

The semantics of numerals*

NILS M. HOLMER

Just as arithmetic among the arts and mathematics among the sciences are the most abstract, so the designations of number used in the various languages are those which least of all words show traces of having at one time had a concrete meaning. At the same time the art of computation and the science of quantification are least evolved among-or one might say least appealing to-those «primitive» peoples whose unartificial system of numbers does not extend beyond the first digits of our decimal system or even beyond the first five numbers in a quinary system or is of a still more rudimentary type (as among some native tribes in South America or among most of the Australian aborigines). On the other hand, the ability to count is comparatively higher among peoples who employ a decimal system (Polynesians, ancient Peruvians, not to speak of those nations whose civilization is a heritage from the peoples of Mesopotamia and Egypt: Hebrews, Arabs and, finally, Indo-Europeans) or a vigesimal system-in reality an evolution of the decimal system, although generally more common in earlier civilizations (Mayans, Mexicans, etc.). Remnants of a vigesimal system occur, of course, in Europe as well.

However, Babylonian numbers, Greek arithmetic and Arabian Algebra, as well as those later specialized branches of mathematics which constitute the dominant and characteristic element in our Occidental civilization, are all based on and derived from the same crude systems of 'one-two-(three)-many' which we have mentioned as typical of the languages of Australian tribes and which are also reflected in the triple or quadruple system of grammatical number (singular-dual-trial-plural) characteristic of the pronominal inflection in some Oceanic languages. And vestiges of this primitive state of things may be discerned in the names of some of the numerals employed in more advanced civilizations still today.

A notably clear example of this is the Austronesian numeral 'five', which in most of these languages is identical with the word for 'hand', corresponding to the number of the five fingers (Malay *lima* 'hand' and 'five', etc.); similarly derived are the Greenlandic *tatdlimat* 'five' (from *taleq*, plural *tatdlit* 'arm') and the Nahuatl *macuilli* 'five' ('what is taken in the hand'). Some Indo-European scholars have been inclined to see a relation between the Indo-European word for 'five' (Latin *quinque*, Greek *πέντε*, Sanskrit *pañca*, Gothic *fimf*) and the Germanic word for 'finger' (Gothic *figgrs*)¹, but this connection is less evident although by no means

* Särtryck ur A R S B O K 1963/1964 utgiven av seminarierna för slaviska språk, jämförande språkforskning och finsk-ugriska språk vid Lunds Universitet.

1. This connection was recognized by Franz Bopp and also appears in Lorenz

*impossible. In the same way the Basque word for 'five' (bost) was formerly combined with a Celtic word meaning 'palm of the hand' (Old Irish boss, whence it would have been borrowed), an etymology which is rendered questionable by the fact that the original form of the Basque numeral seems to have been bortz (still used in Navarre; cf. The Aquitanian form Borsei, a proper name perhaps analogous to such Latin names as Quintus, Pompeius, etc.)². In the vigesimal systems the numeral 'twenty' often means 'man', because when 'twenty' has been reached (in counting on fingers and toes) all the fingers and toes of hands and feet have been counted or, in other words, the 'whole man'. In this way we find the Eskimo (West Greenland dialect) *inuk nâvdlugo* (= 'reaching a man'), the Quiche *huvinak* 'one man' or the Cuna *tulakwen* (idem) used for the numeral 'twenty'³.*

As a matter of fact, in counting units it is customary to use basic nouns which may turn into regular numerals: in Irish *ceann* 'head' is equivalent to 'one' in counting most objects; *súil* 'eye' is used in counting fish. In Irish *beirt* 'a couple' is used for 'two' (especially of persons) and in Scottish Gaelic *paidhir* (from English 'pair'), for 'two (things)'. In English a 'score' is still 'twenty' and this word is (like the French *vingt*) sometimes used as a basis for higher numerals ('fourscore' for 'eighty', etc.). It is hence proved beyond doubt that concrete nouns may be used in the sense of proper numerals.

Before we enter upon our main subject—a semantic analysis of the terms expressing number—it is necessary to say a few words regarding the general character of this particular word class. Numerals are more often than other words «borrowed» or, to use more appropriate terms, they belong to the type of words which show a particularly wide diffusion, which is evidently connected with their being a fundamental element in commerce. It may be noticed in many parts of the world, where commerce is an important cultural factor, that the numerals are those parts of speech which show the closest agreement in otherwise unrelated, or not very closely related, languages. It has been said that the numerals, like the pronouns and names of parts of the body, are among those elements which change the least during the evolution of language and that identical numerals are therefore to be considered as sure criteria of the so-called «genetic» relationship. This is most certainly a wrong point of view: the agreement depends on mutual influence and while any class of words may spread or be «borrowed» over a large area, this holds more particularly for the numerals, at least in advanced civilizations or in civilizations in which trade is important. So, for instance, do we find numerals which are fundamentally the same in Tibetan, Chinese, Thai and Vietnamese, whose «borrowed» character is in most cases manifest. The same holds for the Austronesian languages: whereas other words, including many pronouns, may show great divergences, the numerals are remarkably convergent over an area extending from Madagascar to the Easter Island. We find the situation similar in the Indo-European, Semitic

Diefenbach's *Vergleichendes Wörterbuch der gothischen Sprache* (Frankfort on the Main, 1851), p.374. It is mentioned by Brugmann (*Grundriss*, vol. II, p. 4), but is generally not taken very seriously today (cf. J. POKORNY, *Indogermanisches Etymologisches Wörterbuch*, 2d ed., p. 808).

2. See Luis MICHELENA, *Fonética histórica vasca* (San Sebastian, 1961), p. 363.

3. Cf. Caroline T. STEWART, *The Origin of the Names of Numbers* (in *Beiträge zur Kunde der Indogermanischen Sprachen*, vol. 30. Göttingen, 1906), p. 239, Note 1.

and, to some extent, in the Finno-Ugric languages, that is in general in those languages which have the most evolved numeric systems and where these most obviously have served commercial purposes.

In connection with this it may further be found that the formal agreement between numerals, as they appear in various culture areas, is vague and merely suggestive rather than phonetically exact. Serving primarily the practical purpose of facilitating commerce, forms derived from related as well as unrelated languages may have been adopted and in the latter case the difference existing in grammatical or phonetic structure may have proved inconvenient. It is, for instance, rather clear that a form which is essentially the same has been used for 'seven' in a great many languages in the Mediterranean area, as appearing from their general similarity (cf. Akkadian *sibi*, *sibâ*, Hebrew *ševa'*, Arabic *sab'(un)*, Old Egyptian *sfb*, Coptic *sašf*, Indo-European **septm* and **sepm* and even Basque *zazpi*, Georgian *švidi*, Mingrelian *sk'wit'i*)⁴. Probably one or the other of the Etruscan numerals *semφs* and *cezp-* (in *cezpalχ*) corresponds to 'seven'⁵. Even in the American Indian languages this can be observed: cf. Quechua *pisqa*, Aymara *pheska* ~ *piška*, Chibcha *hysca* or *hycsa*, Cherokee *hiski*, Iroquois *wisk* (Seneca *wis*) 'five' (in the latter languages the phoneme *p* is missing). As usually happens in cases like these, assimilation to the native phonetic pattern or even to the vocabulary itself has taken place: the Hebrew word for 'seven' has the same consonants as the verb 'to swear', the Old Egyptian word, the same consonants as the verb 'to loosen, undress' (whereas the Arabic *sabba'a* only means 'to make sevenfold'; 'to swear' in the sense of 'abuse, revile' is *sabba*) and the Arabic *ḥamân(in)* 'eight' (cf. Old Egyptian *ḥmn*, Coptic *šmoun* idem) has the consonants of *ḥamma* 'to value' or *ḥaman(un)* 'price', etc.

