

131. Numeral Bases

Bernard Comrie

1. Introduction

This map is concerned with one aspect of the mathematical structure of linguistic expressions for numerals, namely the arithmetic base that is used in constructing numeral expressions. By the “**base**” of a numeral system we mean the value n such that numeral expressions are constructed according to the pattern $\dots xn + y$, i.e. some numeral x multiplied by the base plus some other numeral. (The order of elements is irrelevant, as are the particular conventions used in individual languages to indicate multiplication and addition.) A simple example is provided by Mandarin, with base 10, in which the numeral 26 is expressed as in (1).

- (1) Mandarin
èr-shí-liù
 two-ten-six

In Mandarin, the convention is that the numeral before the word for 10 is to be multiplied by 10, while that after the word for 10 is to be added to this product ($[2 \times 10] + 6$). Using this concept of base, plus some additional concepts to be introduced below, six main numeral systems can be identified, of which the second and third in the feature value box can be viewed as subtypes of one superordinate type.

@	1. Decimal	125
@	2. Hybrid vigesimal-decimal	22
@	3. Pure vigesimal	20
@	4. Other base	5
@	5. Extended body-part system	4

@ 6. Restricted	20
total	196

As the Mandarin example shows, the crucial concepts needed to demonstrate that a numeral system has a particular base are addition and multiplication. Beyond this, many numeral systems also make use of exponentiation of the base, i.e. expressions to denote the result of raising the base to various powers. Thus, English has a decimal system, and has a special term for 10^2 , namely *hundred*, as well as one for 10^3 , namely *thousand*. While the use of exponentiation often reinforces the identification of the base, it is not taken here as a defining feature, since some languages use addition and multiplication but without making use of exponentiation; an example is Chukchi (Chukotka–Kamchatkan; eastern Siberia), with a vigesimal (base 20) system, but no special expression for 20^2 , i.e. 400, which is simply expressed as 20×20 . Moreover, the linguistic expression of exponentiation is often opaque — there is nothing in the form of the English words *hundred* and *thousand* to indicate that they are, respectively, the second and third powers of the base. Even the limited transparency provided by numerals like English *bi-llion*, *tri-llion*, with Latinate prefixes for 2 and 3 respectively, is only related by a quite complex formula to the corresponding power of the base 10: $10^{3(n+1)}$ in American usage or 10^{6n} in traditional British usage. For further general discussion, see Comrie (1997), Greenberg (1978), and Hurford (1975).

2. The six types

The **decimal** system has already been introduced by means of the Mandarin example (1); the general structure of numerals in a decimal system is $x10 + y$.

In a **pure vigesimal** system, the base is consistently 20, i.e. the general formula for constructing numerals is $x20 + y$. An

example is provided in (2) by Diola–Fogny (Atlantic, Niger–Congo; Senegal), in which the numeral 51 is expressed as ‘two twenties and eleven’.

- (2) Diola–Fogny (Sapir 1965: 84–85)

<i>bukan</i>	<i>ku-gaba</i>	<i>di</i>	<i>uɲɛn</i>	<i>di</i>	<i>b-əkɔn</i>
twenty	CL6–two	and	ten	and	CL9–one

For practical reasons — in particular, the frequency of the type in the world’s languages — it is useful to distinguish a hybrid vigesimal–decimal system in which the numbers up to 99 are expressed vigesimally, but the system then shifts to being decimal for the expression of the hundreds, so that one ends up with expressions of the type $x100 + y20 + z$, this is illustrated in (3) by the Basque expression for 256:

- (3) Basque (Oroitz Jauregi, p.c.)

<i>berr-eun</i>	<i>eta</i>	<i>berr-ogei-ta-hama-sei</i>
two–hundred	and	two–twenty–and–ten–six

Bases other than 10 and 20 are also attested, albeit rarely, among the world’s languages. Ekari (Trans–New Guinea; Papua, Indonesia) makes use of a base of 60, as illustrated in the expression for 71 in (4); the base of 60 was also used in the ancient Near Eastern language Sumerian.

