# AUSTRONESIAN NUMERALS[1]

Otto Chr. Dahl

A linguistic society possesses the numerals it needs for its daily occupations. A glance at the language of a Bushman tribe living in the outskirts of the Kalahari shows that they have words for 'one' and 'two', but above that only words meaning 'several' and 'many'. It is not that their language is particularly undeveloped. In their noun declination for instance they have forms for singular, dual and plural. Thus perfect accordance between numeral system and noun declination!!

The reason for this limited numeral system is that they do not need to count anything. They are food gatherers, and there is no point in counting the roots and small animals they have got during the day, because their sizes differ. A glance at what has been gathered is enough to estimate if it is sufficient for the family's supper and breakfast.

Cattle breeders need more numerals. They have to count their herds in the evening to see if they have got all the beasts home to the enclosure. But these Bushmen have no domestic animals.

Even very primitive forms of trade call for numerals. Bargaining may begin with a question like this: "How many of the things you have, are you willing to give in exchange for so and so many of the things I have?"

What do we know of the social conditions of the Protoaustronesians and their need for numerals?

They were seafarers. Otherwise they could not have come to the thousands of islands where they are now living. Words for 'canoe' *ua(N)kaŋ, 'ship' *parahu, 'sail' *laiaγ 'paddle' *bəγət'ai, 'oar' *daiuŋ, 'rudder, stern' *udəhi, have a very wide distribution. Seafarers get food by fishing: 'fish-hook' *kauil, 'fish-trap' *bubu, 'fish-poison' *tɭuba are likewise common. A very common word for 'fish' is *Si-kaən, derivative of *kaən 'to eat' with the prefix of the fourth focus. Fish was thus a very important part of their food, and many names of fish species are common. But here the situation is like that of the food gatherer: no need to count the fish.

They were also hunters: *qaɬup 'to hunt', *buru 'to hunt, pursue', *but'uγ, *panaq 'bow, arrow, to shoot'. It was perhaps, therefore, that they kept dogs, *uat'u. They also had pigs, *bəγək, but names of the ruminant domestic animals are not so widespread. Therefore they were probably not cattle breeders.

They were however cultivators: *$t_2$aɬəm 'to plant', *qumaH$_1$ 'cultivated field', *pag'ai 'rice plant', *bəγat' 'husked rice', *ɬət'uŋ 'mortar', *qaS$_1$əlu[H$_2$] 'pestle', *qubi 'tuber, yams', *t$_1$əbuS$_1$ 'sugar cane'. Nor was much counting needed for this occupation.

But they practised trade: *bəli 'to buy'. The causative form of this wordbase is often the expression for 'to sell'. Trade may gradually have prompted the development of their numeral system.

1.

The most common word for 'one' is PAN *t'a, often preceded by *ə or *i. The monosyllabic form is sometimes prefixed to nouns. Mal sə-puluh 'ten' is a compound word, semantically comparable to German ein Zehner. The verb 'to count' is sometimes derived from this wordbase.

In Formosa cognates of *t'a are used in most languages. Only the Atayalic and Tsouic subgroups have other words for 'one' (Ferrell 1969,407). Also outside Formosa there are languages which normally use other words. In some of these *ət'a/it'a is, however, used when counting. Counting is done mechanically without reflection, and therefore an older form is easily maintained. But when the word is emphasized, stressing that a person or an object is single, alone, it may easily be supplanted by a word with this meaning.

This has occured for instance in Ma'anyan and Malagasy. Both in Ma'anyan and in the Merina dialect of Malagasy isa is used only when counting. The word is with s, whereas the ordinary reflexes of PAN *t' are Mny h, Mlg ø. But because the development has been *t' > s > h > ø, s is an older reflex, and may have been maintained by the mechanical and thereby conservative action of counting.

In other relations Mny generally uses eraŋ 'one'. In Mlg érana means 'to fill, pervade', e.g. man-érana ni tranu ni fúfuni 'its odour, fragrance fills, pervades the house', eran-tranu 'pervading the house, everywhere in the house', eran-tsutru 'one spoonful'. The last example shows how the meaning 'one' could develop. It also exists in archaic expressions where the meaning 'one' is clear: eran-aŋkihi 'the breadth of one thumb' = 'one inch'.

But the ordinary word for 'one' in Merina is iray. Northeast and Central east Barito too have erái 'one'. In Ma'anyan erái means 'alone'. The reduplicated form Mny raerái means 'alone, self, own'. In Malagasy the latter has developed into iréri 'alone' with another accentuation. It is thus clear that the meaning 'one' has developed from 'alone'. Similar semantic changes may have taken place in other languages.

2.

The word for 'two' is the same all over the Austronesian area. Dempwolff constructed it in the form PAN *ḍuva. In a special study of this word Dyen changed it into *Dewha (Dyen

1947a, 54), where *D* symbolises the same protophoneme as Dempwolff's *ḋ*. After having studied Formosan languages he changed the *h* into *S*, and writes it thereafter **DewS3a* (Dyen 1965b, 298-99; 1971, 42-43).

His starting point was the Tagalog cognate *dalawá*. He demonstrated that the syllable *da-* was the result of a reduplication of the initial consonant + *a*. Tag and many other languages have such reduplicated forms of numerals, e.g. Tag *ta-tlo* 'three'. The *d* and the *l* should therefore reflect the same consonant. From this he concluded that the Tag reflexes of PAN **D* are intervocalic *-l-* (like Dempwolff's construction), but initial *d-* (in opposition to Dempwolff). This theory of Dyen solved some problems in Dempwolff's constructions, but created others which he was not able to explain (Dyen 1947b).

Ogawa and Asai found reflexes of two *d*'s in Formosan languages, which they wrote PAN $*d_1$ and $*d_2$ (Ogawa and Asai 1935,6). They are not identical with Dempwolff's **d* and **ḋ*. I have found that in Formosan there is also a third set of reflexes from which I have constructed the protophoneme $*d_3$, and the word for 'two' has these reflexes. I have therefore reconstructed it PAN $*d_3uSa$. For more details see Dahl 1973/76, 55-62. PAN $*d_3$ has the reflexes *d-* and *-l-* in Tag, and makes most of Dyen's changes of the initial *d*'s in Dempwolff's constructions superfluous.

