Cardinal Number Systems

David Stampe
Ohio State University

For my parents, Wilson and Ruth Ellen Stampe

Out of the darkness, Funes' voice went on talking to me. He told me that in 1886 he had invented an original system of numbering and that in a very few days he had gone beyond the twenty-four-thousand mark. He had not written it down, since anything he thought of once would never be lost to him. His first stimulus was, I think, his discomfort at the fact that the famous thirty-three gauchos of Uruguayan history should require two signs and two words, in place of a single word and a single sign. He then applied this absurd principle to the other numbers. In place of seven thousand thirteen, he would say (for example) Máximo Pérez; in place of seven thousand fourteen, The Railroad; other numbers were Luis Melián Lafinur, Olimar, sulphur, the reins, the whale, the gas, the caldron, Napoleon, Agustín de Vedia. In place of five hundred, he would say nine. Each word had a particular sign, a kind of mark; the last in the series were very complicated...I tried to explain to him that this rhapsody of incoherent terms was precisely the opposite of a system of numbers. I told him that saying 365 meant saying three hundreds, six tens, five ones, an analysis which is not found in the "numbers" The Negro Timoteo or meat blanket. Funes did not understand me or refused to understand me.

--Borges, *Funes the Memorious.*

Lacking the perfect memory of Ireneo Funes, ordinary men count on their fingers, or on words that are or must once have been the names of their fingers, and where these end, mathematics begins. The variety of ways that languages combine numbers to form higher numbers is amazing. Here are some expressions for eighteen:[1]

Ono	*mete etke so keio mane so ḍitne karewe*	hand 2 and foot whole and toe 3
Sora (a)	*mɨ-jeŋ yagi*	1-foot 3
Welsh (a)	*tri ar bym-theg*	3 on 5-10
Classical Greek	ὀκτω-καί-δεκα	8-and-10
Spanish	*diez y ocho*	10 and 8
Hottentot	[*dĩsi*]-*xhêisa-ca*	[10]-8-and
German	*acht-zehn*	8-10
Vietnamese	*mười-bảy*	10-8
Lithuanian	*aštuna-lika*	8-left
Ainu	*tu-pesan-ishama-wan*	2-from [10]-and-both [hands]

Latin	*duo-de-vi-ginti*	2-from-2-10
Finnish	*kah-deksan toista*	2-from 10-of the 2nd [10]
Sora (b)	*miggəl-tuḍru*	12-6
Breton	*tri-ouec'h*	3-6
Welsh (b)	*deu-naw*	2-9

The list could go on and on.

Despite the analytic translations of these examples, each of them means 18 and only 18. Funes' complaint about *los treinta-y-tres gauchos* is mistaken; despite its composition, the number designates not thirty and three gauchos but an indivisible thirty-three. The sentences

Treinta-y-tres es treinta y tres (33 = 33 and 3)
Treinta y tres son treinta-y-tres (30 and 3 = 33)

are not tautologous, and the singular versus plural verbs reflect the distinction between the single number 33 and the conjunction 30 and 3. *Treinta-y-tres* means 33 and nothing else. The situation is the same in every language: the single accent of the Greek number ὀκτωκαίδεκα 18, beside the multiple accents of the conjunction ὀκτὼ καὶ δέκα 8 and 10, and the unbroken rhythm of the English number *three hundred (and) sixty-five* 365, beside the caesuraed rhythm of the conjunction *three hundred* [pause] *and sixty-five* 300 and 65, show that numbers are integral expressions in pronunciation as well as in meaning.

Yet Borges was right. The number 365, in Spanish *trescientos sesenta y cinco*, is indeed composed of 3 hundreds, 6 tens, and 5. In every language the number 365,

German	*drei Hundert fünfundsechzig*	((3 x 100) + 5 + (6 x 10))
Welsh	*tri chant pump ar dri ugain*	((3 x 100) + 5 + (3 x 20))
Sora	*miggəl-tuḍru-koṛi mɔnlɔy*	(((12 + 6) x 20) + 5)

and only the number 365, is composed of parts whose values add up to 365. There may be another number used to designate the same quantity, like Sora *yagi-sɔa yagi-koṛi mɔnlɔy* ((3 x 100) + (3 x 20) + 5), and it may even be elliptical, like English *three sixty-five* for *three hundred sixty-five*. But these alternate numbers also have parts whose values, explicit or implicit, amount to 365. Every number in every language, apparently, is equal to the sum of its parts. More: it *is* the sum. In each number the many is one.

This mystery poses a familiar challenge. If the number *treinta y tres* has the same *form* as the conjunction *treinta y tres*, there is nonetheless a difference of *value*. To understand numbers we must understand form *and* value.