- (4) Ekari (Drabbe 1952: 30)

<i>èna</i>	<i>ma</i>	<i>gàati</i>	<i>dàimita</i>	<i>mutò</i>
one	and	ten	and	sixty

Some languages of the world have numeral systems that do not make use of an arithmetic base. One such system is the **extended body–part system**, here illustrated by a discussion of Kobon (Madang, Trans–New Guinea), which is quite typical of a number of languages of Highland New Guinea. Languages like

Kobon make use of further body parts to extend the system beyond the ten fingers. In Kobon specifically, the names of the following body parts (on the left-hand side of the body) are used in order to count from 1 to 12: little finger, ring finger, middle finger, index finger, thumb, wrist, middle of forearm, inside of elbow, middle of upper arm, shoulder, collarbone, hole above breastbone. The count can then continue down the right-hand side of the body, from the collarbone to the (right) shoulder as 13 to the little finger as 23. It is then possible to reverse the count, starting from the little finger of the right hand as 24 back up to the hole above the breastbone as 35 and down again to the little finger of the left hand as 46. One effect of this is that the names of particular body parts when used as numerals are multiply ambiguous. For instance, *sidurj* 'shoulder' can denote either 10 (on the left-hand side of the first pass across the body), or 14 (on the right-hand side of that pass), or 33 (on the right-hand side of the return pass across the body), or 37 (on the left-hand side of that pass), or 56 on the left-hand side of the next pass across the body, etc. There are usually means, optional or obligatory depending on the language, to distinguish the second side of the body used in a count from the first, as well as to indicate which pass across the body is being used, but there is no productive means to identify other than a small number of passes across the body. Extended body-part systems are thus typically rather limited in the range of numbers that they can express, but can be used productively at least into the scores.

Finally, some languages have **restricted** numeral systems, by which I mean more specifically a numeral system that does not effectively go above around 20. The most restricted numeral system would of course be one lacking any numerals at all, and according to Dan Everett (personal communication) Pirahã (Mura; Brazil) is a language of just this type. A number of languages of the world have numeral systems that extend only as far as 3 (e.g. Mangarrayi (isolate; Northern Territory,

Australia)), while others show slightly higher but nonetheless heavily restricted upper limits, such as 5 (Yidiny (Pama–Nyungan; Queensland, Australia)).

3. Problem cases

In many cases, the assignment of a language to one or another of the types identified is straightforward, but nonetheless a number of problems can arise, and the following paragraphs will note some of these and, where relevant, indicate the solutions that have been adopted in the data analysis underlying this chapter.

First, it is essential to ascertain that the expressions in question are indeed numerals, since in many languages there are quantifying expressions, including some with quite specific denotations, other than numerals, such as *pair* in English (necessarily denoting a set of 2). The general criterion to be used is that for an expression to qualify as a numeral, it must be the usual way of identifying that number of entities in the language in question within a noun phrase. In modern standard English, *seventy* (as in *seventy years*) is thus a numeral, whereas *three score and ten* is not, even if there may have been periods in the history of the English language, and may still be regional varieties, where the latter expression would qualify as a numeral. Note that it is probably not reasonable to require that numerals be used in counting — some cultures with low numeracy do not engage in counting, although they may nonetheless have a non-empty restricted numeral system.

Some languages have two (or more) numeral systems satisfying the criterion of the previous paragraph. Where both are of the same type, as with the indigenous and Sino–Korean systems in Korean — both decimal — then there is no problem in assigning the language unequivocally to one type. Where the systems are of different types, preference has been given to the

most productive, e.g. the extended body-part system in Kobon, which exists alongside a restricted system.

Many languages combine different bases in the construction of their numeral system, and for the purposes of this chapter various decisions have been taken, some principled, some of a more practical nature, to limit the number of distinct types represented on the map. Only one mixed-base type has been given a separate representation, namely the type that is vigesimal in the range up to 99 and then decimal in the expression of the hundreds, because of its frequent occurrence among the languages of the world. In the case of other mixed systems, preference has been given to the base that is most productive in the construction of numerals in the range 20–400. In all numeral systems with a base of 20 or greater, and in several with a smaller base, numbers less than the base are constructed using smaller bases. For instance, in Igbo (Niger-Congo; Nigeria), with a base of 20, the numerals 1–19 are constructed using 10 as the base, as illustrated in (5) for 32:

