Sonderdruck aus

Proceedings of the
Fourteenth International Congress of Linguists

Berlin/GDR, August 10—August 15, 1987

Editors:
Werner Bahner
Joachim Schildt
Dieter Vehweger

I

Akademie-Verlag Berlin
Remarks on the Syntax of Numerals

Francisco Marcos-Marin
Madrid

Concepts

Mathematicians and linguists do not give the same answer to the question of what numbers are. The concept of number involves an abstraction and its representation as a two-faced code: the one which represents the "number", i.e. the "cipher", and that representing the "name of the number". The first one is an arithmetical sign, whilst the second is a linguistic sign which follows the corresponding rules of each language. However, the representation of a number by its cipher belongs, in the last analysis, to Linguistics.

The study of numeration systems helps us to understand how a linguistic system works, particularly inasmuch as the numerical system needs very few distinct elements. Account should be taken of both aspects: the names of the numbers and the numeration systems. "Natural" numbers are the "computing numbers"; they match, one by one, the elements of a set: even numbers, divisible by two, odd numbers, not divisible by two. Only these numbers are of interest for the linguist, because the arithmetical operations in which they may be involved are addition, subtraction, and multiplication. Division, in fact, is an arithmetical operation which may imply other kind of numbers: "rational numbers", a subdivision of "real numbers."

Majewicz' (1976) assumption that numerals may be formed also by division is not to be taken in the strict sense: Breton anter kant, 50, literally means 'half of a hundred'. Towka hean unudjeme rukite, 30, 'one man (20) and a half', in Haida 4 comes from 2, but 3, 8 and 10 from 5: 2 is sif, 4 st'a'asun, 8 st'a'asəxə, 3 lq'nul, 6 Lq'A'ul, 5 Lə'ii, 10 Lə'A'. In all these cases "half" is involved, they do not account for a true division.

The names of the numbers are substantives (the four, the five), adjectives (two books, the first chapter, the second call), adverbs (bis) or, in languages like Walapai, even verbs. Just like any other member of the linguistic system, they are pairs of "expression" and "content": their elements are related arbitrarily, i.e. conventionally: the "one member set" is called unus in Latin, one in English or wāḥid in Arabic: these are their names. Nevertheless, there are many traces of a partial motivation. Protoaustronesian (Dahl: 1981) *lima expressed the contents "hand" and "five", a very common association (Majewicz: 1981). Malayward tud'u expresses those of "forefinger", "seven" and "to indicate". This association is also frequent in Bantu languages, as Carl Meinhof (1906) demonstrated in his classic book on the subject.

Numerals also constitute a very special paradigm, where the value of each member is fixed by arithmetical rules defined by the numeration system. Therefore, the role played by secondary motivation is very important: it applies from 10 on, exclusively, in all decimal languages and, in a mixed mood, in the vigesimal ones.

Numbers belong, also, to a representation system where the correspondence between expression and content is bidirectional, though conditioned by the numeration system: unus, one, wāḥid, match the cipher 1. This double attitude invests them with a peculiar importance from the standpoint of linguistic study.

A whole system of numerical representation needs only two keys: the positional principle and the symbol "zero". The positional principle, in our decimal system, expresses unit, decade, hundred, and so on. The "zero" (which may be represented by a blank) was unknown to Romans, it was taken from the Arab mathematician Al-Kharizmi, who took it from the Indians, in the first half of the 9th c. a.D. The oldest European document including Indian ciphers is a Spanish one, the Codex Vigilantium (976 a.D.). Arabic sifr 'void' gave Osp. cifra, with the original meaning of 'zero'. Cifra meant later any representation of a number, its present sense in Spanish, there of English cipher sifr originated as well Latin zephyrum, 12th c., abbreviated ze", origin (Spitzer: 1925) of I. zero, an Italian loan word in English and Spanish.

The new ciphers had many advantages, above all, they were much more precise than heretofore. Thus, they probably favored the use of cardinal numbers in the old territories of the Roman Empire, a tendency observed everywhere, at the expense of distributive and multiples mostly, but of ordinals also.