The value of each cardinal number corresponds to its order in counting. Each of the numbers for 18 listed above is the 18th number counted in each respective language. They do not mean 18 solely by virtue of the summing of their parts: the sum of the parts of both *ocho y diez* and *diez y ocho* is 18, but only the latter *means* 18, because only it is the 18th number counted in Spanish. Of course, there are numbers too high to count to, whose value is determined from the values of their parts. But ultimately their parts analyze into simple numbers, units like *one*, *five*, *nine*, and the values of these unanalyzable numbers can only be described in terms of their counting-order. Counting is not only the basic use of numbers, it is also the means by which they are learned. Here, as in all life, phylogeny reflects ontogeny, and form reflects function.

Counting and Quantifying

The primary purpose of counting is to compare sums of different sets of things without having to match them side by side. Menninger (1969:34) tells of a man who bought sheep from the Damara at two twists of tobacco per sheep, and the transaction was carried out without counting, literally exchanging two twists for one sheep. Lining up sheep is difficult, and lining up things past, future, or abstract (like days of work) is impossible. Abstract things can be counted with concrete things: the Veddas of Ceylon are said to count with sticks (Menninger 33). In numbers we have abstract sticks. We count with numbers as we measure with yardsticks or trade with money, but numbers are more portable and easier to come by.

Some cultures have no need of yardsticks or money, and some, like the Veddas, apparently have no need of numbers. The Andamanese (according to Bloch in Meillet and Cohen 1952:519), distinguish *ūba-tūl* 1 from *īkpōr* 2 or more, but count no higher. The Australian language Walbiri (according to Ken Hale) distinguishes *t^{y}inta* 1, *t^{y}irama* 2, *maṇkurpa* 3 or more, and although Walbiris combine numbers as in *t^{y}iramakaRit^{y}iramaka-Rit^{y}inta* (2-another-2-another-1), the stringent limit that memory places on such a counting procedure suggests that they didn't have much to count. Of course, when technology or the marketplace demand numbers, every culture can coin or borrow them. Some Australians have devised higher numbers of the one-hand, two-hand, one-foot, whole-man variety; the Walbiri have adopted *wani*, *ṭuwu*, *t^{y}iriyi*, *puwa*, *payipi*, *t^{y}ikit^{y}i*, *t^{y}ipini*, *yayiṭi*, *ṇayini*, *ṭini*. Here, as in many scores of languages in an over-peopled world, the more widely circulated currency devalues, then drives out, the local coin. And nothing is free. While counting men protest being reduced to mere numbers, there are African herdsmen (Marion Johnson tells me) for whom counting cattle is taboo: a good herdsman knows each animal as an individual.

Numbers are used not only for counting things, but also for quantifying the nouns we use in referring to things we count. Ordinarily the numbers used as quantifiers are the same as the numbers used for counting. Occasionally, however, different numbers are used, or differently inflected numbers. For example, in the Munda language Gata?, *mmwiŋ* 1 would be used in counting, but *mwiŋklig* 1 is the corresponding quantifier. But in quantifying people, *mmwiŋja* is used, and in quantifying cattle, *mmwiŋbha?* (Zide 1973). Here there is a further difference depending on what sort of things are being quantified. Often, however, these differences depend on the sort of noun being quantified, rather than directly on the sort of thing the noun refers to. So in Latin, we have *unus puer* '1 boy', *una puella* '1 girl', *unum donum* '1 gift', but *una agricola* '1 farmer', even if the farmer referred to is male, because the noun that the number quantifies is feminine. Distinctions like these rarely go beyond the units, and more often are limited to the first one or two numbers. Most languages lack such distinctions in numbers altogether.

In many languages one cannot refer to countable things without quantifying them. In English we cannot speak of apples without indicating the number of apples we are referring to, at least whether the number is singular, *one (or an) apple,* or plural, *apples*. The plural covers everything from two upward, or in languages that distinguish a dual, everything from three. This should remind us of number systems like Andamanese, which distinguishes one from two or more; or Walbiri, which distinguishes one from two from three or more. The highest number in such systems is approximate rather than exact, recalling *a couple* or *a few*. At eighteen months my son would ask for *two cookie* and protest if he was given only two; as with many children, he used his highest number as a plural. Cardinality is born of plurality.

Singularity, on the other hand, is born of cardinality. The grammatical singular, in the form of the indefinite article, is commonly derived, historically if not synchronically, from the number *one*. English *a(n)* has this source, but the slight accent given elements present merely by grammatical decree has wasted away its form.

Uncountable things, like *water*, require no quantifier, indefinite article, or plural, and in fact admit none: we cannot speak of *a water*, *two waters*, unless we are speaking elliptically of water in countable form, glasses of water or kinds of water. The distinction of countable versus uncountable things becomes a relatively rigid distinction between classes of nouns, traditionally called count nouns versus mass nouns, though (as in the distinction of gender) it may be partially arbitrary: in English *peas* are quantifiable but *corn* is not. Mass nouns can be quantified indirectly, as in *two ears (grains, stalks, acres, bushels) of corn*, or *one loaf (stick, slice, piece) of bread*, by quantifying a count noun which individuates (*ear*) or measures (*bushel*) what one is referring to. In languages that lack the count/mass distinction, all nouns may be treated as if they were mass nouns, so that even in quantifying countable things, the number must be related to the noun by means of an individuating form, as in Thai *bùrì sɔ̌ŋ muan* 'cigarette two long-thing', i.e. 'two cigarettes', where *muan* recalls *stick* in English *two sticks of licorice (firewood, dynamite)* except that in Thai it is used with what in English would be a count noun. The term classifier is applied to forms like *muan* because in effect such forms categorize all the nouns of languages like this according to their shape, animacy, gender, function, and so forth. The meaning of classifiers, as formal rather than fully functional elements, is sometimes generalized to the point of nonexistence, or may be redundant, as in Burmese *ʔein ta-ʔein* 'house one-house', i.e. 'one house', but they remain classifiers due to their association with their respective sets of nouns.[2]