Even within the Indo-European languages the forms of the numerals do not correspond to one another phonetically: compare the Latin *septem*, Greek *ἑπτὰ* and the English 'seven' (Gothic *sibun*), in which latter the *-t-* of the former is missing (possibly indicating that the stem is originally not **sept-* but **sep-*; cf. the Semitic forms and further below). More particularly we notice that the numerals have a tendency to attract one another in respect of phonetic structure, forming pairs which have either the initial sound or some other phonetic characteristic in common: Latin *quattuor* ~ *quinque*, English 'four' ~ «five», Welsh *pedwar* ~ *pump* (in spite of Sanskrit *catvârah*, *pañca* 'four, five', respectively), Russian *d'evjat'* ~ *d'esjat'*, Lithuanian *devyni* ~ *dešimt* (in spite of Latin *novem*, *decem*), Tokharian (dialect B) *sukt* ~ *okt* ('seven, eight'; cf. Latin *septem octo*), Hungarian *egy* ~ *két* ('one, two'), Lapp *aktse* ~ *kaktse* ('eight, nine', in spite of the Finnish *yhde-*, *kahde-*; that is to say that Hungarian and Lapp have formed their

4. The Mingrelian *-k'* is secondary (as in *č'k'imi* = Georgian *č'emi* 'my', *sk'ani* = Georgian *šeni* 'thy'). The latter forms seem to represent an earlier stage in the evolution of the cognate Armenian *evt'n* 'seven', that is a form in which the initial **s-* had not yet passed into **h-* (and zero; hence **ševt'-?*), in the same way as Georgian *γvino*, Mingrelian *gvini* represent an earlier stage in the evolution of Armenian *gini* 'wine' (in which the original Indo-European **w-* had passed as far as **yw-*).

5. According to Vladimir GEORGIEV, *Hethitisch und Etruskisch* (Académie bulgare des sciences. Sofia, 1962) p. 47, *semφs* is the Etruscan word for 'seven'.

6. Cf. Caroline T. STEWART, *op. cit.*, pp. 229-30.

words for 'two' (or 'eight') from a stem **kekete-* (instead of **kakete-*), as it appears under the influence of **ekete-* 'one'), Southern Lapp *ukkie* ~ *lukkje* ('nine, ten'), *åkkede* ~ *låkkede* ('ninth, tenth'), in which the vocalism of the word for 'nine' appears influenced by that of the word for 'ten'; also notice the Semitic forms *s-b-* ('seven') and *t-s-* ('nine'), which reveal a certain similar structure.

It is probable that numerals have in their origin not been very different from other words, serving the mere needs of ordinary conversation. This is the more likely when we think of (1) the fact that the lower numerals are closely linked up with the system of grammatical number (singular-dual-trial-plural)⁷ and (2) the circumstance that numerals in the beginning evidently lacked the preciseness of designation which they have later acquired (cf. further below)⁸. In primitive systems the highest number, whether 'three', 'four' or 'five', is often equivalent to 'many'⁹. When the necessities of trade and commerce so demanded, as well as further on that of science, a more precise fixation of the value of the various numerals became imperative. Examples of cases in which traces of an earlier vagueness of meaning survives are by no means exceptional. In addition to those given by the author in a previous paper¹⁰, we may add the following ones: in Maori *tekau* is now 'ten' (probably owing to influence of the decimal system), while formerly it meant 'twenty' (cf. *rua tekau* 'forty', *toru tekau* 'sixty', *wha tekau* 'eighty') as in some other Polynesian languages (e. g. Tongan); in Hawaiian, on the other hand, *ka'au* (originally the same word) means 'forty'. The word for 'two' (*niso*) in Cree (an Algonquian language) resembles the word for 'four' (*niso*) in the related Blackfoot and the word for 'three' (*nisto*) in Cree is like the one for 'five' (*nisito*) in Blackfoot. A 'hundred' means 'ten times ten' in Modern English, but in Old Norse the corresponding word (*hundraþ*) meant 'one hundred and twenty' (a «long hundred», just as still today one speaks of 'long and short tons'). In Lapp *låken* (*lukkje*, etc.) means 'ten', but in some dialects *stuore lukkie* (a 'big ten') is equivalent to a 'hundred'. Even today the word 'billion' has a different meaning in different parts of the world, being a 'thousand millions' in France and America, but a 'million millions' in England. Such being the facts, we should not go far wrong in making allowance for this vagueness of meaning even in the field of reconstruction: it used to be supposed, for instance, that the Etruscan numeral *huθ* would answer to 'four' on account of

7. Grammatical number does not correspond with arithmetical number. The dual, for instance, rather expresses 'two together' than the number 'two'; hence the close formal relation to the *comitative* case forms in some languages, for which see the author's *How Does Language Change*, (Annual Report 1961-62, published by the Linguistic Seminars. Lund, 1965), pp. 79-80.

8. Vagueness of meaning is a characteristic of such common words as do not pertain to the *terminology* of any particular science. With the original indefiniteness of meaning in numerals, which has resulted in a vague or different numeric value in different languages, compare for instance the different meaning of originally identical common words, such as Latin *hostis*, *hortus*, *mare*, on the one hand, and the corresponding English forms 'guest' (Anglo-Saxon *giest*), 'yard', 'mere' (= a 'lake'), on the other.

9. Cf. in the cuneiform writing the use of the symbol <<< (related to <<< or ||| 'three') to design plurality.

10. See the author's *Comparativ Semantics: A New Aspect of Linguistics* (International Anthropological and Linguistic Review, vol. 1: 1. Miami, 1953), p. 98.

the fact that the ancient («Pelagian») name of a Greek city Τετράπολις was ἑπτηνία, a conjecture which certainly does not lack ingenuity; if we accept it, there would at least be no semantic objections to a comparison with a form in another language of the «Nostratic» group, viz. Georgian, in which *xut'i* means 'five'. If Hermann Möller compared the Semitic stem *ʿasar-* 'ten' with the Indo-European stem **ok-* (in Latin *octo*, Sanskrit *aṣīti-* 'eighty'), this is unobjectionable from the semantic point of view, although it might be questioned on phonetic grounds. One might even feel justified in comparing the Old Egyptian *hmt* (Coptic *šomt*, *šomnt*) 'three' with the Semitic *ḥ-m-s* 'five' (Arabic *khams(un)*, Hebrew *ḥāmēš*), supposing both forms represent an extension of a more primitive stem **ḥ-m-*.

The purpose of this paper is, however, not to attempt to trace the original concrete meanings of numerals, but to establish semantic connections between words expressing number, either similar or identical in form in the various languages. The following survey cannot, of course, claim to be complete, due both to the lack of space and material and further not least owing to the lack of definite information regarding early associations of numerals with common nouns. The former have ever been mysterious; their rôle in mythology, superstition and magic is well known. Owing to this perhaps, as well as to their having been early adapted into a system of counting and to having frequently passed from one language into another and hence not seldom suffered transformation (cf. above), whatever connection they may have had a long time with the current vocabulary of a language has become dimmed and in the course of time utterly wiped out.

To this process has contributed the not infrequent combination with irrelevant elements (relation words), such as the initial *n-* in a great many Amerindian numerals (Blackfoot *natoka* 'two', probably from *at-* 'again', Dakota *nōpa*, Osage *noⁿba* or *thoⁿba*, Ponca *ēaⁿba*, all meaning 'two'; cf. Nahuatl *ome*, idem), which has sometimes been supposed to be an element meaning 'hand' (in reality not much more than a guess); among such elements we could probably range the Indo-European **kⁿe* (in Latin *quinque*, etc., as against Tokharian (dialect A) *pān*, as if from **pene*; further *quattuor*, etc., as against the Sanskrit *turiya*, *turya* 'fourth' or a Hittite *duyanalli*)¹¹ and **d(e)-* (in Latin *duo*, etc., as against Tokharian *wu*, *we*, *wi* 'two', *decem*, etc., as against *viginti*, *triginta*, etc.)¹² for which see further in dealing with the separate numerals.

We shall now-by applying the method of comparing homonymous morphemes in various languages, as suggested by the present author in earlier papers-make an arrangement of some forms of the basic numerals in the decimal system (those from 'one' to 'ten'), with the Indo-European forms as a starting point¹³. These we shall deal with in order, beginning with the numeral 'one'.

'One'. -In Indo-European the following stems are basic: **se-* (**sem-*,

11. For this latter, cf. Vladimir GEORGIEV, *op. cit.*, p. 51.

12. Cf. the author's, *Note on Slavic sto 'hundred'* (Annual Report 1948-49, by the Slavic Institute at Lund), p. 150.

13. Cf. the author's *Comparative Semantics: A New Aspect of Linguistics* (see Note 10, above), pp. 97-106, and *Semántica y etimología* (Boletín de la Real Sociedad Vascongada de los Amigos del País, vol. XII, Part 4. San Sebastian, 1956), pp. 385-95.