Tag *-aw-* is not a regular reflex of PAN **-u-*. It was to explain this irregularity that Dyen constructed the sequence **-ew-*. The normal reflex of PAN **ə* is Tag *i*, but Dyen found some instances with *a*. He admitted however that no other case with the sequence **-ew-* had been reconstructed before (Dyen 1947a, 51).

In a critique of Dyen's construction Blust has stated:

> In view of the fact that the reconstruction of **-ew-* is now seen to depend on an apparent irregularity in a single lexical item in Tagalog, Dyen's argument might be dismissed on general methodological grounds as insufficiently motivated. (Blust 1974a, 131).

I have given similar structural reasons against the construction of **-ew-* (Dahl 1973/76, 57).

Blust found further that the **ə* was not necessary because an earlier unstressed prevocalic *i* (or *u*) after a non-initial consonant frequently semivocalized in all of the Tagalic languages (Blust 1974a, 133).

Therefore a pre-Tag **darua* could develop into **daruá*, further into **darwá*, and in this form an *a*-epenthesis was possible. The result in modern Tag would be *dalawá* (Blust 1974a, 141-43). But he admitted that there were some difficulties connected with this complicated process.

My explanation is simpler. Another wordform from the same wordbase is Tag *ka-lu-luwá* 'spirit, soul' (the second ego). Here the reflex of PAN **u* is regular in the reduplicated stem. In earlier Tag **da-luwá* the **u* was both preceded and followed by *a*, and being assimilated to these *a*'s it became Tag *dalawá*. We have the same assimilation in Tag *lansaŋ-an* 'street' from PAN **laNk'u* 'stroll, promenade' + **-an*. (Dahl 1973/76, 56-58, Blust 1970: footnote 123).

Dyen has replied to Blust's critique that diphthongs as reflexes attributable to *ew* are found in seven Philippine and one Sulawesi language: Pandan *dawha*, Iriga Bikol *darawa*, Ilongot *d̈ewwá* (but Kakidugen Ilongot *dua*), Keleyqiq Kallahan *dewwaq*, Iraya *darawa*, *darwa*, Ata Manobo *dadawa*, Tagabili *l̈wuh*, Bolaang Mongondow *de(j)owa?*, *deowa*, *dojowa*, *dorowa*, *dua*, *dua?* (Dyen 1975, 1-2).

I do not find Dyen's argumentation convincing. We can oppose to Dyen's eight instances attributable to **-ew-* more than a hundred, known to me, deducible from **u*, and in about 90 per cent of these the reflex is *u* or *o*, normal reflexes of **u*. This brings Dyen to reflect upon double forms in the protolanguage (1975, 2-3).

Because of the overwhelming majority of languages with clear reflexes of PAN **u*, I find it more natural to consider the forms attributable to **əu* as secondary innovations. They have been found only within what has been considered as the Philippine subgroup, but even here most languages reflect **u*. If the abnormal reflexes are secondary, a plausible explanation is that a diphthongization has taken place.

This mostly occurs in accentuated syllables, but Ilongot *d̈ewwá* is accentuated on the ultima. Zorc assumes however that the ultima accentuation of numerals in Philippine languages may be due to counting intonation, and therefore not original (Zorc 1978, 31). 'Two' is a word often pronounced with emphasis, e.g. 'not one but *two*', '*two*, not three', 'there were *two* of them'. A dipthongization in this word in some few languages therefore seems more likely to me than the (at least) very infrequent sequence **ew/əu* without any differing reflex from that of **u* in the great majority of languages.

The languages cited are not close neighbours. A parallel diphthongization is therefore more probable than a common evolution.

As support for the presence of the sequence **-ew-* in PAN Dyen cites the word for 'star', in Pangasinan *bite:wen*, in Buhi (Bikol) *biẗʔwin*, and in Apayaw *bitiwan* from what has hitherto been constructed as PAN $*bi(N)t_1uq$ + *-ən* or *-an* (Dahl 1973/76, 28).[2] These forms should be deducible from **bitewqen* (Dyen 1975, 3-4). But in Tag the cognate is *bituín* with *u* and not with ***aw*. If we had **-ew-* in **DewS3a* and **bitewqen* we should have had the same reflex in both words in modern Tagalog. The cited words for 'star' are therefore an argument against the presence of *ew* in **d3uSa*. The hypothesis of assimilation is sufficient, and because of Tag *ka-lu-luwá* more probable. The irregular forms of $*bi(N)t_1uq$*-ən* in the three languages cited may also be due to diphthongization.

Dyen has constructed the second consonant in the word for 'two' as $*S_3$ (1965, 298-99). Except in Sedeq the reflexes of **S* in this word are the same as of $*S_1$. It is thus only Sed *daxá?* that needs a special explanation.[3]

Both Ferrell (1969, 408) and Tsuchida (1976, 153) write the Sedeq word with *x* (other authors have *daha*, *dahá?*). Both PAN $*S_2$ and **t'* have the reflex *x* in Sedeq, and that may

have influenced the development of the *$S$ in this word. The very frequent use of low numerals may account for irregularities.

In most Philippine languages the reflex of PAN *$S_1$ and *$S_2$ is *h*, and both Charles (1974, 468) and Zorc (1978, 31) reconstruct the PPh form of this word as **duha*. Some Philippine languages have more or less consistently lost this *h*, and so has also Tagalog in final position, but in intervocalic position Tagalog has generally kept the *h*. The loss of it in *dalawá* is thus irregular. But Aklanon, which generally maintains PPh **h* faithfully, and some other Bisayan languages have had metathesis in this word, Akl *dáywah* < **da-duha* (Zorc 1977, 101). The same has also occurred outside the Philippines. Southeast Barito has *rueh* 'two' (Hudson 1967, 76). If Tagalog has had the same, the loss of final *h* is regular. Dyen has recently suggested the same (1971, 43). After the metathesis the *w* has deloped as a glide after *u*.

Sedeq is thus the only language with an irregular reflex of *$S_1$ in this word, and the same irregularity occurs in no other Sedeq word. An irregularity in only one word in a single language is not a sufficient argument for constructing a special protophoneme. We can therefore reconstruct PAN *$d_3uS_1a$, and consider the *x* in Sedeq as an irregularity due to frequent use.

3.