The study of the development of numeration systems may shed light on the comparative history of numerals. Austronesian languages show an etymological community from one to four and a diversity after this numeral. (Bushmen in the Kalahari desert distinguish one, two, several, many). It is, however, preferable to suppose that they had a base five system, as Pazeh still shows nowadays, up to nine. This explanation seems to be confirmed by a private communication from Blust to O. Chr. Dahl: a base five system is to be found in Ilongot, north of the Philippines, as well as in other islands, even in Melanesia, i.e. Manus in New Guinea.

Modern men say usually that we live in a decimal system, the unmarked one. English-speaking countries and French Canadians moved to it. It is, nevertheless, a partial truth. A sexagesimal system is still in force for hours, minutes, seconds and geometrical degrees. A vigesimal system was apparently widely spread: Northern Europe, Celtic countries, Basque, and French. Let us quote examples from the oldest European language: ogei is Basque for 'twenty', ogei(e) ta hamar (20 and 10) is 'thirty', berrogei (2 x 20) 'forty', irurogei (3 x 20) 'sixty'. A duodecimal system accounts for clock hours, dozens, months of the year, old British shillings and pence, feet and inches. Twenty years ago nobody knew much about binary systems, nowadays we are all familiar with the binary unit, the bit. The same could apply to octal and byte.
Classification

The numerals belong to the morphological classes of substantive, adjective and adverb. Syntactically they are either modifiers or modified, whilst semantically they are members of one of the following sets:

Cardinals: which represent an exact quantity. They are the numbers of natural numbers: Lat. duo.

Ordinals: which represent each element in an ordered series: Lat. secundus.

Distributives: these express a division by so many equal parts as indicated by the number of the radical: Lat. bini.

Multiples: which represent the factor of multiplication of the same quantity: Lat. bis.

Multiplicatives: these are the expression of the cipher by which a set is multiplied: Lat. duplex.

Proportionalities: which express the numerical relation between a referent, taken as a basic unit, and its term of comparison: Lat. duplus.

Fractionaries: which express the quotient of a division: Lat. diminu.

Collectives: these express, in languages like Russian, a group of two or more persons, sometimes of young animals, and are also employed with pluralia tantum.

Morphological procedures and basic numerals

Numerals as free forms are the primary object of consideration here. Their syntactic aspects will be discussed later. The phrase “basic numerals” is taken to mean “cardinals”, even if there is general recognition of the fact that derivation from cardinals does not, on its own, explain the whole of this field.

Inflectional languages, like Arabic or Latin, distinguish between “masculine” and “feminine” forms, as well as “neuter”, if they have one. They also differentiate “singular, plural” and even “dual”: Lat. ambo and duo are remains of the old Indo-European dual grammatical number.

Derivative procedures apply universally. The most common are suffixation, prefixation and (re)duplication. The former is widely exemplified in Latin, the latter are well known in Austronesian languages, which do not neglect suffixation, either. Prefixes such as Indonesian se- express the multiplication by one or the numerator one in a fraction, whereas per- forms the denominator, when it is greater than two (Verhaar: 1984). Ke- changes to ordinals cardinals in postnominal position and to collective cardinals those placed before the substantive. Ber- is an indefinite multiplier of certain cardinals, it can also generate new verbs (so do men- and menper-): berduaan ‘to be together’ (two “dual” persons).

Reduplication of the first consonant + a is very common in Austronesian languages, therefore it must be considered as an old feature. The first syllable can be duplicated in its entirety. O. Chr. Dahl explains in this way the formation of Javanese loro ‘two’ from Protoaustroasiatic *duS:S; a > *ru; u evolves into a semiconsonant, in Old Javanese rwa, later *ro, with monophthong; duplication *ro follows; and finally occurs the well known diminution of vibrations in contigous syllables, producing the modern form loro.

New formations may be affected by this process, see Javanese papa, ‘four’.