**sm-*), which probably has a pronominal origin (= 'the same?') or is originally connected with pronominal forms (cf. Greek $\acute{\omicron}$, η , Sanskrit and Gothic *sa* 'he', Greek $\acute{\omicron}\mu\omicron\varsigma$, English 'same', etc.), and **oi-* (to be discussed below)¹⁴. Of the former remnants only exist: Sanskrit *sahasra* '(one) thousand', Latin *semel*, *simplex* (the latter equaling English 'onefold', etc.), except in Greek (where it appears in $\epsilon\acute{\iota}\varsigma$, $\mu\acute{\iota}\alpha$, $\acute{\epsilon}\nu$, $\mu\acute{\omicron}\nu\omicron\varsigma$, etc.) and in Tokharian (where the dialect A has the forms *sas*, *som*, etc., and the dialect B, the forms *še*, *šeme*, etc. 'one'). The meaning is evidently 'the same', which explains both the sense of 'one' and those of 'he, she, it, that', etc. (cf. Greek $\acute{\omicron}$, η ; Celtic *sin*, as in Gaulish sootv 'that'; Old Norse *sem* 'that, which' (relative), originally 'that same'). The same element no doubt recurs in the reflexive pronouns: Latin *se*, Gothic *sik* ('himself, herself'), etc.; an alternative form in **sw-* occurs in Latin *suus*, Greek $\acute{\epsilon}\alpha\theta\tau\acute{\omicron}\varsigma$, Sanskrit *svayam* 'self', etc. (the meaning 'self' representing an evolution of the reflexive sense, as also in Russian *sam* 'himself', from the same stem as English 'same').

A corresponding semantic evolution in Basque is interesting. In this language the stem *be-* covers many of the senses of the Indo-European **s(e)*, **sw-* (cf. above): the reflexive pronoun is *bera* (which is often used for 'he, she'), with the possessive *bere* 'his, her (own)' (the latter is to the stem *be-* as *gure* 'our' to *gu* 'we') and the numeral 'one' is *bat* (in certain cases and dialectally *bet*)¹⁵. *It is further tempting to see the same stem in the adverb beti* 'always', which would then be a formation analogous to the Latin *semper* (*be(t)-* corresponding to *sem-* and *-ti*, a variant of the ablative suffix *-tik*, to *-per*; hence literally 'through one'); the Basque words *bakan* 'simple' and *bakar* 'alone' are, of course, derivations of *bat* 'one'¹⁶.

Now to turn to the second stem occurring in forms of the numeral 'one', namely **oi-* (see above), there is every reason to think that we have here an original nominal stem. Its derivations (Latin *unus*, English 'one', etc.; Sanskrit *eka-*; Avestan *aēva-*, Greek $\omicron\acute{\iota}\omicron\varsigma$ 'one alone') are all declined as regular adjectives. In Greek the alternative forms $\omicron\acute{\iota}\nu\eta$ or $\omicron\acute{\iota}\omicron\varsigma$ are nouns meaning 'ace on dice' and the present writer has always felt that the concrete meaning of this latter word ought to be something like that of English 'pip' (originally 'seed' of some fruits, then also 'pip' or 'spot on dice')¹⁷. The stem

14. Usually, this **oi-* is identified with a pronominal stem, viz. the one in Latin *is*, *ea*, *id* (see BRUGMANN, *Grundriss*, vol. II, p. 7, and J. POKORNY, *op. cit.*, p. 286), but there are two objections to this conjecture: firstly, the pronoun does not seem to have the initial diphthong **oi-* anywhere (while **ei-* and **i-* are common), neither does the numeral seem to have any other diphthong (the Slavic *in* 'other' need not have initial **ei-*, but may easily be equaled to Latin *unus*); secondly, the stem **ei-* does not seem to have the meaning 'the same', which is fundamental in the forms beginning with **s-* (see above) and which underlies and explains the sense of 'one' (cf. English 'one and the same').

15. Cf. *betan* 'together', *bet betan* 'just now', etc.; see J. M. SATRÚSTEGUI, *Vocabulario popular* (Euskera, vol. VIII-IX. Bilbao, 1963-64), p. 262. In many Basque dialects the choice of the vowels *a* and *e* depends on surrounding sounds; in Biscayan the pronunciation is regularly *bet* after an *-i-* or *-u-* in the preceding syllable (*egun bet* 'one day, a day', *arri bet* 'a stone'). The form *bet* is probably also found in the numeral *bederatzi* 'nine' (cf. below).

16. Whether Basque *bete* 'full' could also be considered a derivation of the same stem is more questionable (= 'whole', 'one' ?); in certain cases *bete* is actually used in the sense of 'one': *urtebete* 'one year, a year', *astebete* 'one, a week' (literally '(a) full year, week').

17. Cf. the author's *How Does Language Change?* (see Note 7, above), p. 78, with Note 85, and further a more general discussion in a paper with the title *Some Semantic Problems in Cuna and Kaggaba* (International Anthropological and Linguistic Review, vol. I:

*oi- would belong to an important semantic group, comprising all sorts of concrete nouns commonly used in *counting* ('stones', 'pebbles' –Latin *calculus*– 'seeds', 'kernels', etc.), including such as in a still earlier evolutionary stage were, from the semantic point of view, equivalent, forming a one-time single concept, but later differentiated into words meaning 'bud', 'flower', 'fruit', 'bone', 'egg', 'eye' and even 'star'¹⁸. In some languages the 'spots' or 'pips' on a die are called 'eyes' (German *Auge*, Russian *očko*, Hungarian *szem*) and even though the English word 'eye' itself cannot be etymologically connected with the Indo-European stem *oi-,¹⁹ yet another word, which in Middle English was so similar in sound to that of 'eye' that it had to be replaced by the Norse (or possibly northern) form 'egg', is undoubtedly to be included into the above semantic group²⁰. As a matter of fact, the Indo-European words for 'egg' (German *Ei*, Latin *ovum*, Greek $\omega\omicron\nu\nu$) are in many cases directly derivable from a stem *oi-: Old Church Slavonic (and Polish) *jaje*, Albanian *ve* or *voje*, etc.²¹.

It ought perhaps to be briefly mentioned here that in Austronesian and American Indian languages 'one' and 'another' are generally a single concept (often implying the idea of 'different' or 'strange'): Fijian *tasi*, Maori *tabi*, Arawak (South America) *aba*, which all denote (1) 'one' and (2) 'another'. In Quechua *huc* 'one' (or 'other') connotes 'strange' or 'stranger' (cf. *hucpa ricran* 'in one's or another's arms', that is 'in a stranger's arms'). This semanteme is reflected in Slavic: cf. Old Church Slavonic *inŭ* (= *unus* and *alius*), which is historically the same as English 'one' (in modern Russian *inoj* means 'another').

2-3, Miami, 1953), pp. 199-200, with further references. Notice that both in Basque and Malay eggs are counted by the classifier 'grain': Basque *iru ale* 'three grains (of eggs)', Malay *tělur tiga butir* 'egg, three grains'; in Haitian Creole the phrase is also *you* (or *gnou*) *grên zé* 'one grain (of) eggs', that is 'one egg'.

18. Cf. the author's *Some Semantic Problems* (see Note 17, above). «Egg' in Nahuatl is 'bird-stone' (*totoltell*) and 'eye' is 'stone' (*whatu*) in Maori, where 'stone' in its proper sense is expressed, for distinction, by a derivation *kowhatu* (as also in Hawaiian, 'ohaku').

19. The best etymology of the word 'eye' (German *Auge*, etc.) is the one suggested by Ernst Lewy; v. *Kleine Schriften* (Berlin, 1961), p. 46 and p. 59, with Note 1.

