even tens 60 and 80 have a structure clearer in the modern language in the ordinal than in the cardinal numeral, ‘n times 20’, i.e., 60 is ‘3 times 20’ and 80 is ‘4 times 20’. The odd tens use overcounting, with fractions ‘half of the *n*th’ meaning $[(n-1)+½]$, so that 50 is ‘half of the third (i.e., 2½) times twenty’, 70 ‘half of the fourth (i.e., 3½) times twenty’, 90 ‘half of the fifth (i.e., 4½) times twenty’.

(79)

	Cardinal	Ordinal	
10	*ti*	*tiende*	
20	*tyve*	*tyvende*	
30	*tred(i)ve*	*tred(i)vte*	
40	*fyrre*	*fyrretyvende*	
50	*halvtreds*	*halvtredsindstyvende*	cf. *tredje* 3rd
60	*tres*	*tresindstyvende*	cf. *tre* 3
70	*halvfjerds*	*halvfjerdsindstyvende*	cf. *fjerde* 4th
80	*firs*	*firsindstyvende*	cf. *fire* 4
90	*halvfems*	*halvfemsindstyvende*	cf. *femte* 5th
100	*(et) hundred(e)*	*hundrede*	

[Danish (Allan et al., 1995:122)]

However, Danes have explained to me that they do not see the numerals in this way, but simply as irregular non-compositional formations for 50, 60, 70, 80, and 90 in a decimal system; those who know Russian draw the analogy to Russian *sorok* (cf. (21)), a non-compositional rendering of 40 in a decimal system.

Georgian provides a particularly nice example of an instance where a form, that for 1000, is seemingly transparent, but where the apparent internal structure is disregarded in constructing numeral expressions that include 1000; here one actually has empirical evidence that the apparent internal structure is disregarded. Consider the numerals in (80).

(80)

10	*at-i*	
20	*oc-i*	
100	*as-i*	
1000	*at-as-i*	ten-hundred-NOM
2000	*or-i at-as-i*	two-nom-ten-hundred-NOM
	(not **oc-as-i*)	(*twenty-hundred-NOM)

[Georgian (Hewitt, 1995:51-54)]

The numeral 1000 has a transparent internal structure ‘ten-hundred’. However, for the higher thousands, one does not construct expressions of the type ‘twenty-hundred’, but rather of the type ‘two-[ten-hundred]’, i.e., treating ‘ten-hundred’ as a non-compositional whole.

10. Closing remarks and further reading

The default numeral system for a natural language, at least at the present stage of cultural evolution, makes use of a base, typically 10 but with many other possibilities; higher numbers are formed by means of addition and multiplication, sometimes supplemented by exponentiation, applied to the base and, indirectly, its products and powers. Sometimes, higher multiplicative bases that are not powers of the lowest multiplicative base are used, while asymmetric extended body-part numeral systems in Papuan languages do not fit neatly into classification in terms of a base. These arithmetic processes are all “retrospective” in that they build consistently on lower numerals to express higher numerals. Less commonly, “prospective” constructions are found for some numerals, whereby a numeral is expressed in terms of a higher numeral: subtraction, division (actually, multiplication by a fraction), and overcounting. There are also restricted numeral systems, typically going up to the limit of subitizing around 4, often with no internal structure, sometimes making use of base 2 and addition. Numeracy is in many respects a cultural phenomenon, and language contact, increasingly in an era of globalization, has led and is leading to the loss of much of the diversity that is known or can be hypothesized to have existed in the world’s numeral systems.

Chan (2022) provides an online open-access database of numeral systems of the world’s languages – about 4800 in May 2022.

In works dealing with the structure of numeral systems, Greenberg (1978[1990]) postulates a number of constraints, absolute or as tendencies, on natural language numeral systems. Hammarström (2010) examines rare features of numeral systems that test the boundaries of what is possible. Hanke (2005) examines phenomena going beyond the arithmetic properties of numeral systems, including such features as the choice of marker to express addition (‘and’, ‘on’, etc.).

Dehaene (2011) examines numerals (and numbers) from a psychological perspective. Of works adopting an evolutionary perspective, Hurford (1975; 1987) concentrates on the development of formal complexity, while Everett (2017) views numeracy from a cultural vantage point. For language endangerment, reference may be made to Comrie (2005).

Some works dealing with particular parts of the world are also of general interest because of the unusual properties of numeral systems of the area. For New Guinea, Owens & Lean (2018) incorporates the pioneering work on New Guinea numeral systems by the late Glendon Lean, while Saxe (2012) traces the development of the Okaspmin numeral system from contact to the early 21st century. For the Americas, Closs (1986) and Puente (1998) sample from the whole area, while Yasugi (1995:77-105, 241-341) concentrates on Mesoamerica.