Lexico-logical procedures are regularly applied to form other types of numerals. Addition, subtraction, coordination and juxtaposition will be considered under the heading of Syntax. Arabic and Latin certainly have no prefixation like that typically found in Austronesian, with *ka- in the ordinals, han- in the multiples and tsi- in the distributives. Suffixation, instead, is common to many kinds of languages. So in Latin, -tus, -tus, -bmes, -dus are used for ordinals, -plex for multiplicatives, -plus for proportionsal and -erus for distributives. Various suffixes are used in Austronesian, Old Malay offers -i for ordinals borrowed from Sanskrit: pancam-i, saptam-i (‘fifth, seventh’). Indonesian -an means either ‘indefinite multiples of the base’, as in the example: ratu-isan ‘hundreds of’, approximatively: ‘dua ratu-isan ‘about two hundred’ or ‘distributive’, as in the examples: waing ratu-isan ‘money in notes of a hundred’ (ruppes); lusas ‘per dozen’. There exists, also, a suffix of unicity: -nya: ‘the only document’, naskah satusatunya.

Syntax

Few things seem to be so aseptic as numerals. In spite of that, Norwegian authorities (Tauli: 1968, Hagege: 1983) decreed, irrespective of the results, a change of the pattern of the numeration system, from

A) unit + coordinator + decade
   to-og-tese ‘two and twenty’ (22) 
   fem-og-trede ‘five and thirty’ (35) 
   to the pattern

B) decade + unit
   tiise-to ‘twenty-two’ (22)
   trettifem ‘thirty-five’ (35)

These schemes are but two of all the possible combinations (Bonet: 1981), a third possibility should be added:

C) decade + coordinator + unit
   Sp. treinta y cinco ‘thirty and five’ (35)

Apparently there are languages which match all possible patterns. Latin and Arabic seem to be specially rich in this regard. The numerical increment can be achieved by “addition, subtraction” and “multiplication”. Addition, the first pattern, is apparently the most natural. Latin, nevertheless, offers examples of the second. So do Aima urawa ‘3’ (subtracted from 10’ = 10). Yoruba and Romany, quoted by Majewicz (1976, 94). Traces of subtraction can be found in other Indo-European languages, f. i. Russian, where the etymology of (90)
   dziugomesto, provides us with an ending -stot, which is the name of (100). The subtractive solution ‘ten up to hundreds’ is favored by the similarity between (9) dzisatkaj and (10) dzisatkaj and connected with the analogy existing in other Slavic languages, like Polish, where (9)
   dziesieciu took the initial dz from dziesiety. Russian speakers thought that (90) should be a derivative of “nine” when they ceased to understand the original subtraction. Spanish cinco from Latin quinque offers a contamination with the end of the word cuatro, ‘four’.

Addition could be produced either by juxtaposition or by coordination, both possibilities existing in Latin. In both of them either the larger or the smaller number can come first, as Latin also shows. Both patterns are the logical consequence of arithmetics: the combination of decades (multiples of ten) and units cannot yield any multiple of hundred, the arithmetical operations needed are additions and subtractions, exclusively. The corresponding syntactic patterns are coordination and juxtaposition.
Rules: After 20 cardinals with the article are employed, instead of the ordinals, for the decades, although the ordinal is kept in units combining with these tens: 
\[ \text{al-sa'aru-l-tāfā'īn wa-l-ışrān} \]
the 24th hour (lit. the hour the fourth and the twenty)

Syntax helps us to determine whether a number is a cardinal or an ordinal, sometimes in combination with lexicological procedures. Ordinals follow the substantive in Indonesian, \textit{buku keempat} 'book fourth', lit. In Spanish, postposition is compulsory when the cardinal is used in the place of an ordinal: \textit{Alfonso diez, siglo doce} (lit. 'Alphonse ten, century twelve'). It is impossible to say: \textit{diez Alfonso, doce siglo} (lit. 'ten Alphonse', twelve centuries), although the cardinal, when used as such, is placed before the noun: \textit{diez Alfonso, doce libros} (lit. 'ten Alfonso', ten individuals of that name, 'twelve books'): proper ordinals can be placed freely either before or after the substantive: \textit{libro cuarto, cuarto libro} (book fourth, fourth book).