20. It would perhaps be more appropriate to assume an adaptation of a purely English word (cf. Anglo-Saxon *ecg* 'edge') than to speak of plain borrowing of the Norse word meaning 'egg', although we ought probably not to go so far as to say that a connection exists between the ideas of 'egg' and 'edge'; this curious fact nevertheless obtains in Amerindian languages (cf. the author's paper *Some Semantic Problems*, quoted in Note 17, above, p. 200) and is reflected in the Malay use of the word *mata* 'eye' of the 'cutting edge (of a knife)'. Within the field of Indo-European it almost appears a freak of the linguistic evolution that Old Irish *og* (in no way formally related to either English 'egg' or 'edge') does mean (1) 'an egg' and (2) 'an edge or point of a weapon' or that the Czech *vejce* appears a diminutive of *oj* or *voj* 'adze' (which latter actually is from the above stem *oi-, probably representing a more primitive sense 'stone ax').

21. Cf. the author's *Qualitative and Quantitative Evolution of Initial Vowels in Balto-Slavic* (in Annual Report 1948-49, by the Slavic Institute at Lund), pp. 139, 147. The current idea that Latin *ovum*, Greek $\omega\omicron\nu\nu$ are related to the word for 'bird' (Latin *avis*, Greek $\alpha\iota\omega\nu\nu\omicron\varsigma$, cf. J. POKORNY, *op. cit.*, pp. 783-84, and WALDE-HOFMANN, *Lateinisches etymologisches Wörterbuch*, p. 231) is as unrealistic as it is typical of the «Neo-Grammarians» school of the past century. Latin *ovum*, Greek $\omega\omicron\nu\nu$ presents certain phonetic difficulties for a connection with our basic *oi- 'egg'; the former, however, might possibly be to expected **uwum* (that is the Avestan stem *aēva*- 'one, from *oiwo-) as *deus* is to the stem *divo*- (from **deiwo*-). More directly connected with a stem *oiwo- would be the Latin *uva* (from *oiwā; cf. WALDE-HOFMANN, *op. cit.*, p. 849).

If the comparisons made in the preceding sections are valid, the Indo-European **oino-* (**oiwo-*, **oiko-*) would come closest in meaning and function to the Malay and Indonesian *suatu* or *satu* 'one' (from **sē-watu* 'one stone'; cf. Hans Kähler, in *Grammatik der Bahasa Indonésia*, Wiesbaden 1965, p. 67). It must however not be thought that the «counter», or counting word ('seed', 'stone',; etc.) is restricted to denoting the numeral 'one'. In Nahuatl, for instance, 'two' (*ome*) seems originally identical with the noun stem *omi-* (cf. *omitl* 'bone') and in the Austronesian languages 'three' *tělu* shows a strange similarity in sound, on all levels, with the noun stem **telul* 'egg' (cf. Malay *tělur*)²². In Quechua 'two' is *iscay*, which apparently contains the common Amerindian «root» **is-* 'eye', whereas the element *-ca-* is a common word for 'two' (cf. Maya *ca* 'two') and *-y* a common nominal termination in Quechua, letting us suppose that the whole complex simply expresses 'the two eyes' (cf. the use of 'eye' for 'one', in one of the initial sections of this paper).

'Two'. –The Indo-European base which corresponds to the number «two» (or rather to duality in the sense described above) appears to have been originally no more than the element **w-* (**wo*, **wi-*, **u-*, etc.), still seen in the Latin *viginti*. That the quite common forms **dwo-*, **dwi-* (Latin *duo*, *bis*, Greek *δίο*, *δίς*) are not basic is evident not only from Latin *viginti*, Greek *εἴκοσι*, *εἴκατι*, etc., Sanskrit *vimsati-* (which could reasonably mean nothing but 'two tens'), but also from the Tokharian forms, which are *wu*, *we* (= *duo*, *duae*) in the dialect A and *wi* (cf. Latin *bis*) in the dialect B. Hence the initial *d-* (in *duo*, etc.) would be a secondary accretion (comparable to the syllable *de-* in *decem*, etc.; cf. above). The diffusion and varied employment of the simpler forms in **w-* betray their great antiquity; a form **wi-* might even underlie both the Basque and Sumerian words for 'two'²³. More certain is however the connection of our basic **w-* with certain other Indo-European forms, in the first place the dual suffix *-u* in some languages (cf. Sanskrit *dvau* 'two', = Vedic *dvā*, *duvā*, Sanskrit *vrkrau* 'two wolves', = Vedic *vrkā*, etc.); the same element is possibly also found in Old Norse *tvau* 'two' (cf. however Brugmann, *Grundriss*, vol. II, p. 10). As an initial element it appears in the Indo-Iranian forms for 'both' (Sanskrit *ubhau*, Vedic *ubhā*, Avestan *uva*, *uba-*)²⁴. The numeral 'two' is inflected as a dual in Indo-European: the Sanskrit *dvau* therefore expresses the idea of 'two' by a repetition of the element **w-*, first in the stem itself (*-v-*) and secondly in the dual suffix (*-u*). The same holds for the Semitic languages (Arabic *ipnāni*, Hebrew *šnayim*, feminine *štayim* 'two') and Old Egyptian and Coptic (*snau*)²⁵.

22. Cf. OTTO DEMPWOLFF, *Vergleichende Lautlehre des austronesischen Wortschatzes* (Berlin, 1934-38), vol. III, p. 134. In Melanesian (Solomon Islands) the words for 'egg' and 'one' are still more similar: Nggela *tolu* 'three' and 'egg', Sa'a and Ulawa *olu* 'three' and *sa'olu* 'egg'. The formal association of the two stems is accentuated by the Krama form of Old Javanese *hantělu* 'egg', which is *tigan* because *tiga* is the Krama word corresponding to Old Javanese *tělu* 'three'.

23. Basque *bi*, Sumerian *min* are both derivable from **wi-* (in the latter languages there appears to exist a variant *man* (derivable from **wo-*); both languages lack the phoneme **w*, which appears regularly evolved in the words quoted).

24. Cf. BRUGMANN, *Grundriss*, vol. II, p. 11.

25. While it seems impossible to assign a concrete (or substantival) meaning to the dual

'Three'. -In the Indo-European languages 'three' is expressed by the base **tr-* (usually by extension **tri-*, as in Greek τρεῖς, τρι-, Sanskrit *trayas*, *tri-*, Greek τρίτος 'the third'; probably also Latin *ter*, from **tris*, and *tertius*, from **tritjo-*); the shorter stem (**tr-*) is to be postulated for Sanskrit *trtīya-* 'third' (so that **tri-* is actually to **tr-* ('three') as **wi-* is to **w-* ('two'; cf. above).

The morpheme **tr-* (**ter-*, **tor-*, etc.) is an important grammatical element in the Indo-European languages, expressing among other things comparative (and analogous) forms (as in Greek βέλτερος 'better', Latin *alter*, Sanskrit *antara-* 'other'-cf. *alius*, Sanskrit *anya-*, idem-English 'other', etc.), certain terms of relationship (as in Latin *pater*, *mater*, *frater*, *matertera* 'mother's sister') or finally agent nouns (Latin *actor*, Greek ῥήτωρ etc.), which all reveal but little connection with the notion of 'three'²⁶.

There are also concrete nouns in the Indo-European languages which are doubtless connected with the stem of the numeral 'three' (apart, of course, from such obvious derivations as 'triple', 'trinity', 'trivet', etc.). In social systems divisions into a specified number of sections is a frequent occurrence (often reflected in the toponymy: *Tripolis* 'the Three Cities', *Tetrapolis* 'the Four Cities', Gaulish *Tricorii* 'the Three Peoples', etc.). In early Rome a *tribus* (= English 'tribe') was according to tradition one of *three* divisions (*Ramnes*, *Tities* and *Luceres*) into which the Romans were divided and the word has commonly been understood to have a connection with the numeral 'three' (*tres*, *tria*)²⁷. However, although Latin *tribus* cannot be assumed to mean the 'third part (of the Roman people)'-at least not according to the current type of Latin or Indo-European derivation-a connection with the numeral 'three' is not excluded. On the other hand, the Latin *tribus* appears connected with the Celtic **trebā-* (Old Irish *treb* 'tribe,

w-*, this is perhaps less so in the case of the corresponding Semitic and Hamitic stems (p-n-* and **śn-*, respectively). At least the latter might have a connection with Old Egyptian *śn* 'brother' (Coptic son), *śnt* 'sister' (Coptic sōne); these words would be related to the numeral in about the same way as English 'twin' is to 'two'. The Semitic root *p-n-y* means 'bend' or 'twist' (there is also an Old Egyptian verb *śnh* 'bind, tie', Coptic sōnf, *śnśn* 'unite'). In Eskimo, 'two' (*mardluk*, in which *-k* is a dual suffix; cf. Indo-European) is held to be connected with the verb *maligpā* 'follows him' and would thus simply mean 'following ones' or 'such as follow each other (together)'.