Abbreviations

ABS	Absolute	NOM	Nominative
CL	Class Marker	ORD	Ordinal
LNK	Linker	PREF	Prefix

References

Allan, R., P. Holmes & T. Lundskær-Nielsen. 1995. *Danish: A comprehensive grammar* [M]. London: Routledge.

Andrews, J. R. 1975. *Introduction to Classical Nahuatl* [M]. Austin: University of Texas Press.

Avelino, H. 2006. The Typology of Pamean Number Systems and the Limits of Mesoamerica as a Linguistic Area [J]. *Linguistic Typology*, 10:41-60.

Bouquiaux, L. 1970. *La Langue Birom (Nigeria Septentrional): Phonologie, morphologie, syntaxe* [M]. Paris: Les Belles Lettres.

Carlson, R. 1994. *A Grammar of Supyire* [M]. Berlin: Mouton de Gruyter.

Chan, E. (ed.). 2022. *Numeral Systems of the World's Languages* [OL]. https://lingweb.eva.mpg.de/channumerals/Naute.htm.

Chao, Y. R. 1968. *A Grammar of Spoken Chinese* [M]. Berkeley: University of California Press.

Chrisomalis, S. 2010. *Numerical Notation: A comparative history* [M]. Cambridge: Cambridge University Press.

Closs, M. P. (ed.). 1986. *Native American Mathematics* [M]. Austin: University of Texas Press.

Comrie, B. 2005. Endangered Numeral Systems [A]. In J. Wohlgemuth & T. Dirksmeyer (eds.). *Bedrohte Vielfalt: Aspekte des sprach(en)tods* [C]. Berlin: Weißensee Verlag, 203-230.

Comrie, B. 2013. Numeral Bases [OL]. In M. S. Dryer & M. Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/131.

Comrie, B. 2020. Revisiting Greenberg's "Generalizations about Numeral Systems" (1978) [J]. *Journal of Universal Language*, 21(2):3-84.

Dahl, O. C. 1968. *Contes Malgaches en Dialecte Sakalava: Textes, traduction, grammaire et lexique* [M]. Oslo: Universitetsforlaget.

Davies, J. 1981. *Kobon* [M]. Amsterdam: North-Holland Publishing Co.

Dedrick, J. M. & E. H. Casad. 1999. *Sonora Yaqui Language Structures* [M]. Tucson: University of Arizona Press.

Dehaene, S. 2011. *The Number Sense: How the mind creates mathematics* [M]. Rev. edn. Oxford: Oxford University Press.

Derbyshire, D. C. 1979. *Hixkaryana* [M]. Amsterdam: North-Holland Publishing Co.

Dixon, R. M. W. (comp. & ed.). 1991. *Words of Our Country: Stories, place names and vocabulary in Yidiny, the Aboriginal language of the Cairns-Yarrabah region* [M]. St Lucia: University of Queensland Press.

Döhler, C. 2018. *A Grammar of Komnzo* [M]. Berlin: Language Science Press.

Drabbe, P. 1952. *Spraakkunst van het Ekagi, Wisselmeren, Nederlands Nieuw Guinea* [M]. The Hague: Martinus Nijhoff.

Eiseman, F. B. Jr. 1990. *Bali: Sekala and Niskala, Volume II: Essays on society, tradition, and craft* [M]. Singapore: Periplus.

Everett, C. 2017. *Numbers and the Making of Us: Counting and the course of human cultures* [M].