For 'three' Dempwolff constructed PAN **təlu*. Ogawa and Asai also found two different *t*'s in Formosan languages, which they wrote PAN *$t_1$ and *$t_2$ (Ogawa and Asai 1935, 6). I use the same symbols. They are not identical with Dempwolff's **t* and **ṭ*. Dyen wrote them **t* and **ć*. In all Formosan languages the initial consonant of this word is a reflex of *$t_1$.

To this construction Tsuchida has added a final *$H_2$. The criterion for this is Ami *tulu*, Takitudu Bunun *tauh*, Sais *tuu?*, Paz *turu?*, Sed *təlu?* 'three'. The Aklanon cognate corroborates this reconstruction with *tá-tloh*. The original form of the word is thus PAN *$t_1$*ə*$luH_2$ (Tsuchida 1976, 135).

4.

For 'four' Dempwolff constructed PAN **ə(m)pat*. In the Formosan languages the reflex of the final consonant is that of *$t_1$.

Dyen has found that an initial *s* in some Formosan languages explains irregular reflexes of Dempwolff's initial *ə in some Philippine languages. In Tagalog the cognate is *?a:pat* and in Hiligaynon Bisayan *a:pat*, but in these languages the regular reflex of *ə is not *a*. Dyen gave the initial consonant the symbol $x_2$, and supposed the same model of reduplication as in Tag *da-lawá*. Then we would get the form PAN *$x_2a$-$x_2$*əpat*. If *$x_2$ was deleted in both positions, we would get **aəpat*, and with crasis of *aə* the result would be the modern *a:pat* (Dyen 1962, 215). Tagalog has developed hard onset to initial vowels, so the initial ? is no problem.

In Formosa some languages have a monosyllabic form **pat*$_1$, others a form derivable from the disyllabic *$S_1$*əpat*$_1$: SqAtay *paiat*, CiAtay *pazat* (with infixed *-ai-* or *-ia-*), Ruk dialects *patə*, Thao *pá:t*, Taok, Pap, Hny *pat*, Puy *pat*, Bun *paat*. Sed *spat*, Tsou *səptə*, Kan *u-súpatə*, Sar *upatə*, RukBudai *səpátə*, Fav *na-spaat*, Paz, Sais *səpat*, Paiw *s(ə)pat'*, Ami *s(ə)pat*, Kav *spat*, Siraya *hpat*, Ketagalan *spat* 'four' (Tsuchida 1976, 220. Ferrell 1969, 410).

Outside Formosa the eastern languages have generally the monosyllabic form, the western mostly the disyllabic. Which of them is the original one?

Javanese has the short form *pat*, and this has undergone the reduplication into *pa-pat* both in Old and New Javanese. The short form could then be the old one subjected to the ordinary reduplication. However, Javanese is the only MJ language having the short form. The others show reflexes of **ə(N)pat* (Nothofer 1975, 193). It is therefore clear that PMJ had the disyllabic form.

This form is so widespread that it is inconceivable that all these languages or subgroups should have independently developed the same syllable before a PAN **$pat_1$. We therefore have to consider the long form to be PAN. In all the Formosan languages which have the first syllable, the initial consonant is a reflex of *$S_1$ (see Tsuchida 1976, 159-60). I therefore reconstruct the word as PAN *$S_1$*əpat*$_1$.

How has it then happened that the initial syllable has been deleted independently in so many widespread languages?

Zorc has shown that in PPh the accent fell on the ultima, if the penult vowel was *ə (Zorc 1978, 70 and 91). In Malay it is the same, and it is likely that this phenomenon was PAN. The syllable *-*pat*$_1$ was therefore the strong syllable in the word. In rapid counting a weak syllable is sometimes omitted. In my mother-tongue Norwegian, 'four' is *fire* ['fi:rə]. But when counting rapidly we say: *en, to, tre, fir, fem* ... I assume the loss of the weak initial syllable in this word may be accounted for in the same way. We shall see the same when examining the word for 'six'.

Both in PPh and in PMJ the regular reflex of PAN *$S_1$ is *h*. But this word has ø reflex in both subgroups. I assume that the reason is similar. Low numerals have a very frequent use, and thereby a phoneme may be worn off. The *h* has a weaker articulation than any other consonant, and if that articulation is weakened even further, it is absorbed by the adjacent vowels.

Zorc has reconstructed this word with initial glottal stop in PBis (1977, 305). But Bisayan has developed hard onset to initial vowels. Therefore the glottal stop is generally not a reflex of a consonant in the protolanguage. First PPh **h* < PAN *$S_1$ has been deleted, and thereafter the hard onset has developed. In Aklanon it is the form with developed hard onset that has had the reduplication: **?a-?*ə*pat*. Thereafter it has undergone deletion of *ə and metathesis, and has thus acquired the modern form *?áp?at* (Zorc 1977, 101).

Up to 'four' the numerals are the same all over the AN area, with only insignificant exceptions. Beyond that numeral there is no

such uniformity. It is thus possible that at a certain point of the development of these languages, the Austronesians had numerals only up to 'four'.

5.

The most common word for 'five' found all over the AN area, is PAN *lima. To Dempwolff's reconstruction of this no alternative has been proposed, as far as I know.

In many languages *lima also means 'hand'. The reason for this is obvious. When we count on the fingers, 'five' is 'the whole hand'. It would therefore not be correct to consider *lima 'five' and *lima 'hand' as two homonym words. 'Five' and 'hand' are allosemes of the sameme.

However, in many modern languages the cognate of *lima means only 'five', and other words have been adopted for 'hand' to avoid ambiguity. In Ngaju Dayak and Fijian dissimilation of the nasal has eliminated the ambiguity: NgD *lime* 'five', *leŋe* 'hand, arm', Fi *lima* 'five', *liŋa-* 'hand, arm'.[4] Many other languages have PAN *taŋan 'arm, foreleg (of animals)' also with the meaning 'hand', thus the word for 'arm' used for the most important part of it.

Another word is Tag *kamây*, Ifugo Batad *ŋamay* 'hand', Ivatan *ka-kamay* 'finger' (Reid 1971, 88 and 81), perhaps also Tontemboan and Tombatu *kama* 'hand' (Sneddon 1970, 22). The expressions 'to do something with the hand' or 'to do it with the fingers' may designate exactly the same action. Here we may seek the semantic link between 'hand' and 'finger'.