If the form or inflection of numbers, or the classifiers with which they are used, distinguish among classes of nouns, then the nouns and the things to which they refer must be of the same class. This implies, as the saying goes, that we don't add apples and oranges. And in fact it is quite strange to speak of *two apples and oranges*; *two pieces of fruit* sounds much better, although it isn't as informative, but the use of the common term *fruit* avoids the conjunction. Or else we must sum the classes individually, as in *two women and three children*, where we have no word for the five. It is interesting to note, though I think the explanation lies elsewhere, that our reluctance to apply a single number to a conjunction of nouns, as in *two apples and oranges*, is matched by a reluctance to do so to a conjunction of numbers, such as *two hundred-and-one* $2 \times (100 + 1)$, i.e. 202.

Numbers quantify nouns, just as adjectives qualify them, and in the case of the unit numbers, they are not themselves quantified. (Mathematical expressions like *two threes* are a special case.) Higher numbers are quantified, however, as in *three hundred*, and in this regard they resemble nouns. It is perhaps this difference between the adjectival unit numbers and the nominal higher numbers that accounts for constructions like Welsh *un ci ar ddeg* 'one dog on ten', i.e. 'eleven dogs', or older English *four men and twenty*, with the unit number modifying the noun and the higher number conjoined to them, giving a structure resembling that of *green grass and sunshine*. In any event there is a tendency for the units, particularly the lower ones, to be inflected as adjectives, reflecting the gender and case of the noun they modify (*una pulchra puella* 'one pretty girl', *uni mali pueri* 'of one bad boy'). The numbers which can be modified by other numbers, and particularly the higher of these, tend on the other hand to be inflected as nouns, with intrinsic gender and plurality marked (*unum centum* 'one hundred', *tres centi* 'three hundred' parallels *unum donum* 'one gift', *uni* *doni* 'three gifts'), and the noun they quantify may be put in the genitive (*tres centi puerorum* 'three hundreds of boys', matching the rather old-fashioned English expression *three millions of dollars*, and recalling expressions like *three ears of corn*).

But it is not unusual, even in a language where nouns and adjectives are richly inflected, for most numbers to be uninflected. Often only the number one, or one and two, take adjectival infection, and in these instances one suspects that these numbers are playing the role of articles, at least in part; as German shows, article inflection can be very tenacious. As for numbers proper, there are several reasons for them to go uninflected. First, inflections are not as important for identifying which noun a number modifies as they are for adjectives. This is because adjectives are not only attributive (*the tall girl*) but also predicative (*The girl is tall*); predicative adjectives can occur at quite a distance from their head nouns, and inflections help clarify which adjective goes with which noun. Numbers, on the other hand, are ordinarily attributive (*three girls*) rather than predicative (**The girls are three*), and occur in closer proximity to their head nouns, so that inflectional help is rarely necessary to clarify which noun a number modifies. Second, adjectives readily modify conjunctions of nouns, and as the classic example *old men and women* shows, in the absence of inflections it is not clear whether the adjective modifies one or more of the conjoined nouns. Numbers, as we have seen, rarely modify conjunctions of nouns (**two apples and oranges*), and therefore do not require inflections to indicate their scope. Third, the adjective/noun status of numbers is not very clear, and it is therefore unclear what should determine their inflection. This unclarity stems from various properties of numbers: a number may simultaneously modify and be modified (*hundred* in *three hundred thousand*); in phrases with elliptical heads (*the first two*), in partitives (*two of them*), and in distributives (*two by two*), even the normally adjectival units play a nounlike role; and in counting, numbers are removed from their sentential niche altogether. There are probably additional reasons, but I will leave it to the reader to supply him- or herself with his or her own, one of which might begin with the observation that inflection can be a damned nuisance.

The ordering of numbers with regard to the nouns they modify often follows that of other modifiers, either preceding the head noun, as in English *two tall trees*, or following it, as in Zulu *imithi emide amaɓili* 'trees tall two', but there is a strong tendency for numbers to precede even if adjectives follow, as in French *deux arbres hauts* 'two trees tall', and in all languages I know numbers precede classifiers even if they follow the noun (as in the Thai example on the preceding page). These facts were surveyed by Greenberg (1966), and deserve much more study than I have been able to give them. When numbers modify other numbers, as in *three hundred* they follow the head numbers only in languages where adjectives follow their head nouns (Zulu *imithi emide engamashumi amabili* 'trees tall ten two', i.e. 'twenty tall trees'); in the vast majority of languages the modifying number *precedes* the modified one.