Cambridge, MA: Harvard University Press.
Everett, D. L. 2005. Cultural Constraints on Grammar and Cognition in Pirahã: Another look at the design features of human language [J]. *Current Anthropology*, 46:621-646.
Franklin, K. 2001. Kutubuan (Foe and Fasu) and Proto Engan [A]. In A. Pawley, M. Ross & D. Tryon (eds.). *The Boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* [C]. Canberra: Pacific Linguistics, Research School of Pacific Studies, The Australian National University, 143-154.
Georg, S. 2007. *A Descriptive Grammar of Ket, Part 1: Introduction, phonology and morphology* [M]. Folkestone: Global Oriental.
Gordon, E. V. 1957. *An Introduction to Old Norse* (2nd Edition) [M]. Oxford: Clarendon Press.
Greenberg, J. H. 1978. Generalizations about Numeral Systems [A]. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (eds.). *Universals of Human Language, Volume 3: Word structure* [C]. Stanford: Stanford University Press, 249-295. Reprinted (with corrections) in K. Denning & S. Kemmer (eds.). 1990. *On Language: Selected writings of Joseph H. Greenberg* [C]. Stanford: Stanford University Press, 271-309.
Greenhill, S., R. Blust & R. D. Gray, 2008. The Austronesian Basic Vocabulary Database: From bioinformatics to lexomics [J]. *Evolutionary Bioinformatics*, 4:271-283.
Hammarström, H. 2010. Rarities in numeral systems [A]. In M. Cysouw & J. Wohlgemuth (eds.). *Rethinking Universals: How rarities affect linguistic typology* [C]. Berlin: De Gruyter Mouton, 11-59.
Hammarström, H., R. Forkel, M. Haspelmath & S. Bank. 2022. *Glottolog 4.6* [OL]. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://glottolog.org.
Hanke. T. 2005. *Bildungsweisen von Numeralia: Eine typologische Untersuchung* [M]. Berlin: Weißensee Verlag.
Hewitt, B. G. 1995. *Georgian: A structural reference grammar* [M]. Amsterdam: John Benjamin's Publishing Company.
Holton, D., P. Mackridge & I. Philippaki-Warburton. 1997. *Greek: A comprehensive grammar of the modern language* [M]. London: Routledge.
Holzknecht, S. 1989. *The Markham Languages of Papua New Guinea* [M]. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
Honti, L. 1993. *Die Grundzahlwörter der Uralischen Sprachen* [M]. Budapest: AkadémiaiKiadó.
Hurford, J. R. 1975. *The Linguistic Theory of Numerals* [M]. Cambridge: Cambridge University Press.
Hurford, J. R. 1987. *Language and Number: The emergence of a cognitive system* [M]. Oxford: Blackwell.
Karpuškin, B. M. 1964. *Jazyk Orija* [M]. Moscow: Nauka.
King, G. 1993. *Modern Welsh: A comprehensive grammar* [M]. London: Routledge.
Lean, G. & K. Owens. 2018. Appendix B: Details of counting systems discussed in chapter 8 [A]. In K. Owens, G. Lean, P. Paraide & C. Muke (eds.). *History of Number: Evidence from Papua New Guinea and Oceania* [C]. Cham (Switzerland): Springer, 297-353.
Lojenga, C. K. 1994. *Ngiti: A Central-Sudanic language of Zaire* [M]. Cologne: Rüdiger Koppe.
Martzloff, J. -C. 1997. *A History of Chinese Mathematics* [M]. Berlin: Springer. [Translated from French original, 1988].
McGregor, R. S. 1972. *Outline of Hindi Grammar with Exercises* [M]. Oxford: Clarendon Press.
Merlan, F. 1982. *Mangarayi* [M]. Amsterdam: North-Holland Publishing Co.
Owens, K., G. Lean, P. Paraide & C. Muke (eds.). 2018. *History of Number: Evidence from Papua New Guinea and Oceania* [C]. Cham: Springer.
Pica, P., C. Lemer, V. Izard & S. Dehaene. 2004. Exact and Approximate Arithmetic in an Amazonian Indigene Group [J]. *Science*, 306:499-503.
Polański, K. & J. A. Sehnert. 1967. *Polabian-English Dictionary* [M]. The Hague: Mouton.
Potet, J.-P. G. 1992. Numeral Expressions in Tagalog [J]. *Archipel*, 44:167-181.
Puente, F. B. 1998. *Los Sistemas de Numeración Indoamericanos: Un enfoque areotipológico* [M]. México: Universidad Nacional Autónoma de México.
Rajaonarimanana, N. 2001. *Grammaire Moderne de la Langue Malgache* [M]. Paris: Langues & Mondes – L'Asiathèque.

Saxe, G. B. & I. Esmonde. 2012. *Cultural Development of Mathematical Ideas: Papua New Guinea studies* [M]. New York: Cambridge University Press.

Schapper, A. & H. Hammarström. 2013. Innovative Numerals in Malayo-Polynesian Languages Outside of Oceania [J]. *Oceanic Linguistics*, 52:423-456.

Skorik, P. Ja. 1961. *Grammatika Čukotskogo Jazyka, Čast′ pervaja: Fonetika i morfologija imennyx častej reči* [M]. Moscow-Leningrad: Izd-vo Akademii nauk SSSR.

Smith, G. P. 1988. Morobe Counting Systems [A]. In *Papers in Papua New Guinea linguistics (Pacific Linguistics A-76)* [C]. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University, 26:1-132.

Smyth, D. 2002. *Thai: An essential grammar* [M]. London: Routledge.

Steenwijk, H. 1992. *The Slovene Dialect of Resia: San Giorgio* [M]. Amsterdam: Rodopi.

Swanton, J. R. 1940. *Linguistic Materials from the Tribes of Southern Texas and Northeastern Mexico* [M]. Washington: United States Government Printing Office.

Whitney, W. D. 1889. *Sanskrit Grammar, Including both the Classical Language and the Older Dialects of Veda and Brāhmaṇa* (2nd Edition) [M]. Leipzig: Breitkopf & Härtel.