Zorc has reconstructed PPh *ʔa-lí:ma[h] with long and stressed penult vowel (1978, 79). But in PBis *lǐmá 'five' is with short penult and stress on the ultima (Zorc 1977, 101). Zorc assumes however, as already mentioned, that the numerals received the pattern with stressed last syllable through counting intonation (Zorc 1978, 72 and 96). Both in this way and by the addition of an initial *ʔa* the original word with two allosemes has been split into two words: *(ʔa-)lí:ma* 'hand' and *lǐmá* 'five'.

PAN *li:ma is however not the only word for 'five' found in Formosan. Four northwestern neighbours have another word: Paz *xasə́b*, Sais *asəb*, Taok *hasap*, Fav *achab* 'five' (Ferrell 1969, 411). From these I reconstruct the protoform *$\gamma at_2 əb$. PAN *γ is the only protophoneme with the reflexes Paz *x* and Sais ø. Both *$t_2$ and *$S_1$ have the reflex *s* in Paz and Sais (Tsuchida 1976, 305-11). In Fav we have PAN *γ > ø, e.g. PAN *$baqəγuH_1$ > Fav *bao* 'new' (Ferrell 1969, 358), and the reflex of PAN *$t_2$ written *ch*, e.g. PAN *$mat_2ai$ > Fav *macha* 'to die', but that of PAN *$S_1$ written *s*, e.g. PAN *$iS_1əq$ > Fav *isi* 'urine' (Ferrell 1969, 281 and 244). Taokas is an extinct language of which very little is known.

It is possible that *$\gamma at_2 əb$ may have been an original PAN numeral for 'five', and that it has been supplanted by the word for the (whole) 'hand' because of counting habits. If so the four northwestern languages could represent the first group of AN immigrants in Formosa, and *li:ma have supplanted *$\gamma at_2 əb$ in the centre of the AN society shortly after their emigration. But much comparative study would be needed before this suggestion could amount to more than a very tentative working hypothesis.

Pazeh is the only AN language known to me that has a quinary numeral system.[5] The numbers from 6 through 9 are composed of *xasə́b* + the numerals 1, 2, 3, 4, linked with an *i* before a consonant (Ferrell 1969, 412-15). The three other languages have special numerals for these numbers, but not those of the majority of AN languages.

6.

The most common word for 'six' was constructed by Dempwolff as PAN *ənəm. Outside Formosa it has the same wide distribution as the preceding numerals. In Formosa too it is the most common.

In Tagalog and Hiligaynon Bisayan Dyen has found the same irregular *a* in the first syllable of this word as in PAN *$S_1 əpat_1$, and he has given the same explanation. The word must originally have had an initial consonant which he gave the hypothetic symbol $x_3$. This had been reduplicated with **a* like the other numerals, and thereafter been deleted. Finally **a* + *ə underwent crasis into *a*: PAN *$X_3a$-$x_3ənəm$ > *aənəm > a:nəm (Dyen 1962, 215). This hypothesis seems sound, but what was the real articulation of the initial consonant?

Dyen used the same symbol $x_3$ for the initial protophoneme of Paz *yakú* 'I' (Dyen ibid.). But this is the vowel PAN **i-*, which is the initial component of personal pronouns of 1st and 2nd person in many AN languages, not only in Formosa (Ferrell 1969, 186-91) but also outside, e.g. Mak *i-kau* 'you', *i-katte* 'we' (Mills 1975, 212). NgD *y-aku* 'I', *i-kau* 'you' (Hardeland 1858, 97). SakMlg *z-ahu* 'I', *i-ha* 'you', MerMlg *i-z-ahu* 'I' (Dahl 1951, 234-42). Sak *z-* is from a non-syllabic *i*, like NgD *y-*. Mer *-z-* reflects a glide between a syllabic *i* and a following vowel. Sangir has *i-* before all personal pronouns (Adriani 1893, 236). As we have seen, before a vowel this *i* often becomes non syllabic. It cannot be the same protophoneme in the word for 'six'.

Forms with initial *a-* are not the only irregularity occurring in this word. We find initial *u* in Kan *unə́mə* (Tsuchida 1976, 182), Paiw *unəm*, Puy *ʔunə́m*, Ami *unum* (besides *ʔnəm*), Kav *ʔu:nnə́m* (and *ʔnəm*) 'six' (Ferrell 1969, 412). I therefore reconstruct this numeral as PAN *uənəm (Dahl 1973/76, 47-49), and the reduplicated form of it *ua-uənəm. Here too Dyen's $x_3$ thus represents a non syllabic vowel like in **i-aku* (Dyen 1962, 215), but not the same vowel.

A non syllabic **u* has often been deleted (Dahl 1973/76, 50-52). This explains Dempwolff's construction *ənəm. In the cases above with *unəm*, *unum* it has not disappeared, but crasis with the following *ə has taken place *uə > u, and thereafter in some languages assimilation of the *ə of the following syllable.

As in PAN *$S_1 əpat_1$ the first syllable has been deleted in some languages both in and outside Formosa. I assume that here too the weak syllable has been lost in rapid counting.

PAN *uənəm is not the only word for 'six' in Formosan. The same languages which have *$\gamma at_2 əb$ 'five' also have other words for 'six'.

Pazeh has *xasə́b-uza?* '5+1' as already mentioned, Fav *na-taap*, Taok *tahap* (thus a common word), and Sais *bosi* 'six' (Ferrell 1969, 412). These are probably local creations.

The Atayalic subgroup and Thao have chosen another way to enlarge their numeral system. Sedeq has *ma-tə́ru?*, CiAtay *ma?-tu?* 'six' from Sed *tə́ru?*, CiAtay *tu-gal* 'three'. Here the prefix *ma(?)-* must indicate a doubling. SqAtay has *ciu-gal* 'three' and *c-z-iu* 'six' (Ferrell 1969, 409 and 412). The infixed *-z-* seems to have the same effect as the prefix *ma(?)-*. It is tempting to see in the suffixed *-gal* in Atayal the second syllable of PAN **tuNgal* 'single, alone', so that '3' is 'single three' and '6' 'double three'. But PAN **1* has been deleted in Atayal, as in Sq *ciu-*, Ci *tu-* < PAN $*t_1$*əluH*$_2$. Atay *1* reflects PAN **ɨ*. However, Ma'anyan has *toŋkan* 'alone' and MerMlg *tú̂kana* 'single, alone' with a final *n* which may reflect PAN **ɨ*. It is therefore possible that PAN **tuNgal* has had an alternative form with final **ɨ*. If so, Atay *-gal* may be a reflex of this form and mean 'single'.