This has brought us to the question of the internal structure of complex numbers, the topic of the rest of this paper. Our discussion of counting and quantifying is over, except for a brief return to the topic we started from: the order of cardinal numbers in counting, and the fact that the value of each number corresponds to its place in the counting order.

Order is a universal of counting. The things counted need not be ordered: each counting imposes an order, but the sum is the same regardless of this order. The numbers we count *with* are strictly ordered, however, so that the value of each corresponds to the value of the last plus one. Without this natural ordering, counting would be impossible and numbers without value.

It is this same ordering that permits cardinal numbers to be used as ordinals, to express the order of things. Ordinals are secondary in function in that they require things to be ordered as well as summed, and they are secondary in form: although a few ordinals may have their own form (*first*, *second*), most are derived from cardinals by special morphology (*four-th*) or special syntax (*row three*) or both (*Henry the Eighth*).

First-Order Combination

Somewhere between 2 (the minimal plural) and 20 (the sum of the fingers and toes), and usually at 10, every language runs out of simple numbers expressing consecutive integers (units). To count higher it is necessary to start over again at 1, somehow marking the units of the second cycle to distinguish them from the first. The usual way to do this is by combining the highest unit counted on the first cycle with the successive units of the second.

However, it sometimes happens that the number from which counting starts over at one is not the highest consecutive simple number. In Walbiri, after counting through the simple numbers 1 *t^{y}inta*, 2 *tirama*, 3 *maṇkurpa*, the fourth number is not based on *maṇk^{ur}pa* but *tirama*: *tirama-kaRi-tirama* (literally 2-another-2). This seems inefficient, because by starting with a base 3 the Walbiri could count to 6 (**maṇkurpa-kaRi-maṇkurpa*) with combinations of two units, whereas the base 2 only allows such combinations to reach 4. Probably because *maṇkurpa* can mean '3 or more', as was noted earlier, it is felt to be too approximate to serve as an exact number base. This is supported by the existence of an alternative, *tirama-kaRi-t^{y}inta*, meaning exactly 3.

Similarly, the English numbers are unanalyzable from *one* through *twelve*, but the 13th number is not based on 12 but on 10: *thirteen* combines forms of *three* and *ten*, *fourteen four* and *ten*, and so on. The explanation in this case is historical: *eleven* and *twelve*, like German *elf* and *zwölf*, were originally *ain-lif-an 'one-left' and *twa-lif 'two-left', i.e. left after 10, and thus they were not units. Their original literal meanings are lost, but their original places in the counting order and their numerical values have been preserved. And the counting system remains based on *ten*.

The *base* number of a number system, then, must be defined as that number from which counting starts over. In the vast majority of languages, at least originally, it is the highest of the simple numbers. Taking the highest available number as base, combinations of base plus successive units can yield the highest possible sum. But over the years the parts of composite numbers can be so affected by language change as to become unrecognizable, as in the case of *eleven* and *twelve*, and thus certain composite numbers above the originally highest simple number, the base, may *become* simple.

In the Munda language Sora, whose sisters all count on a 10 base, the numbers *gəlmuy* 11 and *miggəl* 12 were originally compounds with **gəl*(now *gəlji*) 10, perhaps **mi*ʔ*-gəl-muy* (1 - 10 - 1) 11 and **mi*ʔ*-gəl-bar* (1 - 10 - 2) 12 (Zide 1973). Obsolescence of some of the constituent forms (**mi*ʔ survives only in a few compounds, like *misi* '[one] hand') and rhythmic shortenings of these words left them completely opaque. This led to a restructuring of the counting system. *miggəl* 12, now the highest apparently simple unit, replaced *gəlji* 10 as the base for counting: 13 is *miggəl-bɔy* (12 - 1), 14 *miggəl-bagu* (12 - 2), and so on.

This confirms rather dramatically the principle that the highest unit number is the optimal base. But a restructuring like this is unusual if it is not supported by the total structure of the number system. Most Munda languages have decimal-vigesimal counting: they count 10, 20, 20 + 10, 2 x 20, (2 x 20) + 10. Sora changed from a decimal to a duodecimal (12) base within this vigesimal structure. Soras therefore add units to 12 to reach 19 *miggəl-gulji* (12 + 7); then count 20 *bɔ-koṛi* (1 x 20) and add units to reach 32 *bɔ-koṛi-miggəl* ((1 x 20) + 12), to which are added units to reach 39 *bɔ-koṛi-miggəl-gulji* ((1 x 20) + 12 + 7); 40 is *ba-koṛi* (2 x 20), and so on, in a Stravinskian alternation of twelves and eights unparalleled in any known language.[3] 12 as a base is not unheard of--we have the hours, the months, the dozen and the gross--and it corresponds to counting with the thumb the three joints of each of the four fingers, a procedure I have observed in tribal markets in India. But embedded in a vigesimal system, the duodecimal base requires counting in unequal cycles of units. I have heard Sora children count 19 *miggəl-gulji* (12 + 7), 20 **miggəl-tamji* (12 + 8), 21 **miggəl-tinji* (12 + 9), missing the cut-off at 20 because of the accustomed succession of the units *gulji*, *tamji*, *tinji*.