26. Cf. Caroline T. STEWART, *op. cit.*, p. 248. To trace a semantic connection between the various morphological functions of the element **tr-* (cf. above) is a perilous undertaking, but might be ventured. The base **tr-* seems to imply 'something exceeding' or 'farther away from (something else)': 'better' is 'exceedingly good' or 'more than good', *matertera* expresses a degree 'farther away than *mater*', etc. From this sense is derived the one of position (*pater*, which is the sociological term corresponding to the more vague and conversational *pappus*, Greek πάππος 'any old man' or 'grandfather') or order (*alter*, *uter*, etc.); the restriction to the number 'two'-as also in English 'elder'-is probably secondary. The use of the same suffix here as in forms of the comparative degree (in Greek and Sanskrit) strongly recalls the analogous use of comparative and superlative forms in Lapp terms of relationship: cf. southern Lapp *áččeb* ('father'), *cižž' emes* ('mother'), the former (having the comparative suffix) used in relation to *one* child, the latter (having the superlative suffix), in relation to *two* or more children (cf. the respective use of 'elder' and 'eldest' in English).

27. This was maintained as early as Varro. It is accepted unconditionally by Frederik MULLER (*Altitalisches Wörterbuch*, pp. 494-95); in Ernout and Meillet's *Dictionnaire étymologique* (p. 1241) it is qualified as a «simple supposition».

gens', Welsh *tref* 'town')²⁸ and the Teutonic **þurpa-* (as in German *Dorf* 'village' or the English '-thorpe' in place names). Since there is no reason for us to assume that the Latin word is borrowed from Celtic, we have to count on two basic stems: **trib(h)-* (for the Italic words; cf. Umbrian *trifu*, *trifo*) and **trb(h)-* (for the Celtic and Teutonic words), which both appear to correspond in some way to the basic forms of the numeral 'three' (as postulated above)²⁹ and would hence prove a semantic connection between the ideas of 'three' and 'tribe' (or '*social section').

We shall however adhere to the principles of «comparative semantics» in so far that we shall not attempt to analyze the further nature of this relation³⁰, but limit ourselves to point to an analogous relation in another field. In Basque 'three' is *iru* (*irur-*; in French Basque *hiru*, *hirur-*), of which a more primitive base **ilu(r)-* can be reconstructed. To this base appear to belong such ethnic terms as *Iluron*, *Illurgavonenses* (or *Ilergavonenses*), *Ilergetes*, etc., which seem in some way analogous to the Gaulish *Tricorii* (for which cf. above). Now the name for a 'city' in Basque is *iri* (in French Basque *hiri*; the Biscayan form *uri* is doubtless a secondary one) and if we were allowed to compare these forms with those of the numeral 'three', we should arrive at a connection analogous to the one between Latin *tribus* and *tres* (cf. above). As a matter of fact, the Basque word for 'city' has become rather famous in the linguistic literature as having been used in attempts to demonstrate the one-time expansion of the Basque-Aquitania linguistic area throughout large parts of the Iberian peninsula by connecting the Basque *iri* with the first element in the old name *Iliberis* (surviving in the present time in the place name Elvira, in Granada), which was ingeniously interpreted as meaning the 'new town' (Basque *iri berri*), that is the same as the French *Neuville* or the English 'Newto(w)n'. The Basque word *iri* may very well represent the evolution of an earlier **ili* and the connection with the above-mentioned **ilu(r)-* is reasonable in the light of the comparisons made above³¹. A connection might even exist with the place names *Irún* and *Iruña* (the Basque name of Pamplona, in Navarre).

In connection with the Basque numeral *iru* 'three', another semantic combination suggests itself, viz. the one with certain verbs denoting a handling of 'three' (parts or strands, etc.). The Basque word for 'to spin' is *iruin* or *irun* (in French Basque also *hirun*; hardly related to Latin *filum* and Spanish *hilar* 'to spin') and in Latin *torquere* has an identical meaning; the relation of the latter to the base **tr(i)-* 'three' becomes conclusive if we further compare the Germanic verbs for 'to turn, twist' (German *drehen*,

28. Cf. Alfred HOLDER, *Alt-Celtischer Sprachschatz* (Leipzig, 1904), vol. II, pp. 1908-9.

29. The Teutonic words, however, might have been borrowed from Celtic (?).

30. See the author's *Comparative Semantics* (referred to in Note 10, above), p. 99.

31. In Basque *iru(r)*, the element *-u-* is an increment (as is *-i-* in Indo-European **tri-*), as seen from the derivations *eren* 'third part', *erenegun* 'the day before yesterday'. If Basque has any connections at all-apart from structural ones-with the ancient languages of the Caucasus, the Near East and the Mediterranean coasts and islands, a comparison with the ethnic term *Illyrii* (*Ilurii*) might be considered-which would have been the 'three' ancient tribes of the Illyrians?-if not also one with the ancient name of Troy (Latin *Ilium*, Greek Ἴλιος or Ἴλιον, representing a Pre-Indo-European **ili*). However, it should be noted that Greek Ἴλιον originally has an initial digamma.

English 'throw', along with the noun 'throe'), on the one hand, and the indubitable connection of the English verbs 'to twist', 'twine', 'twirl' with the numeral 'two', 'twain'; the idea is simply the one of winding together respectively 'three' or 'two' strands in making a string³².

'Four'. –As in the case of the numerals 'two' and 'three', the Indo-European languages show traces of the existence of duplicate stems for the numeral 'four' also: *k^hetur-, which is the normal base, and a shorter form *tur-, which however may be the more primitive one (cf. above); the latter is especially represented in the Sanskrit ordinal number turīya- or turya- 'fourth'³³.

The semantic connections of the concept 'four' have been treated somewhat at length by the author in an earlier article *Semántica y etimología*³⁴. The essential points of that article were the connections existing between the number 'four' and various spatial concepts, reflected in the use of the English word 'square' (ultimately from Spanish *escuadra* or the Italian *squadra*, being derivations of a Latin verb *ex-quadrare* 'to square', in various senses). Some of these words are definitely-like the nominal derivations of 'three' dealt with in a preceding section-based on the ordinal-partitive sense ('fourth') e.g., the Spanish *cuarto* in the sense of a 'room' (originally a 'fourth part', like 'tribe' = a 'third part'); also compare the Arabic *rabba'a* 'to square', *tarabba'a* 'to square one-self', that is 'to spread oneself out, as when sitting' (which is from the same «root» *r-b-* as the numeral 'four'). The semantic congeners of 'four' mentioned in the article referred to (see Note 34) do not by any means pretend to represent the primitive concrete meaning of the numeral (any more than the ideas of 'tribe' or 'twist' or 'spin'-see above-can pretend to be primary in relation to those of 'two' and 'three'). Some of the associations with the mentioned spatial concepts may, however, indicate a considerable age of the semantemes concerned. It is interesting to note the relation between the Basque *lau* 'four' and *lana* or *laba* 'a plain' or *labe* 'oven', the latter no doubt originally, like the English 'stove', the same as a 'room' (cf. German *Stube*; Sanskrit parallels were quoted in the mentioned article). We shall not enter here upon reasons for the connections which seem to exist between the ideas of 'four' and 'room', etc., but content ourselves, as above, to take stock of the associations found in the various languages. Starting from the fact that designations of 'room', 'square', etc., have comparatively modern associations with Latin *quattuor*, *quartus*, one might venture to trace an analogous connection between other

32. The Latin verb is, of course, related to the Greek *τρέπω* 'turn', which does not have the sense of 'twist'; a suspicion that such may have been its original connotation is, however, confirmed by the existence of a Greek noun *τροπός* meaning 'a twisted leathern thong'.