Thao, Atayal's neighbour in the south, has *ka-tú:ru?* 'six' from *tú:ru?* 'three'. The prefixed *ka-* must also indicate a doubling.

Sundanese has *gənəp* for 'six'. This is cognate to Mal *gənap* 'full, complete'. The hand was full at 5, therefore this word is used to design the next numeral (Gonda 1975, 444).

7.

For 'seven' Dempwolff constructed PAN **pitu*. From the Formosan languages it is clear that the *t* is PAN $*t_1$. Madurese has *petto?*, *petto(h)* with geminate *t*, which generally indicates a short penult vowel. Whether this is PAN or the vowel has been shortened locally in Madurese through a counting intonation, is not clear.

In Formosa $*pit_1u$ is present in Atayalic with Sq *pitu?*, Ci *ma-pítu?*, Sed *?m-pitu?*. The word thus seems to have the prefix *ma-* here too. May it be from an analogy with 'six'? It is not possible to find any doubling effect of it here. -Thao has *pí:tu?*.

The only Formosan languages with other words for 'seven' are the four in northwest: Pazeh has *xasəb-i-dusá?* '5+2', Sais *yö?ähä?*, Taok *yweto*, Fav *(na)ito*. The Saisiat word seems composed from **yu + *?at'a?*. Both of the glottal consonants *?* and *h* change the articulation of adjacent **a* and **u*, **a* > *ä* and **u* > *ö* (Tsuchida 1976, 311). Sais *?ähä?* is 'one', thus from PAN **t'a* with a different initial vowel than the other languages, perhaps an assimilation of **ə* > *ä*. But I have no idea of what the initial *yö* < **iu* may be. A numeral composed with 'one' is strange here. Taok *yw-eto* may however be composed of the same elements. 'One' in Taokas is *ta:(nu)*, thus PAN **t'* > Taok *t*. The second element in Fav *(na)ito* may be the same, 'one' in Favorlang being *na-tta*. But the final *o* in the word for 'seven' is enigmatic (Ferrell 1969, 413 and 406).

There are thus not many discrepant words for 'seven' in Formosan, but in this word we also find discrepancies outside Formosa. Malay has *tud'uh*, Mak *tud'u*, and there are cognates of the same in several languages of North Sarawak (Blust 1974b, 19 and 38) and in Nabay Murut *turu?* (Prentice 1971, 280), in Tundjung *tujuu* (Hudson 1967, 90).

Dyen has found the etymology of this word in PAN $*t_2u(N)Zuq$ 'to point' (1951, 539). This is certainly correct. Both in Madagascar and Formosa I have seen people, when counting, begin with the little finger of the left hand, and after 5, 'the whole hand', countinue with 'six' on the thumb of the right hand. 'Seven' is then on the forefinger of the right hand, the finger used to point with. If the same practice in counting is also used in the islands in between, the connection between 'seven' and 'to point' becomes sure. The same etymology of 'seven' is found in some Bantu languages (See Meinhof 1906, 59).

The fact that modern Malay has another word (probably from the same etymon) for 'forefinger': *d'ari t-əl-uñd'u?*, does not pose any problem. We have seen that when **lima* 'hand' is used for 'five', other words have often replaced **lima* in the meaning 'hand'. So also here, when *tud'uh* is used for 'seven', another form has supplanted it in the meaning 'forefinger'.

In Malay *tud'uh* is an innovation, because all the other MJ languages have cognates of PAN $*pít_1u$.[6] In other languages where *tud'u(h)* is found, it may be a loan from Malay. But because it is caused by a common habit in counting, it may also have been created independently in more languages. In the East Barito languages Tabojan, Lawangan and Dusun Dejah it is at any rate not a loan. Here we have *turu*, *toru* 'seven' with the regular *r* < PAN **Z* (Hudson 1967, 76). So also in Nabay Murut *turu?* above.

Other discrepant forms also occur in some languages, but probably as local innovations.

8.

For 'eight' Dempwolff constructed PAN **valu[]*. He describes his **v* as a voiced bilabial fricative (1934, 59). Dyen has given it a subnumerated *w*, $*w_2alu$, because of different reflexes in some Formosan languages. He has also a $*w_1$ and a **W* (1965b, 301).

I have proposed another explanation of these discrepancies, viz. that PAN /u/ could be realized as syllabic or non syllabic according to its place in a syllable centre or at a syllable limit (1973/76, 17-18 and 50-52). Before another vowel a non syllabic *u* has often been deleted, as we have seen in PAN **uənəm* 'six'.[7]

With syllabic *u* I have found Kav *uwaló?* (besides *va:rú?*, *waru*) (Ferrell 1969, 414), TB *uwalu* (v.d.Tuuk 1867, 217), *ualu* (Warneck 1977, 279), Bar *uayu* (regular), *ualu* (archaic), *wayu-* (in composita) (Adriani 1928, 489). The glide *-w-* in Kavalan and Toba Batak (by v.d. Tuuk) shows clearly that the *u-* is syllabic. The form *wayu-* in Bare'e is used only as the first component of compound numerals, and is thus further away from the accentuated syllable than in forms with *u-*. It thus loses syllabicity more easily. The initial *u* in Sar *ualu* (Tsuchida 1976, 146) is not neces-

sarily a reflex of this *u, as this language may add an initial u to most numerals.

Moreover, outside Formosa I have found it with u- in Ibanag *ualu*, Sangir *ualu*, Tondano *ualu*, Tonsea *uadu*, from the two last languages constructed as PMin **ualu* by Sneddon (1978, 128).

The following Formosan languages have lost the initial *u: Kan *a-alu*, Paiw *alu* (Tsuchida 1976, 146), Hny *(mi)alu* (Ferrell 1969, 414). It has also been lost in Muna (south of Sulawesi) *alu* (Adriani 1914, 259).