The regular and recurrent cycle of the units is one of the things that makes counting possible. It is in order to achieve this that a certain number (the base) is held constant, while the successive units are combined with it. (Just imagine a system in which, after *ten*, one counted *five-and-six, seven-and-five, four-and-nine,* etc.) Except in Sora, the cycles of units are equal in length, because the base is a factor of the higher base. So in English and German, the base is *ten* rather than *twelve* because it is a factor of *twenty* and ultimately of *hundred*, and thus the full cycle of units recurs in counting.

The base represents an end as well as a beginning, as is clear from Germanic **ain-lif(-an)* and Lithuanian *wienó-lika* 11 (one left [beyond 10]). As such it can be anticipated, as in Finnish *kahdeksan* 8 (2-from 10), *yhdeksan* 9 (1-from 10). Ainu forms all its numbers from 5 to 10 by subtraction (Menninger 69):

1	*shi-ne*	6	*i-wan*	(4-10)
2	*tu*	7	*ar-wan*	(3-10)
3	*re*	8	*tu-pesan*	(2-down)
4	*i-ne*	9	*shine-pesan*	(1-down)
5	*aschik-ne*	10	*wan*	

The rarity of subtractive counting is undoubtedly due to the fact that the succession established in the simple numbers has to be reversed. Although Finnish counts by anticipation, it subtracts only for eights and nines; for the rest it counts upward toward the next 10, as if counting the successive units by decade: 11 *yksi-toista* (1-of the second), 12 *kaksi-toista* (2-of the second), and likewise 21 *yksi-kolmatta* (1-of the third), and so on. (In similar fashion we say that 1976 is in the 20th century.) In this way reversal is avoided except in drawing near the goal, where only a couple of numbers need be reversed, and where the tendency to subtract is strong--compare Latin *duo-de-viginti* 18 (2-from-20), *duo-de-triginta* 28 (2-from-30). (We give the time as *two-fifteen*, *two-thirty*, but as the hour draws near we subtract: *quarter till three*, *five till three*.) Anticipatory counting is not commonplace, perhaps because the laws of chance more often put what we are counting closer to ten than to twenty, and the probability of anticipation becoming the basic pattern of counting is therefore small.

The term base seems particularly appropriate in expressions like Welsh *un ar ddeg* 11 (1 on 10), or the corresponding expressions in Russian, *odin-na-dzatj*, and in Rumanian, *un spre zece*. The preposition expressing the relation between the constituent numbers determines their order.

Order is not determined by conjunction, the most common way of expressing the relation between numbers in first-order combination. Dogs and cats are cats and dogs, and in addition, one and two are two and one. However, in most languages the base number is put before the unit, and in general a higher number before a lower, when they are conjoined. In English we have *sixty-five* (60 + 5), *three hundred sixty-five* (300 + 60 + 5), *nine vigintillion· · nine million nine thousand nine hundred ninety-nine*($(9 \times 10^{?})\cdots+\cdots(9 \times 10^6) + (9 \times 10^3) + (9 \times 10^2) + (9 \times 10^1) + (9 \times 10^0)$), and this order is typical, at least in the higher numbers, of all languages. Exceptions, if they occur at all, come in the lower numbers: the teens in English (*fourteen*), the teens and decades in German (*vierzehn*, *vierundzwanzig* 24). These appear to involve compounding, and I will return to their order in the final section.

Frequently the exceptions to the higher-plus-lower ordering are marked with an explicit conjunction, while the regularly ordered expressions are unmarked: contrast regular *twenty-four* with exceptional *vierundzwanzig* in German or *four-and-twenty* in older English; in classical Greek ten plus unit could be expressed without a conjunction (δεκαοκτω) or backwards (οκτωκαιδεκα), but not both (*οκτωδεκα). Many further kinds of evidence could be cited. The normal order of conjoined numbers is clearly higher plus lower.

I think we can understand this by recalling that number composition involves a constant part and a variable part, so that each successive number in counting differs minimally from the last: *twenty-éight, twenty-níne, thírty, thirty-óne, thirty-twó....* In each successive number we have an old part and a new part, which I have marked in the above examples with an accent, since as it happens the accent falls on the new part. There is a well-known principle of speech, discussed by Mathesius among others, according to which old (known, familiar, already under discussion) material precedes new, other things being equal. Compare, for example:

It's not just cold, it's snowing.
It's snowing, it's not just cold.

What did you get Ed? - *I got Ed a tie.*
- *I got a tie for Ed.*

Why did she buy a car? - *She bought the car for you.*
- *She bought you the car.*

Not just John--John and Martha.
Not just John--Martha and John.