33. The current opinion is that the forms in *tur- represent a syncope of original *k^h(e)tur (whether or not this be warranted by any phonetic laws or tendencies), especially in consideration of the Avestan *āxtūivm* 'unto the fourth (time)', that is 'four times' (cf. BRUGMANN, *Grundriss*, vol. II, p. 14; J. POKORNY, *op. cit.*, p. 643). Similar shorter forms are however found in Greek compounds also (*τρέπεζα*, *τροφάλεια*; v. BRUGMANN, *Grundriss*, vol. II, p. 15; BOISACQ, *Dictionnaire étymologique de la langue grecque* (1950), p. 988). While sketching this paper, it occurred to the writer that the *k^he- of *k^hetur might possibly be a trace of an enclitic 'and' at one time added to the preceding numeral ('one, two and three, four and a span') without knowing that the same thought had occurred to Caroline T. Stewart more than fifty years earlier (see *op. cit.*, p. 239, Note 3).

34. *Semántica y etimología* (referred to in Note 13, above), pp. 391-92.

stems of which the meaning is similar and the following combination is perhaps suggestive. As indicated in the above-mentioned *Semántica y etimología* (pp. 391-92), the Basque numeral (*lau, laur-*) is—as we assumed above for the numeral ‘two’ *bi*—possibly related to the corresponding Sumerian numeral, which is either *limmu* or *lammu* (and in which also the *-m-* may represent **-w-*; cf. above). The primitive form underlying the numeral ‘four’ in these languages (which both belong to the early Pre-Indo-European and Pre-Semitic stratum) would then have been something like **lawu(r)-* (cf. the reconstructed form **ilu(r)-* for ‘three’, above). Now there exists in Greek a well-known word, evidently belonging to the Pre-Hellenic period (that of the «Pelasgian», «Minoic» or «Eteocretan» languages), ending in the typical suffix *-vθos*,³⁵ namely the word for ‘labyrinth’ (in Greek λαβύρινθος). This was the name of a large kind of building with innumerable rooms and intricate passages typical of an early period of the Mediterranean culture, especially in Crete. These palatial structures would have been the kind of «squares» denoted by the Pre-Indo-European term **lawur-nt-* (**laβur(i)nθ-*), being a derivation of an old word for ‘four’. If we assume that *-nt-* (*-nθ-*) furthermore corresponds to the collective suffix in Hittite, we should even be able to translate this mysterious Pre-Hellenic word by ‘collection of rooms’ (and declare it the equivalent of the Hittite *parnant-*; cf. Note 35).

‘Five’. —It has been supposed since the time of Franz Bopp that the last syllable of the Indo-European **penkⁿe* ‘five’ (that is Sanskrit *pañca*, Greek πέντε, etc.) is the enclitic **-kⁿe* ‘and’ (Latin *-que*, Greek *-τε*, etc.), used in counting (‘three, four *and* five’)³⁶ a theory which might find further support in the existence of a Tokharian form *pāñ* (‘five’) in the dialect A (the dialect B has the form *piś*), in which the would-be enclitic element seems absent³⁷. Any possible connection of this form (**penkⁿe* or **pen-*) with any other Indo-European word, except the Teutonic word for ‘finger’ (Gothic *figgrs*; cf. J. Pokorny, *Indogermanisches Wörterbuch*, p. 808), seems excluded³⁸. Usually the Western Indo-European languages—owing to assimilation either to the initial sound of **kⁿetur-* ‘four’ or to that of the last syllable (**-kⁿe*)-have transformed this word into **kⁿenkⁿe* (which explains the Latin *quinque*, Oscan *pomp-*, Old Irish *cóic*, Old Welsh *pimp*)³⁹.

35. Possibly surviving in the collective suffix *-nt-* in Hittite (analogous formations are met with in Tokharian and Celtic as well), e.g. *parnant-* (from *parn-* ‘house’), which Johannes Friedrich renders by *Hauswesen*.

36. See especially Caroline T. STEWART, *op. cit.*, pp. 229, 239.

37. Unless the dialect A form could be explained as a regular evolution of an original **pāns* (corresponding to dialect B *piś*), which is however unacceptable as long as the final *-ns* is a possible word-final phonetic sequence in this language (cf. the plural of *wāl* ‘king’, which is *lāns*). It must further be observed that the corresponding ordinal ‘fifth’ is *pānt* in dialect A, which appears to be an original **pento-* rather than **penkⁿto-* (as in Greek πέμπτος), because Indo-European **oktō* ‘eight’ becomes Tokharian *ok(ā)t*; the dialect A form *pānt* is further confirmed by the existence of a numeral *paⁿta* ‘fifth’ in Hieroglyphic Hittite (see J. Pokorny, *op. cit.*, p. 808; the Tokharian dialect B, however, has *pinkce*, corresponding to the Greek ordinal).

38. Theoretically, a shorter stem **pen-* could be connected with the verbal stem **pen-* (as in Old Church Slavonic *peti* ‘span, hang’), of which the variant **spen-* recurs in English ‘span’ a ‘span’), as the equivalent of the extended fingers of the hand would be an appropriate designation of the number ‘five’ as would be ‘hand’ or ‘fingers’ (cf. above).

39. From the Old Celtic (or Gaulish) **pempe* (πεμπε-) come the Teutonic forms

With 'five' begins a new series of simple numerals (viz. in the quinary system), for which a name ('hand', 'fingers', 'span', 'fist')⁴⁰ was no doubt given, analogous to that of the first unit 'one' (viz. 'pebble', 'seed', 'grain', etc.; cf. above)⁴¹. When the 'span' is completed, the following numerals may be formed by addition of the words for 'one', 'two', 'three' and 'four' (as in the Sumerian *imin* 'seven', from **ia min* 'five-two', *elimmu* 'nine', from **ia limmu* 'five-four', or in the Roman system of notation: VI, VII, VIII). Apart from the quite hypothetical **pen-* (or **pene-*; see above), the element **kmt-* (**komt-*), if correctly connected with the Germanic word for 'hand' (cf. above)⁴², might have been employed, although it is actually found to be limited in its use to the number 'ten' (since the same stem could naturally be used in the sense of 'both hands' as well as in the sense of 'hand')⁴³.

'Six'. –The numerals above 'five' are characterized in two ways: (1) by showing traces of expressing a relation either to the numeral 'five' (as in Sumerian and Aztec) ⁴⁴ or the numeral 'ten'; and (2) by showing signs of having been borrowed from other languages which have earlier developed a full decimal system. In the Indo-European languages, the numerals 'six' and 'seven' appear to have been borrowed or else influenced by the corresponding Semitic and Hamitic forms. The numerals from 'six' on, in the various Indo-European languages, show less internal agreement than the lower ones⁴⁵.

In the Indo-European languages there are two slightly divergent forms of the numeral 'six': **seks* Greek, Latin, Germanic, Baltic-cf. Lithuanian *šeši*, Latvian *seši*-and Tokharian)⁴⁶ and **sweks-* (Indo-Iranian, Armenian, Slavic, Celtic)⁴⁷. In the Semitic languages the word for 'six' most often

(Gothic *fimf*, etc.); cf. the author's *How Does Language Change?* (referred to in Note 7, above), p. 75.

40. The latter ('fist') has actually been derived from the Indo-European **penkⁿe* (viz. from **pnkⁿti-*; cf. Sanskrit *pañkti* 'collection of five', then also 'line, row').

41. In Malay a 'unit' is called *bidjan* (from *bidji* 'a seed'), just as a 'digit' is called *puluhan* (from *puluh* 'ten').

42. Cf. BRUGMANN, *Grundriss*, vol. II, p. 4.

43. It may actually be said that it is used of 'a hundred' also, for the Latin *centum* is of the same stem; the latter may in reality depend on an abbreviation of a longer form meaning 'ten tens' (by analogy with Gothic *taihuntehund* 'hundred'). In southern Lapp *stuore-lukkie* ('big ten') means 'a hundred'.

44. In Sumerian, 'seven' (*imin*) and 'nine' (*elimmu*) are compounds clearly made up from 'five' and 'two' or 'four', respectively (cf. above). In Nahuatl the numerals from 'six' to 'nine' are: *chiquacen*, *chicome*, *chicuey*, *chicunau*, in which *chico-* means 'across', that is 'over to the other (or right) hand' and the elements *-cen*, *-ome*, *-uey*, *-nau*, respectively 'one, two, three, four'; this system is reflected in the Mayan numerical notation (for which cf. Caroline T. STEWART, *op. cit.*, p. 231, Note 2).