Numerals as loan words

Borrowing and lending are normal in the numeral pattern. In Japanese and Korean there exist, at least, two sets of numerals, the autochthonous and the Chinese set. From 'ten' on Japanese employs the Chinese numerals, which are the only possible in Korean after 'ninety nine'. There are lexical borrowings, even 'syntactic borrowings'. These are clearly exemplified by the Japanese numerals which follow the Chinese pattern, not only as lexical items, but also in both their syntactic construction and the order of their constituents. Another example of syntactic borrowing may be found in the expression of dailytime in Malay (according to the English pattern) and Indonesian (casted in that of English).

Let us now consider the way in which Japanese numerals follow the Chinese pattern. A number such as '2537' will be in Japanese \textit{nissei hyaku sanjiiunnan} that is \((2 \times 1000) + (5 \times 100) + (3 \times 10) + 7\), a perfect reproduction of Chinese \textit{liang qian wu bai shan shi qi}. It may be irrelevant here that the typological class which Japanese belongs to is SOV, because nearly the same structure can be found in another SOV language, namely Latin: \textit{duo milia quingenti triginta et septem}. What makes the structure of these Japanese numerals akin to Chinese and different from purely Japanese phrases is the juxtaposed construction, with total absence of the morphemes which Japanese well-formed phrases always must employ: \textit{kyōju-no ie} ('of a professor a house' = 'a house of a professor'), \textit{watashi-wa atarashi hon-o kaimam} ('I-subject new book-object buy-present', 'I buy a new book'). Morphemes such as \textit{no}, \textit{wa}, \textit{o}, \textit{ma} do not occur at all in those numerical structures borrowed from Chinese. The syntactic pattern of these Japanese numerals is, in consequence, typical of that of an isolating language. It is a borrowed syntactic structure.

We shall now focus on lexical borrowing. The history of languages is generous to scholars who study numerals, providing many examples in the oldest documents of any tongue. We have already seen the Arabic origins of \textit{zero and cipher}, let us consider the oldest Malayan inscription (\textit{Codices. 1930; Jones, 1984}), that of Keduan Bukit, Sumatra, A.D. 682.

\[ \text{avasti śriśakavarṣaśū 604 ekādēśi śuklapakśa vulan vaiśākha dapturayāni niyik} \]
di śāṃnau manapal siddhāvatyā di saṭtāmi śuklapakśa vulan iṣayā paṇḍura hiyam
māramp vairī maṁvam manāva yam vāla duḻakā daṭṭa kon duṣṭhūr cīra di śāṃnau daṭīn āśāvī turāṇus sapula du āśāvāh daṭṭa di mata
sukhacita di paṭcāmi śuklapakśa vula[n àsāḍha] mudita daṭṭam marvult vannu

Hrīvijaya java siddhāvatyā subhiksā . . .
Prosperity! Good Fortune! In the expired year 604 saka, the eleventh day of the bright fortnight of the month Vaisākha (23 April, A.D. 682), our divine Lord embarked to carry out a successful expedition. On the seventh day of the bright fortnight of the month Jyestha (19 May, A.D. 682), our divine Lord left Minānā Tāmvan; he led an army of twenty thousand [men] plus two hundred following by ship, and one thousand three hundred and twelve [men] by land. All arrived at... satisfied at heart. The fifth day of the bright fortnight of the month Āśadhā (16 June, A.D. 682) ... light-hearted, joyous, arrived to make the country... Śrī Vijaya, victorious, successful in his expedition, endowed with plenty...