Most of the languages both in and outside Formosa have a consonantal reflex: *w*, *v* or *f* (Ami *falô?*).

By regular development Old Javanese *wwalu* has become New Jav *wòlu*. The first *w* may be syllabic and the second a glide. But I think it more likely that it is a contraction from a reduplicated form **wa-walu* having lost the first *a*. Madurese has *ballû?*, *ballu(h)* with geminate *l*, which indicates a short penult vowel, as we have seen in the word for 'seven'. Here too the question is open whether this is PAN or a local development.

But there are also words for 'eight' with another etymology both in and outside Formosa.

Ciuli, Sedeq and Thao have forms of $*S_1$*əpat*$_1$ 'four' with the same prefixes as the words for 'six' from $*t_1$*əluH*$_2$ 'three': Ci, Sed *ma-spât*, Thao *ka-spat*. In the interior of the word Ciuli has conserved the $*S_1$, which it has lost in initial position in *p-az-at* 'four'. Squliq has *spat*, without *ma-* (as in the word for 'six'), but also with initial *s* like Ciuli *ma-spat*; compare Squliq *p-ai-at* 'four'.

Here the doubling in the sense in the word for 'four' with *ma-* is used also in Fav *maa-spat* 'eight'. Saisiat has both *ma-* and *ka-* linked with *-i-*:*ma-y-ka-spat*; compare the linking *-i-* in Paz *xasəb-i-turû?* 'eight'. Taokas has *ma-hal-pat*, which also must be a doubling of the meaning in Taok *pat* 'four'. Papora (between Taok and Fav) has *ma-hal*, but I have no explanation of the *hal* in these languages. All these languages are or have been western neighbours of Thao and Atayal. We thus find these doubled forms in a restricted geographic area (Ferrell 1969, 414).

Although PAN **ualu* is the most common word for 'eight' outside Formosa, there are words for 'eight' with other etymologies in Indonesia too. Makassarese has *sagantud'u* from *sa-agáŋ-tud'u* 'one with seven' (Matthes 1858, 50). Instead of adding to 7 Buginese has subtracted from 10 with *a-ruâ* 'eight' < **ka-*[*d*]*ua* 'the second (from ten)' (Mills 1975, 89 and 686). So has also Malay in *delapan* and Sundanese in *dalapan* 'eight' from **dua-alap-an* with reduction of *dua-a-* into *də/da-* before the accentuated syllable and in rapid counting. PAN and Mal, Sund *alap* means 'to take away'. The literal meaning of the word is thus 'two taken away (from ten)' (Gonda 1975, 442).

9.

For 'nine' Dempwolff constructed PAN **t'iva*. Dyen writes it **siwa*, and I **t'iua*. In the languages outside Formosa the cognates to this word have reflexes of PAN **t'*, mostly *s*. Even in Ma'anyan where **t'* generally has become *h*, this word has *s*: *suei*, and likewise in Malagasy, where **t'* has often been deleted, we have Mer *sivi*, Sak *sive*.

In Formosa we find the following cognates to this word: Tsou *sío*, Kan *si:ya*, Sar *ku-sia/u-sia*, Pap *(me)siya*, Hny *(a)sia*, Paiw *síva*, Puy *iwá*, Ami *siwá?*, Bun *síva?*, Kav *siwá?*, Ketagalan *siwa* 'nine' (Ferrell 1969, 415). Among these languages only Tsou, Puy, Bun and Kav have *s* as regular reflex of PAN **t'*. Kan and Sar have *s* as reflex of Tsuchida's *θ. In the other languages **t'* and *θ have merged. Tsuchida therefore constructs PAN $*\theta_1 iw_1 a$, with subnumerated symbols because of irregular reflexes (Tsuchida 1976, 129-31).

In the other languages cited PAN **t'* and *θ have the following reflexes: Pap, Hny *t* (e.g. *ta(nu)* 'one' < PAN **t'a)*, Paiw *t*, Ami, Ketagalan *ts* (e.g. Ketagalan *tsa* 'one').

However, in all these languages except Sar the initial *s* in this word corresponds perfectly with PAN $*S_1$, and so also Puy ø-. See the reflexes in the following examples: PAN $*S_1$*apaγ*, Tsou *s-m-opro*, Kan *s-um-sapárə*, Bun *ma-pa-sapal* 'to lay mats'. PAN *luS*$_1$*əq*, Paiw *lusəq*, Ami *lusaq*, Puy *urəh* (metathesis from **ruəh*) 'tears' (Tsuchida 1976, 235 and 164). PAN *bukəS*$_1$, Kav *buqəs* 'hair' (Ferrell 1969, 214). I have not been able to determine the reflexes of $*S_1$ in Papora, Hoanya and Ketagalan as our knowledge of these languages is too limited.

In all the Formosan languages with *s* < **t'* this *s* has merged with *s* < $*S_1$. It has probably been the same in Saaroa too, because here both **t'* and $*S_1$ have become ø. But **t'iua* has *s* in Sar *-sia*. Puyuma has **t'* > *s* and $*S_1$ > ø and ø in *iwá?*. Outside Formosa PAN $*S_1$ has become *h* or ø. Provisionally putting Saaroa aside, we then could say that in Formosa the word for 'nine' is from PAN $*S_1$*iua*, outside Formosa from PAN **t'iua*.

It is evident that this is an impossibility. In one of the groups the word must be a loan from the other. But which is the borrowing and which is the lending group?

Whatever the protophoneme, it must have been realised as [s] at the time of borrowing. It seems to me that the irregularity in Saaroa is best explained if we suppose that the Formosans have been the borrowers. It must then have taken place after the change PAN **t'* > *s* outside Formosa, but before the change PAN $*S_1$ > Puy ø. The Formosans have then identified the *s* of **siua* with their *s* < $*S_1$. But the change PAN $*S_1$ > Sar ø must have taken place before. Therefore Saaroa could not identify the *s* with this phoneme. This language has, however, an *s* < PAN *θ and has identified the *s* of **siua* with this phoneme. The borrowing must then have taken place after the changes PAN **t'* > Sar ø and PAN *θ > Sar *s*. We thereby get a relative chronology of these changes in Puyuma and Saaroa.