45. When speaking about «borrowing» in this case, one must remember that it is impossible to point to any particular case in which any particular Indo-European numeral corresponds phonetically to any particular Semitic or non-Indo-European form. It may even happen that an Indo-European «etymology» be found for any of these numerals (cf. Caroline T. STEWART, *op. cit.*, pp. 264-65), but the influences from outside are still unmistakable from the vague similarity which the Indo-European words for 'six' and 'seven' do reveal with non-Indo-European forms to be discussed in this paper.

46. The Tokharian (dialect B) form is *škas* (from **seks*); cf. dialect A *škäšt* 'sixth' (Latin *sextus*).

47. In Sanskrit and Slavic the initial consonant (*ṣ-* and *š-*, respectively) is perhaps that of Iranian (in Avestan *xšv-*), which in its turn may depend on an ancient reduplication

shows a form with two identical sibilants (Hebrew *šēš*, Akkadian *šiššu*, with which may also be compared the words for 'sixty': Hebrew *šiššim*, Akkadian *šuššu*)⁴⁸ and the case is similar in Old Egyptian: *šyś(w)* 'six' (cf. Coptic *soou*, *sau* feminine *so(e)*, *sa*)⁴⁹. The two sibilants clearly answer to the *s*-sounds of the Indo-European forms (**seks*, **sweks*), in which by analogy with Semitic and Hamitic a middle consonant appears to have evolved secondarily (in the latter languages evidently to conform with the three-consonantal word pattern)⁵⁰. It is worth noticing that such an additional consonant is missing in the Basque form of the numeral 'six' (*sei*, in which the final *-i* is the same as in the other numerals between 'five' and 'ten': *zazpi*, *zortzi*, *bederatzi*). If the Etruscan *ša* corresponds to 'six', as seems *a priori* likely and was formerly supposed, the sibilant alone is the characteristic consonant.

Regarding the semantics of the numeral 'six' in certain American Indian languages, the author would like to refer to his paper *Some Semantic Problems in Cuna and Kaggaba* (International Anthropological and Linguistic Review, vol. I: 2-3), pp. 196-97, in which the relation between this numeral (in sign language corresponding to the raised thumb of the right hand) and the word for 'man' (with a corresponding symbol in the sign language) was pointed out. According to the same method of indicating the number by signs 'seven' ought to be indicated by raising the index finger of the same hand. The phonetic resemblance of the Malay word for 'index finger' (*tundjuk* or *tēlundjuk*, with which cf. the Tongan *tuhu* 'finger', especially 'index finger', and the Malay verb *tudju* 'aim, direct') and the word for 'seven' (*tudjuh*) is suggestive. In Arabic the verb *šaba'a* 'to point with finger' has almost the same consonants as the numeral 'seven' (*s-b-*), of which the first one is slightly different (the Old Egyptian *db*, that is Coptic *tēēve*, 'finger', which also differs in the first consonant, is worth noting in this connection).

'Seven'. -What has been stated for 'six' holds still more for the numeral 'seven'. The Indo-European words are based on the forms **sept-* and **sep-*, of which the latter appears more original on account of the fact that several numerals above 'five' contain an element *-t-* (apparently having an accretive character: English 'eight', Latin *octo*, Sanskrit *aṣṭau*-notice that the *-t-* is absent in the form for 'eighty', *aṣīti-* in Sanskrit and Lithuanian *dēšimt*,

(**sweksweks*, falsely restituted into **ksweks* (?); such reduplicated forms are not commonly found in Indo-European, although they may be common elsewhere (cf. Georgian *samisami* 'three each', *ot'xot'xi* 'four each; Haitian Creole *dis dis* 'ten by ten'). In this way the Georgian forms *ek'vsi* and *ek'vsek'vsi* 'six (each)' could be explained as derived from those of an Indo-European language.

48. BRUGMANN (*Grundriss*, vol. II, p. 5) thinks Sanskrit *ṣaṣṭi* 'sixty' might be borrowed from Akkadian *šušši*, *idem*.

49. The middle consonant of this latter form (*-y-* or the «feather») is presumably a mere expression of the middle vowel (*-i-*, as in Akkadian *šiššu* 'six'; cf. also the Hebrew *šiššī* 'the sixth' and *šiššim* 'sixty'). To conform with the tendency toward three-consonant words, however, a middle consonant has been evolved in some Semitic languages (cf. especially the Arabic ordinal *sādis(un)* 'sixth').

50. Cf. the preceding note. Since in Tokharian (dialect B), the word for 'seven' (*sukt*) seems to have its medial *-k-* from the following numeral 'eight' (*okt*), it is perhaps permissible to assume that the numeral 'six' may have its middle *-k-* from the word for 'eight' (**oktō*) also.

etc.)⁵¹. Supposing this to be the case, the relation to the Semitic forms becomes evident: cf. Arabic *sab'(un)*, Hebrew *ševa'*, etc. Also the Egyptian form *sfb(w)* (Coptic *šasf*) shows a general similarity in sound, allowing us to conclude that it in some way or other copies the Semitic *s-b-*. For other forms derived from or influenced by Semitic, cf. in an earlier paragraph. In the Finno-Ugric languages the Indo-European forms seem to have served as a model: Hungarian *hét* represents an evolution of the stem **sept-* and the Finnish *seitsemä* shows an indubitable case of adaptation to the same Indo-European base, yet under the influence of some native word stem (cf., for instance, *seitse-* and *seitso* 'net'). –For the semantic connections of 'seven', see further in the preceding paragraph (in connection with 'six').

'Eight'. –The numeral 'eight' similarly shows a longer and a shorter stem: the most common one is **okt-* (Latin *octo*, etc.) and the more exceptional one **ok-* (as in Sanskrit *aṣīti-* 'eighty'; cf. above)⁵². The most interesting detail in respect of this numeral is perhaps that it has the appearance of being a dual (Sanskrit *astau*, Greek *ὀκτώ*), which has led some etymologists to believe that the meaning of this word is originally 'two fours' (cf. Caroline E. Stewart, *op. cit.*, p. 248). There seems, as a matter of fact, to be evidence enough of the presence of a **-w-* in certain forms of this numeral (cf. Latin *octavus*, Gothic *ahtau*), but this does not support the theory of the dual signification of **oktō(w)-*, for final **-w* as a dual suffix is not universally Indo-European (being restricted to the languages of India) and furthermore the Sanskrit *aṣtau* is not declined as a dual (the instrumental, for instance, being *aṣṭābhis*, as of a plural, and not **aṣṭābhyām*, as of a dual).

In Old Egyptian there may be a faint trace of 'eight' having been thought of as 'five and three', for the forms of 'three' (*hmt*) and 'eight' (*hmn*) have an identical beginning⁵³. In many languages, however, 'eight' is conceived as 'ten less two' (just as 'nine' is 'ten less one'). This is seen most clearly in the Finnish forms *kahdeksan* 'eight' (cf. *kahde-* 'two') and *yhdeksän* 'nine' (cf. *yhde-* 'one') and has, probably rightly, been supposed to be the case of Malay *délan* 'eight' (cf. *dua* 'two') and *sēmbilan* 'nine' (cf. *sē-* 'one')⁵⁴. The same principle recurs in Latin, where 'eighteen' is designed by *duodeviginti* and 'nineteen' by *undeviginti*, etc. (reflected in the numeric notation IIX (= VIII), IX, XIIX (= XVIII), XIX) as well as in Sanskrit *ekonaviṃśatih* along with *navadaśa* ('nineteen'). The latter type of construction (representing a compound with the word *ūna-* 'wanting, deficient')

51. The general impression among Indo-European scholars is that **sept-* is the original form and that **sep-* depends on a secondary evolution, perhaps in derivations of the type **septmo- > *sepmo-* 'seventh' (cf. BRUGMANN, *Grundriss*, vol. II, p. 18; J. POKORNY, *op. cit.*, p. 909; Sigmund FEIST, *Vergleichendes Wörterbuch der gotischen Sprache*, Leiden 1939, p. 417).

52. Hermann MÖLLER (*Semitisch und Indogermanisch*, Copenhagen 1906, p. 63) combined this **ok-* with the Semitic **aš-* (in Arabic *ašr(un)* 'ten'); this might be acceptable semantically, but whether an Indo-European palatal (or palatalized) **k* is likely to answer to a Semitic **š* or not is still a question.