Tozzer, A. M. 1921. *A Maya Grammar with Bibliography and Appraisement of the Works Noted* [M]. Cambridge, MA: Peabody Museum.

Yasugi, Y. 1995. *Native Middle American Languages: An areal-typological perspective* [M]. Osaka: National Museum of Ethnology.

Appendix: Index of Languages

The classification follows Glottolog 4.6 (Hammarström et al., 2022). The three-letter abbreviations are ISO 639-3 codes; parentheses indicate a partial match.

Language	***Code***	***Classification***	***Location***
Adzera	*adz*	Oceanic, Austronesian	Morobe Pr., Papua New Guinea
Balinese	*ban*	Malayo-Sumbawan, Austronesian	Bali, Indonesia
Berom	*bom*	Benue-Congo, Atlantic-Congo	Plateau State, Nigeria
Chukchi	*ckt*	Chukotko-Kamchatkan	Chukotka, Russia
Classical Nahuatl	*nci*	Uto-Aztecan	Mexico [extinct]
Coahuilteco	*xcw*	isolate	Texas, USA; Mexico [extinct]
Danish	*dan*	Germanic, Indo-European	Denmark
Ekari	*ekg*	Paniai Lakes, Nuclear Trans New Guinea	Papua, Indonesia
English	*eng*	Germanic, Indo-European	England; USA; etc.
Foi (Foe)	*foi*	East Kutubu	S. Highlands Pr., Papua New Guinea
French	*fra*	Italic, Indo-European	France; etc.
Georgian	*kat*	Kartvelian	Rep. of Georgia
German	*deu*	Germanic, Indo-European	Germany; etc.
Haruai	*tmd*	Piawi	Madang Pr., Papua New Guinea
Hindi	*hin*	Indo-Aryan, Indo-European	North-Central India
Hixkaryána	*hix*	Cariban	Amazonas, Brazil
Italian	*ita*	Italic, Indo-European	Italy
Japanese	*jpn*	Japonic	Japan
Ket	*ket*	Yeniseian	W. Siberia, Russia
Kobon	*kpw*	Madang, Nuclear Trans New Guinea	Madang Pr., Papua New Guinea
Komnzo	*(tci)*	Morehead-Maro, Yam	Western Pr., Papua New Guinea
Latin	*lat*	Italic, Indo-European	Rome [extinct]
Malagasy, Nosy Be (variety of Sakalava Malagasy)	*(skg)*	Basap-Greater Barito, Austronesian	Northwest Madagascar
Malagasy, Standard (variety of Plateau Malagasy)	*plt*	Basap-Greater Barito, Austronesian	Madagascar
Mandarin Chinese	*cmn*	Sinitic, Sino-Tibetan	China
Mangarrayi	*mpc*	Mangarrayi-Maran	Northern Terr., Australia
Mayan		[language family]	Mesoamerica

Modern Greek	*ell*	Greek, Indo-European	Greece; Cyprus
Mundurukú	*myu*	Tupian	Amazonas/Mato Grosso/Pará, Brazil
Nauete (Naueti)	*nxa*	Central Malayo-Polynesian, Austronesian	East Timor
Ngiti	*niy*	Central Sudanic	Orientale Pr., DR Congo
Northern Mansi	*(mns)*	Mansic, Uralic	W. Siberia, Russia
Northern Pame	*pmq*	Pamean, Otomanguean	San Luis Potosí, Mexico
Oksapmin	*opm*	Asmat-Awyu-Ok, Nuclear Trans New Guinea	Sandaun Pr., Papua New Guinea
Old Norse	*non*	Germanic, Indo-European	Scandinavia; Iceland [extinct]
Oriya (Odia)	ory	Indo-Aryan, Indo-European	Odisha, India
Pirahã	*myp*	isolate	Amazonas, Brazil
Polabian	*pox*	Slavic, Indo-European	Germany [extinct]
Resia Slovenian	*(slv)*	Slavic, Indo-European	Udine Pr., Italy
Russian	*rus*	Slavic, Indo-European	Russia
Sanskrit	*san*	Indo-Aryan, Indo-European	India [extinct]
Spanish	*spa*	Italic, Indo-European	Spain; Latin America
Supyire Senoufo	*spp*	Senufo, Atlantic-Congo	Mali
Tagalog	*tgl*	Central Philippine, Austronesian	Philippines
Thai	*tha*	Tai-Kadai	Thailand
Welsh	*cym*	Celtic, Indo-European	Wales, United Kingdom
Yaqui	*yaq*	Uto-Aztecan	Mexico
Yidiny	*yii*	Pama-Nyungan	Queensland, Australia
Yucatec Maya	*yua*	Mayan	Yucatán, Mexico