We have already seen that the higher the numbers the more non cognate numerals we find in Formosa. We shall see in a moment that for 'nine' still more languages have non cognate words than for 'eight'. This is also an argument for considering Formosa as the borrowing and not the lending part.

But before considering these words we shall examine other irregularities in **t'iua*. In some

languages the non syllabic *u has been deleted. Dyen and Tsuchida use the symbol $*w_1$ in such cases. In Kan and Sar the deletion is regular. In Tsou it is not (Tsuchida 1976, 144-46). It is also lost in Papora and Hoanya, if regularly or not we do not know. The deletion of *u also occurs outside Formosa, e.g. in TB *sia* 'nine'.

In other languages the *i has disappeared, e.g. Gilbert *rua*, the southeast Barito languages in Kalimantan *suei* 'nine'. The latter have added a final *i*, but this is not due to a metathesis. We shall see that Malagasy, which belongs to this subgroup, has had a final *i* in *sivi* 'nine'. I suppose that both when *u and *i have been lost, they have first become non syllabic before a vowel, and then disappeared having a weaker articulation (Dahl 1973/76, 17-18 and 50-52).

In Formosa more languages have non cognate words for 'nine' than for the lower numerals: Atayal, Sedeq and Rukai, which hitherto showed words with PAN etymology, here have local words. So also Thao *tana:θu?* 'nine', but Fav *tannacho* and Taok *tanasu* id. are cognate to this word, and we may reconstruct Proto-northwest-Formosan $*t_1anat_2u$ (Ferrell 1969, 415).

The PBis word for 'nine' is *siyám (Zorc 1977, 101). Compared with e.g. Tag *siyám*, Yami *siam* (Ferrell 1969, 415) this gives PPh *siyám (Zorc 1978, 30). The word also occurs in Sabah: Murut *siam* 'nine' (Prentice 1971, 172). It might be possible to consider this to be a final *m* added to PAN *t'iua with loss of *u and in some languages development of a glide. But as long as no explanation of the *-m* is found, it is more prudent to consider the word as a Philippine creation.

For 'nine' too we find words expressing subtraction from 'ten'. Makassarese has *salapaŋ*, Sundanese *salapan*, from *sa* 'one' prefixed to the same *alapan* which we found in the Malay and Sundanese word for 'eight', thus 'one taken away (from ten)'.

Matthes found the etymology of this in a Malagasy word, *ma-lafa* 'to take'. I have never met this word, and do not think it exists. Matthes has taken it from Humboldt's Kawisprache (Matthes 1858, 50-51). But Humboldt's source for Malagasy was a vocabulary written by the English missionary John Jeffreys, who used the vowels with their English pronunciation. Humboldt's *mi-lafa* is therefore probably Mlg *mi-lefa* 'to escape, flee, run away' < PAN *lə(N)pat', and gives no explanation to *salapan*.

Malay has *sambilan* from *sa* + *ambil* + *an*. As Mal *ambil* too means 'to take away', the formation and meaning of the word is exactly the same as in *salapan* (Matthes 1858, 50-51). Achenese has *sa-kuraŋ* 'nine' from *kuraŋ* 'to diminish, lessen' < PAN *kuraŋ, also the same idea (Gonda 1975, 442).

Tungho Sais *ää?hä?* 'nine' is strange compared with *?ähä?* 'one'. It could be a metathesis of ? and ä. But the other Saisiat dialect, Taai, has *Lää?hä?* id. (Li 1978, 196). Tungho has deleted Taai *L* regularly (a voiced lateral flap < PAN *l or *γ). If a composed word, *?ähä?* 'one' is preceded by an unknown *Lä*. But it may also be an independent word.

It seems as Rade has gone the opposite way of Malay, using addition instead of subtraction. 'Nine' is *dua-pan* whereas 'eight' is *sa-pan*. These words give the impression that *dua* 'two' and *sa* 'one' have been added to a word for 'seven', but 'seven' is not *pan*, it is *kjuh*. Up to six Rade has the ordinary numerals. This *kjuh* may be a local creation, but it is also possible that it reflects the last syllable of $*pit_1u$. There are cases of *t* > Rade *k-*, e.g. PAN $*t_1a\cancel{l}ək$ > Rade *kna?* 'to cook', PAN $*t_2u(N)Zuq$ > Rade *ktro* 'to point'. As to *pan*, Egerod suggests however in private communication that *delapan* and *salapan* may have been borrowed into Rade, and thereafter reduced to monosyllabism like most words in this language. The remaining *pan* then lost the meaning "taken from". In order not to get homophone words for 8 and 9, the words for 'one' and 'two' have been added so that the words mean *pan* no. 1 = 8, and *pan* no. 2 = 9. This seems a plausible explanation.

Till now I have neglected the extinct language to the north of Saisiat, Luilang, because only the three first numerals in this language have clear AN cognates: *sa(ka)* 'one', *tsusa* 'two', *tulu* 'three'. But 'four' *suva* is astonishingly like *t'iua 'nine' with loss of the *i (Ferrell 1969, 407-10).

This evokes the following reflections: Is *t'iua too a subtractive term, meaning originally 'one subtracted from the higher unity'? In a decimal system the higher unity is 10, in a quinary it is 5. If Luilang originally had a quinary system, this might be an explanation. But our knowledge of Luilang is far too limited to offer any hope of a reliable answer to our question. Moreover, the fact that *t'iua must be a loan-word in Formosan, makes the comparison of *suva* with *t'iua dubious.

10.

The word for 'ten' was constructed as PAN *puluh by Dempwolff, and corrected with good reasons by Dyen into *puluq (Dyen 1953, 28). This word has a very wide distribution outside Formosa, often preceded by *sə-*, *sa-* or *sam-*, thus 'one unit of ten', in the last case bound to *puluq with a nasal.

In Formosa we have it only in Ruk *porók*, Paiw *púluq*, Puy *pulúq*, Ami *polo?*, SqAtay *m-puu*, and probably also in Ciuli *mala-pzo* 'ten'.

In northwestern Formosa there is another etymon: Paz *?isít*, Taok *(ta)isid*, Luilang *isit*, Fav *zchiett*, Hny *(miata)isi*, perhaps also Pap *(me)tsi* 'ten'. From these I tentatively construct $*it_2id$.