53. In Sumerian 'eight' does not seem to have any relation to 'three', as far as the form goes (the former being *ussu*, the latter *eš*).

54. It was formerly often thought that the final part is **-deksan*, a borrowing from proto-Iranian **dek'an* 'ten', but this seems unlikely.

gives a hint as to the natural interpretation of similar formations in other languages. When in Basque 'nine' is *bederatzi*, which we have assumed (see above) to contain the element *bet-* (= *bat* 'one'), there is every reason to think that this numeral, as well as possibly also 'eight', are formed according to the above principle. The Basque *zortzi* 'eight', which does not (like *sei* 'six' and *zazpi* 'seven') appear to have any connection with forms in Semitic, Hamitic or Indo-European, could only be associated with one concrete word stem in Basque, viz. that of the word *zurtz* 'orphaned' (especially in compounds such as *omezurtz* 'orphan')⁵⁵. This leaves, of course, the element 'two' unexpressed or understood, which may be explained as due to a tendency toward abbreviation not uncommon in numerical expressions (cf. Latin *deunx* 'eleven twelfths', which in reality does not say any more than 'an ounce missing'). One would expect an analogous construction for 'nine' in Basque, but here the final element is a different one (*-eratzi*), which may either be a corruption of **-zortzi* or quite another word (cf. the verb *eratzi* 'to lower'?). The numerals 'eighteen' and 'nineteen' in Basque are formed by analogy with the other numerals between 'ten' and 'twenty' (that is from 'ten', Basque *ama(r)-*, combined with the corresponding lower numeral, as in *amabi* 'twelve', *amairu* 'thirteen', etc.), but it is noteworthy that 'nineteen' shows the final element only (that is *-eretzi*) of the numeral 'nine' (*bederatzi*) and that both 'eighteen' and 'nineteen' have the form *eme-* in place of *ama-* (*emezortzi*, or sometimes *ama-zortzi*, and *emeretzi*, respectively), perhaps a reminiscence of a former (as it would seem) deviating behavior of these two numerals (the vocalism of *emezortzi* may possibly imitate that of the following numeral).

'Nine'. –Certain details regarding the number 'nine' were touched upon in dealing with the preceding numeral. The numeral 'nine' is made up from a base **newn-* in most Indo-European languages, except in Greek, where a base **enewn-* must be supposed to explain the form ἔννεά (the double *-vv-* is no doubt due to the influence of the ordinal number ἔννατος or εἴατος, representing a primitive **enwnto-* 'ninth'). The only concrete word stem similar in form to be found in Indo-European is **new-* (the one in English 'new'; cf. Latin *novem* and *novus*), with which a connection is, however, hardly warranted semantically⁵⁶. If we pay some attention to the occurrence of an initial vowel in the Greek form (ἔννεά), another interpretation of this numeral suggests itself. Remembering the tendency of expressing the number 'nine' in relation to that of 'ten' as something having a unit *wanting*, a combination with the Indo-European preposition meaning 'without' (Greek ἄνευ, Gothic *inuh*, of which the latter at least seems based on a primitive stem **enu-*) would be more appealing. Like the Basque *zortzi* (cf. above), the numeral 'nine' would simply mean 'orphaned (ten)', that is a 'ten' from which a unit is wanting, and as in the case of the Basque numeral the number wanting to make up ten is left unexpressed.

55. As a matter of fact, it could be the same word as *zortzi* 'eight': in Basque *o* and *u* are originally not different phonemes (cf. *ortze* ~ *Urtzi* 'heaven, thunder, God') and the termination *-i* (in *zortzi*) is that of the other numerals between 'five' and 'ten' (cf. above).

56. Early explanations of 'nine' as being the 'new number' (cf. Caroline T. STEWART, *op. cit.*, p. 243, Note 3) do not seem to make much sense.

'Ten'. –This is the unit of the decimal system and the corresponding numeral might be expected to have had a concrete meaning at one time, by analogy with what we have supposed to be the case of the numbers 'one' and 'five'. In this way the Lapp word for 'ten' (*lāken*, in the southern dialects *lukkje*) is nothing but the Finno-Ugric word for a 'number' (cf. Finnish *luku* 'counting, number, sum', etc.)⁵⁷. For the Indo-European words, a connection with a base meaning 'hand' (**kmt-*, **komt-*; cf. Gothic *handus*) has often been suggested and seems rather satisfactory⁵⁸. For the explanation of the initial **de-* in most Indo-European forms of the numeral 'ten' (Latin *decem*, Greek δέκα, etc.), cf. above⁵⁹.

The way in which numbers have been expressed by written symbols is often instructive. If 'one', for instance, is represented by a dot or small circle (as in the Mayan or Mexican systems), this symbol is not simply to be interpreted as an abstract sign, but rather as a picture of any of those small objects ('seeds', 'pebbles', *calculi*) which could be used in counting (cf. above); the strokes (as in the Roman notation) would stand for sticks used in the same way (Caroline E. Stewart, *op. cit.*, p. 231, Note 2, supposes that they represent fingers). The arrangement of the numerical symbols in the Roman notation corresponds either to a quinary system ('six', 'seven', 'eight' = 'five' plus 'one', 'two', 'three', respectively, expressed by VI, VII, VIII) or else according to the system of counting from the higher basic numeral ('nine' = 'ten short of one', that is IX, etc.). When the concept of zero was invented, it was significantly expressed—at least in some systems—by a circle, being the symbol of emptiness⁶⁰. If it is true, as has sometimes been supposed, that the Roman symbol for 'ten' (X) represents a picture of the two hands, this might not be without importance for the understanding of the concrete meaning of the word stem used to express the number ten. When in the Babylonian representation of numbers the same three wedges could be used for 'three', 'thirty' and 'three times sixty', etc., this is important for the understanding of how the higher numerals were formed even linguistically on the basis of the lower ones (as we have seen in the case of 'six' and 'sixty' or 'ten' and 'hundred', above). Here Jacques van Ginneken's theory of the priority of sign to articulated sound may have some application (see *La reconstruction typologique de la langue archaïque de l'humanité*, in Verh. Kon. Akad. van Wetenschappen, n.r., afd. letterk., vol. 44 (1939), especially, §§ 88, 122).

This paper is not intended to be «glottogonic», but in the first place a treatise on semantics. Its purpose is to point to associations of ideas in regard to the numerals in various languages and to constructions which seem to be copied in more than one language. These observations may quite naturally

57. In a vigesimal system 'twenty' may be expressed in this way; cf. the Nahuatl *cemponalli* 'twenty' (literally 'what is counted', from the verb *poua* to count').

58. Cf. Caroline T. STEWART, *op. cit.*, pp. 239, 254.

59. Accepting a word meaning 'hand' as a designation of 'ten' instead of 'five' would be the same as to assert that this word could also mean 'both hands': this is in fact no more strange than when 'one eye' is called *leath-shúil* ('half-eye') in Irish, implying that *súil* 'eye' could be understood alternatively as 'the two eyes'.

60. As a matter of fact, it was at the same time the sign of 'all' or 'whole' (cf. the Basque *uts*, which means at the same time 'empty' and 'whole, entire').

to some extent help in establishing «etymologies» of the numerals (some of which have been suggested in the preceding paragraphs by way of illustration). It is however the present author's opinion that the etymologization of numerals is an extremely precarious undertaking, on account of their having undergone so many and so radical modifications, due to analogy, assimilation (or attraction) and to the desire to arrange them into a symmetric system. Such systems are, however, late in appearing, as seen from the circumstance that words which seem identical or related in form have often been found in different languages or in different systems to have been given a different value. An important principle is to be derived from this fact: if, as we have suggested in a few cases in this paper, two numerals should happen to be identical in two languages not otherwise showing any general similarity, it must by no means be concluded *that the whole numerical system* is related or, still less, that two such languages must needs belong to the same linguistic family. This, of course, holds for all cases in which individual elements may be found to be identical in any two languages, but it should perhaps be more strictly observed in the case of the numerals, which are particularly flighty and transitory creations of language, in spite of their universal reputation of belonging to the central and fundamental stratum of the linguistic structure.