In each pair of sentences, the second, which fails to put old information before new, is at least stylistically inferior to the first. It is possible to salvage some of the examples by accenting the new information, but this actually helps establish the point, since in numbers the new material takes the accent, too. This can be observed by counting by tens from *twenty-eight: twénty-eight, thírty-eight, fórty-eight;* the order of the parts of these numbers is fixed, but the placement of the main accent is not, and it goes on the changing part of each successive number. (The principle of accenting new material will enter the discussion in the final section of the paper.) As for the ordering principle of old before new, it seems to me this is what keeps higher numbers ordered before lower ones in conjunction. In counting, the lower conjuncts, like the rightward digits of an odometer, change more rapidly than the higher ones; the lower numbers are the new material, and therefore they are ordered later--at the climax, as it were, of the whole number.[4]

Second-Order Combination

There seem to be languages in which, having added all the units to the base, one simply stops counting. There are other languages, like Walbiri, cited earlier, or the Australian language Aranda, in which one keeps counting by continuous addition. Thus Aranda counts 1 *nyinta*, 2 *tara*, 3 *tara-ma-nyinta* (*ma* 'and'), 4 *tara-ma-tara*, 5 *tara-ma-tara-ma-nyinta*. (Note that even here the new-material-last principle governs the order of added elements.) Although this system is formally infinite, it is functionally quite severely limited by the brevity of human memory. Finger-counting systems go a little higher, but more important they present the notion of quantifying quantities: two hands are two fives, and (leaving anatomy) three hands are three fives. Here are sums of sums, and multiplication is born.

As the example suggests, the natural, and apparently universal, grammatical expression of multiplication has a variable multiplier (the successive units) quantifying a constant multiplicand (the base): an adjective-noun construction. (This is the source of the adjective/noun dichotomy in number words discussed earlier).

When multiplication is joined with addition, the resultant expressions are, in mathematical terms, non-associative: (2 x 100) + 3 does not equal 2 x (100 + 3). The expression 2 x 100 + 3 is ambiguous, meaning either 203 or 206 depending on how its structure is interpreted. This ambiguity is avoided by a universal restriction on numbers: the head noun (the multiplicand) must be a simple constituent, never a compound one. Thus 2 x 100 + 3 is universally interpreted as (2 x 100) + 3 *two hundred(s) and three*, and never as 2 x (100 + 3), *two hundred-and-threes*.

The universality of this restriction requires some explanation. Adjectives are not in general prohibited from modifying conjunctions of nouns (*brave men and women*), nor are numbers (*ten men and women*), providing the conjoined nouns refer to similar sorts of things. Why are numbers prohibited from modifying conjunctions of numbers? Again, the explanation comes from counting, which, as we noted in explaining why addition proceeds with a constant base, requires minimal variation in successive numbers. If 202 and 204 were expressed as 2 x (100 + 1) and 2 x (100 + 2), 203 would require a totally different expression, because it does not equal any multiple of whole numbers. The structure (2 x 100) + n, on the other hand, permits 202, 203, 204 to be expressed uniformly, with variation confined to the unit expression n. Therefore compound heads in general are avoided in numbers.

This restriction carries over to complex multiplicative expressions like *two hundred thousand*, which intuitively has the structure ((*two hundred*) *thousand*) rather than (*two* (*hundred thousand*)). Either structure would express the same product, 200000, but the structure with a simple head is selected. Our intuition is reflected in our accentuation of the number, which matches that of ((*two dozen*) *apples*) rather than that of (*two* (*golden apples*)). Corresponding expressions are parsed the same way in all the languages I have examined.

We are reluctant to accept *two million million* as a proper number, even if we do not know the alternative expression *two billion* (*two trillion* in America and France), and this reluctance to take a number times itself is universal. This is clearly just a special case of the restriction against compound heads. Since one cannot count to $(n + 1) \times n$ without counting en route the prohibited $n \times n$, this has the further consequence that the quantifier in a number cannot be equal to, *or higher than*, its head. We may have *hundred thousand* but not *thousand hundred* because to arrive at it we would have to pass *hundred hundred*.

There is a corresponding restriction on addition. In systems without multiplication, like Walbiri or Aranda, one is forced to add the base number to itself. But with multiplication, having counted the base plus all the units, the base is multiplied by two. Therefore we have a generalization: in both addition and multiplication the constant number (the augend and the multiplicand) is higher than the variable (the addend and the multiplier). We may have *ninety-nine hundred and ninety-nine* $((99 \times 100) + 99)$, but never *hundred hundred and hundred*, or *hundred-and-one hundred and hundred-and-one*.