The connection between Sanskrit and Malay languages is very close. Sanskrit was for a long time the language of religion and science. The aforementioned inscription displays a high number of loan-words. In the Indonesian language low numbers of Sanskrit origin appear mostly in compounds, employed in restricted vocabulary:

- one: eka- (Sk. eka)
  - first: [yang] pertama (Sk. prathama)
- dual: dwi- (Sk. dri)
- triple: tri- (Sk. tri)
- five: pańca- (Sk. pānca)
- seven: sapta- (Sk. sapta)
- eight: asta- (Sk. asta)
- ten: dasa- (Sk. dasa)

There are also some high numbers of Sanskrit origin which suffered semantic changes. I owe to Professor Werner Winter an explanation of this evolution, related to the syntactic structure of the numeration system. Originally, only 10, 10², and 10³ were needed:

- (10) sa-pulu
- (10²) ratus
- (10³) sa-tulu

When the numerical needs of the community grew, 10⁴, 10⁵, and 10⁶ were borrowed from Sanskrit, not with their original meaning, but adapted to the needs of the receiving language:

- (10⁴) ten thousand: lakṣa- (Sk. lakṣa, a hundred thousand)
- (10⁵) a hundred thousand: keśi- (Sk. koti, the highest number, ten millions)
- (10⁶) million: juta- (Sk. ayuta, ten thousand)

This caseness for adopting loan numerals and reforming the system followed an arithmetic basis for their syntax (including semantic incidence) appears in the first attested text, as above. We see forms like ekakā śatimśa, saptamśa, 'seven', or pāncaśamśa, 'fifth', or the second part of the compound word dhusalakṣa, 'twenty thousand' (lit. 'two ten thousand'), all of them Sanskrit loan-words.

The form dhu, 'two' (Dyen: 1947, Blust: 1974) comes from Protoaustronesic *dhu-sa, in spite of its apparent similarity to Sanskrit dvi. Some of the forms mentioned in the inscription are also Austronesian, including ābraha, 'three hundred' (ābra is the Malayan form of Protoaustronesian *brahā and ratus is the common Austronesian word for 'hundred'), sapuļu, 'ten', and sariu, 'thousand'. These last two show the prefix sa-, which has the sense of 'one unit of' or a 'multiplier by the unit'.

The adaptation of loan-words became complete, giving way to morphemic reinterpretations of some of them. This happened with Indonesian pertama, 'first', analyzed as per + tama, which generated an adverb with duplication of the second element: Indonesian pertama-tama, 'in the first place':

### Conclusions

Cardinals are the basic numerals, but it does not mean that all the other forms are derived from them.

The cardinal system consists of:

1. a basic lexicon composed from:
   a) names of numbers up to a low cipher: (10) in Arabic and Latin, (12) in English and German, (15) in Modern Spanish, (16) in Old Spanish and Modern French.
   b) names of the tens
   c) names of exponents of ten (hundred, thousand, million, etc.)

2. a limited number of syntactic rules.

These syntactic rules match the arithmetical operations of addition and subtraction if and only if either units alone or units and decimals are concerned. When the combination affects units and tens with other exponents of ten, the corresponding arithmetical operation is multiplication.

The syntactic rules of the cardinal numbers are determined by the arithmetical of addition, subtraction and multiplication.

Division cannot be included at the same level, because it yields either natural or rational numbers. Some languages have specific numerals representing these results: partitives, distributives, fractionaries, but they seem to be a minority. Suffixes, like partitive -ahu in Spanish, are also employed. Nevertheless, a solution which implies the combination of a cardinal and a lexical morpheme indicating the division is widely used. Spanish offers this possibility, together with the previous two: dos mitades, cuatro porciones, cinco partes.

A standard of grammaticalization is not to be found.

Numerals are the only elements of a language which show an undeniable interconnection of economic and linguistic resources. The connection of arithmetical and syntactic rules gives a nearer profile to these specific parts of the lexicon.

### References

Barthe, Renate. The Semantics and Syntax of Number and Numbers, (from MLA data bank Festschrift article in AN 73-7-74), 51–93.


Word-Order Syntax (W-Syntax) as an Autonomous System