- (5) Igbo (Green and Igwe 1963: 37)

ohu nà ìrì nà àbụọ
twenty and ten and two

In Supyire (Gur, Niger-Congo; Mali), with a base of 80, the lower scores are expressed vigesimally, while numbers below 20 are expressed using a mixed quinary-decimal system (base 5), as illustrated in (6), with 399 expressed as '80 x 4 + 20 x 3 + 10 + 5 + 4':

- (6) Supyire (Carlson 1994: 169)

ɛ̀kwuu sícyɛɛré 'ná béé-tàànrè
eighty four and twenty-three
ná ké 'ná báári-cyèèrè
and ten and five-four

In some cases, an alternative base is used only in the construction of a small proportion of relevant numerals, and in such cases this alternative is simply disregarded in assigning the numeral system overall to a particular type. In French, for instance, the numerals in the range 80–99 have a vigesimal structure, as in the expression for 97 in (7), but since the system is otherwise entirely or almost entirely decimal, French has been assigned to the decimal type.

- (7) French
quatre-vingt-dix-sept
 four–twenty–ten–seven

While some languages, like Mandarin illustrated in (1), have completely or almost completely transparent structures of the type $xn + y$, other languages have various departures from the ideal formula in their numeral expressions through the appearance of morphophonological idiosyncrasies or portmanteau forms, and it is clearly necessary to be able to distinguish such exceptions from an instantiation of a different numeral system type. Relatively minor morphophonological idiosyncrasies can usually be identified without difficulty, as in relating English *fif-ty* to *five* and *ten*, with the identification of the first morpheme more transparent than that of the second. In some cases, completely different phonological forms may be used in certain combinations, as in the expression of the tens in Spanish, where the suffix *-enta*, as in *och-enta* 80 (eight–ten), bears no formal resemblance to *diez* 10, but is nonetheless reasonably consistent in the expression of the tens. When only a handful of forms are portmanteau in an otherwise transparent system, they can be disregarded, as in the case of Russian monomorphemic *sorok* 40 in comparison with *pjat'-desjat* 50 (five–ten). In some languages most or all of the products of the base are portmanteau forms, but we still identify them as such in that there is a separate such form for each product of the

base and intermediate numbers are expressed by adding to that form; thus, in Turkish (Turkic; Turkey) each of the tens in the range 20–50 is monomorphemic (*yirmi* 20, *otuz* 30, *kırk* 40, *elli* 50; cf. *iki* 2, *üç* 3, *dört* 4, *beş* 5), but numbers between the tens are expressed by addition of the remainder, as in (8), which expresses 21:

(8) Turkish (Kornfilt 1997: 428)

<i>yirmi</i>	<i>bir</i>
twenty	one

4. Geographical distribution

Even a cursory perusal of the accompanying map serves to show that, at least as far as numeral systems are concerned, we live in a decimal world, with the decimal type dominant in nearly every part of the world. Some of the other types are highly restricted geographically, in particular the extended body-part type, found as a basic numeral system in Highland New Guinea, and the restricted system, largely confined to Australia and Amazonia. Bases other than 10 or 20 are extremely rare in the modern world, the examples in our sample being Supyire in West Africa and Ekari in Indonesian Papua.

However, the vigesimal system, whether pure or combined with the decimal system above 100, is still found in a number of different areas in the world, and is particularly frequent in some specific areas, such as Mesoamerica. Mesoamerica is in fact indicative of a worldwide historical trend for the dominant decimal system to encroach on and replace other systems. Pre-Conquest Mesoamerica was largely vigesimal, with the prototypical example being Classical Mayan. The influence of Spanish after the Conquest led to many indigenous languages adopting the Spanish numeral *ciento* 100 and with it the decimal system for the expression of the hundreds. In many languages, replacement by Spanish forms has percolated even further down

the system, and it is not infrequent in contemporary accounts of Mesoamerican languages to read that in practice Spanish numeral expressions are used for all but the lowest numbers. Non-decimal numeral systems are even more endangered than the languages in which they occur.