Examples like *ninety-nine hundred* $(((9 \times 10) + 9) \times 100)$, or *nine hundred ninety-nine thousand* $(((9 \times 100) + (9 \times 10) + 9) \times 1000)$, show that there is no corresponding restriction against compound, or complex, quantifiers. There are languages that do not use them, however. Sanskrit, a ten-based system, avoided compound quantifiers
3 by providing a name for every consecutive integral power of ten, so that the quantifier never had to transcend nine. This is equivalent to having a name for each digit of a written decimal number, and it may well be that this, together with the discovery of zero, is what enabled the Indians to invent decimal writing. The drawback is that a much larger inventory of head-names is required for counting a given distance if complex quantifiers are avoided, since otherwise the head-names can ascend exponentially. Ancient treatises disagree on the order (value) of the Sanskrit head-names, and modern Indian languages have dropped most of them, and instead form numbers with complex quantifiers. Few languages are consistent in this. English has the bases *ten, hundred* (ten tens), *thousand* (ten hundreds, rather than a hundred hundreds), *million* (a thousand thousands, rather than ten thousand).

To these restrictions on the structure of numbers can be added restrictions on their order. The most basic restriction is that constituents be continuous. Perhaps in Welsh *un ci ar ddeg* 'one dog on ten', meaning eleven dogs, *un ar ddeg* 'eleven' is a single constituent interrupted by *ci* 'dog'. But no number constituent is interrupted by another number. We do not find the likes of *un cant ar ddeg* 'one hundred on ten' meaning hundred and eleven, or eleven hundreds.

As for the order of multiplier and multiplicand, since these are adjective (or adjective-phrase) and noun, we would expect them to follow the ordinary order of other adjective and noun constructions in the language. And they do, as in the *twenty tall trees* examples from English and Zulu, above, but not universally. Adjective-noun languages always have the order multiplier-multiplicand, but so do many noun-adjective languages. The marked preference for this order invites various explanations. As Greenberg (1966) notes, there is a preference for the order quantifier-noun, and in fact a preference for the order modifier-head in general. In compounds--and lower multiplier-multiplicand phrases often are compounded--this preference for modifier-head order seems even stronger; in fact as I write I am unable to think of a single language in which *cheesecake* would be a kind of cheese rather than a kind of cake. The general order preference would explain the order multiplier-multiplicand, but it will still require explanation itself, and that is far beyond the limits of this study.

However, there is an independent motivation for the preference of the order multiplier-multiplicand. We have seen that multipliers are universally lower numbers than multiplicands, and augends are universally lower numbers than addends. We have also seen that, by the old-before-new principle, the higher augend precedes the lower addend: *hundred ninety-nine* (100 x 99). The preferred order multiplier-multiplicand, *ninety-nine hundred* (99 x 100), puts lower before higher. So we have opposite orders: higher + lower, versus lower x higher. This means that in a composite number lacking any overt expression of addition or of multiplication, the order of elements alone tells us what is added and what multiplied. With the other principles described here, the total structure and value of any number is determined without ambiguity.[5]

Simplifications and Complications

As grammarians should know, making things simpler usually makes them more complicated. Our principles run into complications wherever there is ellipsis, since the value of an elliptical number is not the sum of its overt parts. One might consider ellipsis a superficial phenomenon, maintaining that in underlying form numbers obey the rules. But we have seen no evidence for a distinction of levels, and in any event we want to understand how ellipsis works, especially since it works without introducing ambiguity. Ellipsis of head numbers follows the same pattern as ellipsis of head nouns, as in *The men came back and three (men) stayed. Three three,* for example, is sometimes heard for *three thousand three hundred. Three three* cannot mean *thirty-three* or *three hundred thirty* because *thirty* is a single word, and in general the morphemes of single words are not subject to ellipsis, rearrangement, or separation. It cannot mean *three hundred three, three thousand three, three million three,* etc., because in general ellipsis of heads only occurs between contiguous 'digits' of a number: compare also *three (hundred) thirty-three, thirty-three (hundred) thirty, thirty (thousand) three (hundred)* (with a rhythm distinct from the compound *thirty-three*), *thirty-three (thousand) three (hundred)*. The same restriction occurs is Vietnamese and other languages, and I think it reflects a principle of rounding-off: we round numbers by leaving off the lesser details, and a number like *three thousand three hundred* is rounder than *three thousand three*.

What constitutes a round number depends on the number system. Sora *miggəl* 'twelve' is rounder than English *twelve* because it is a base, and could be construed as *miggəl* plus an unspecified unit; Sora *gəlji* 'ten' is not as round as English *ten*. Sora *miggəl-koṛi* 'twelve-twenty' is round because it is a number with few constituents and thus little detail; the corresponding round number in English would be *two hundred* or *three hundred*, since these have fewer constituents than *two hundred forty*.

The number of constituents plays a role in the choice between alternant numbers. 18 in Welsh is more often *deu-naw* (2-9) than *tri ar bym-theg* (3 on 5-10), although the latter follows the regular pattern (compare *deu ar bym-theg* 17), because the former has one major constituent while the latter has two. Normally, as we pointed out above, numbers formed on higher bases are preferred: we say *two thousand*, not *twenty hundred*. But *twenty-four hundred* has one major constituent while *two thousand four hundred* has two, and it is the more frequent variant.[6]

As constituents of numbers, individual numbers (*ninety-nine, hundred*) are unified into phrases (*ninety-nine hundred*); phrases may be unified into compound words (*ninety-nine*), and compounds into simple words (*ninety*). In English unification is signalled by 'musical' means: at each step the duration of the whole is roughly halved, the accent of one part is subordinated to that of the other, the melody becomes more indivisible. In this way *nine, nineteen, ninety-nine* may become metrical equivalents; this can best be sensed by comparing them in context, e.g in *almost* _____ *hundred*.

The decades (*twenty, thirty*, etc.) are deeply unified constituents because they are attributive rather than conjunctive phrases (contrast *thirteen*), because their parts are simple (contrast *thirty-three*), and because they are relatively frequent (contrast (*three hundred*). Unified conjunctives like *thirty-three* do occur, but most deeply in the teens, because here the *ten* constituent does not have a modifier (except in the 'underlying structure' of some generative grammarians).

In many languages compounds have their accent fixed on the first constituent. Compounded numbers subject to this rule include Sanskrit *trayo-daša*, Latin *tre-decim* (whence Spanish *trece*, French *treize*), Irish *tri-deec*, German *drei-zehn* (Yiddish *draitsn*), all meaning 13. We noted earlier that the lower number represents new material in counting, and it should therefore follow the higher number in addition. All these examples are exceptions. But we also noted that new material *must* take the accent. In the languages cited the first element must take the accent. The exceptions are explained: the lower number is put first in these compounds to keep it under the accent.[7]

Notes

[1]I have inserted hyphens within words to separate morphemes. Unless noted otherwise, my examples are from standard grammars, supplemented by Meillet and Cohen 1952. The Sora examples are from my field notes; see also Zide 1973, and footnote 3 below. The literature on number systems is vast, and I have not had access to some of the major compendia. For general bibliography, see the culture-historical survey of Menninger (1969), and Hurford (1975), which includes a critical survey of recent generative work on number systems.

[2]The Thai and Burmese examples are from Greenberg 1972, a study of the grammatical role of classifiers. On their lexical role compare Denny 1976.

[3]Soras do not, whatever the reader may be thinking, have twelve fingers and eight toes; this facile hypothesis is demolished by an alternative counting system, previously unreported, that goes 1 *ʔbɔy*, 2 *bagu*, 3 *yagi*, 4 *unji* (so far the ordinary numbers), 5 *mɨ-si* (one-hand), 6 *mɨ-si-bɔy*, 7 *mɨ-si-bagu*, 10 *bagu-si*, 15 *mɨ-jeŋ* (one-foot), 20 *bɔ-daŋgu* (one-stick), 40 *bagu-daŋgu*, etc. 30 can be *bɔ-daŋgu mɨj-tal* (one-stick one-half). My Sora guide was Monosi Raika of Koraput District, Orissa, India.

[4]However, in the Welsh Bible (analyzed by Hurford 1975) and the Hebrew Old Testament (examples from which were drawn to my attention by Jay Pollack), conjunctive constituents of numbers may be remarkably scrambled. This seems to be a stylization, and probably was not matched in ordinary counting. Numbers in Arabic are written left-to-right against the right-to-left stream of words and I have heard that at one time the digits were read off right-to-left; if so, this has not survived in spoken Arabic.

[5]Numbers with head-multiplier order occur mostly in languages in which modifiers are inflected according to the categories of their head noun, but I do not have enough details on such systems to say whether these inflections play the same structural role as ordering does.

[6]Hurford (1975) notes that this form violates the highest-base condition in his 'Packing Strategy'.

[7]There are lower-higher conjuncts in which the second element is accented, e.g. teen compounds in Persian or English (*thìrtéen*) and decadal compounds in German (*drèi-und-zwánzig*). These are due to accent shifts after the morpheme order became fixed. In Persian the accent of *all* words shifted. The German shift is in progress in

longer compounds, e.g. *àusgezéichnet, Schàffháusen*. In English it is confined to numbers: *thìrtéen, twènty-thrée* receive rising intonation in counting, and I think this has been reinterpreted as rising accent; note the otherwise unexplainable accent retraction of *eléven*, Old English *ánlevan*.

Acknowledgments

Thanks to Ken Hale for Walbiri (1965 was it?), Leena Hazelkorn for Finnish, Norman Zide for a copy of his compendious manuscript on Munda numbers, Arnold Zwicky for loaning me innumerable books for interminable periods, and to all who talked numerology with me.

References

Denny, J. Peter. 1976. What are noun classifiers good for? This volume.

Greenberg, Joseph. 1966. Some universals of grammar. In Greenberg, ed., Universals of language, Cambridge, Mass.: MIT Press.

Greenberg, Joseph. 1972. Numeral classifiers and substantival number. Working Papers on Language Universals 9, Stanford University.

Hurford, James R. 1975. The linguistic study of numerals. Cambridge: Cambridge University Press.

Meillet, Antoine, and Cohen. 1952. Les langues du monde. 2d edition. Paris.

Menninger, Karl. 1969. Number words and number symbols. Cambridge, Mass.: MIT Press.

Zide, Norman H. 1973. Studies in the Munda numerals. Manuscript, University of Chicago.