A third word is found in central Formosa: Sed *máhal*, Bun *más?an*, Tsou *maskə*, Kan *ma:nə*, Sar *(ku)ma:ɬə* 'ten'. All these have regular reflexes of *mat'aɬ. Papora has *mahal* 'eight', and Taok *mahal-pat* id. (Ferrell 1969, 416 and 414).

From 'eleven' through 'nineteen' the numerals are composed of 'ten' + the numerals from 'one' to 'nine', with or without a link in between. The links differ very much, so that none of them seem to be PAN.

20.

The numerals from 20 to 90 are generally formed with *puluq preceded by the numerals from 2 to 9. A nasal often links the two components, and before *puluq the nasal is very often m. But this is probably due to an assimilation, and not original. In Sangir we have n (Adriani 1893, 228), in Chamorro na (Costenoble 1940, 261). Gorontalo has lo, but here is l < PAN *n and o < PAN *ə, e.g. Gorontalo olomo 'six' (Breukink 1906, 78). Tagalog has m after vowels, na after consonants, originally it may have been ŋ (Ramos 1971, 39).

In Malagasy PAN *p has become f, except when preceded by another consonant; then the p has been maintained. Mlg fulu 'ten' is thus regular, but Mérina has rua-pulu 'twenty'. The reason is that PMlg had the linking m, but the cluster mp has later been reduced to p in Merina. In Sakalava the m is still present: ruam-pulu. But after a PMlg final consonant the m is not used, therefore both dialects have efa-pulu 'forty' < PMlg *epat-pulu. Here the assimilation *t > p maintains the p. However, in Merina 'ninety' is sivi-fulu with f. As the m was lacking only after a preceding consonant, sivi must have been a word ending in a consonant. In Kalimantan the southeast Barito languages have suei 'nine' (Hudson 1967, 76). The i of this dipthong must thus have been treated as a consonant, and we may write the PMlg form *siwey. But only original (final) stops were totally assimilated to the following consonant and hindered the change *p > Mlg f, like in efa-pulu. Therefore the semivowel *y could not influence the following *p, which therefore became Mlg f, and *ey became Mer i. Because Malagasy has not lost the *i of the first syllable, and therefore earlier southeastern Barito had this i, it is clear that the final i in SEBarito suei is not due to a metathesis from the first syllable, but it has been added to the original form. However, the word has s- even in the languages which have changed s > h. It is therefore possible that it is a loanword here too, like in Formosa. Another possibility is that the s has been maintained by mechanical counting.

In Formosa only two languages have 'twenty' in this form: Paiw dusa-puluq, Ami tusa?-polo?. Favorlang and Pazeh have the same construction with their own word for 'ten': Fav na-rua-tsxiet, Paz dusa?-isít 'twenty'. Pazeh has thus its quinary system only up to 9, and from 10 on a decimal system.

But many Formosan languages have a special word for 'twenty': SqAtay pusal, CiAtay ma-pusal, Sed ?m-púsal, Tsou m-púsku, Kan ma-pusauu, Sar ma-poaɮə, BudaiRuk ma-pusál, MagaRuk mama-pualu, Thao ma-pu:saš, Bun ma-púsan (Ferrel 1969, 417), all cognates of proto-Formosan *ma-puS$_1$aɮ. I am not able to deduce this from any other AN numeral. In other language families too the word for 'twenty' is not always recognisable from the model of the higher numerals, e.g. Greek eikosi, Latin viginti.

100.

For 'hundred' Dempwolff constructed PAN *γatut'. Dyen writes it *Ratus. It has a very wide distribution in Austronesian languages, but is not found in any Formosan language.

1000.

For 'thousand' Dempwolff constructed PAN *ḷibu. Dyen uses r instead of ḷ for the same protophoneme, and I do the same (Dahl 1973/76, 86). It has a very wide distribution in western languages, but as far as I know not in eastern AN, if Motu gerebu 'ten thousand' is not cognate (Dempwolff 1924-25, 302).

Most of the languages show clear reflexes of PAN *r, but old Javanese has iwu and modern Javanese ewu 'thousand'. As far as I know Dempwolff has never explained the loss of *r in this word. Wolff, who does not accept the protophoneme *r, constructs *γibu, which has then regular reflexes both in Javanese and in Mal ribu. The presence of what he considers as an irregular r in Malagasy and many other languages is considered as due to loans from Malay (Wolff 1974, 80).

I do not accept the exclusion of *r from the PAN phoneme inventory. It has special reflexes in many languages, although it is admittedly infrequent. But in this word the construction of *γ is corroborated from several languages in central and southern Sulawesi. Typical is Bar sowu 'thousand' < PAN *t'a 'one' + *γibu. After the regular deletion of *γ crasis of *a + *i has produced *ə, and this has the reflex o in Bare'e. We find similar forms in surrounding languages: Parigi sobu (Adriani 1928, 773), Besoa, Bada sabu, Lalaki sowu, Mekonka sobu 'thousand' (Adriani and Kruijt 1914, 129; 230; 236). These languages have ø reflexes of *γ, but not consistently. It is therefore possible that they have borrowed the word from Bare'e when needing higher numerals.

For PSS Mills has constructed *sɨbbu from Bug sɨbbu, Mak, Sad etc. sa?bu 'thousand' (1975, 820). (Mills' ɨ represents the same phoneme as Dempwolff's ə). The word can therefore not be directly inherited from *t'a-γibu, but it could be a loan from the Toraja languages. It is not likely, however, that a high numeral should be borrowed into Buginese and Makassarese from a Toraja language. Mills has suggested that the South Sulawesi languages may be immigrant languages on a Toraja substratum, and that a major motive of the immigration was trade (1975, 499-500; 513 ff.). This may explain the presence of the word for 'thousand' in an old Toraja (Bare'e) form. The first words traders learn in a foreign society are the numerals.

As a parallel to the Sulawesi forms deducible from *t'a-γibu I cite Mlg a-ribu 'thousand' with regular loss of *i', but irregular r < *γ, the latter due to old borrowing, probably from Malay.

We also find this borrowing from Malay in South Sulawesi. Mills has constructed PSS *riwu$_1$ 'hundred thousand' from Bug riwu, Mak si-riu, Sad sa-riu id., but he adds: "perhaps borrowed < Mal ribu 'thousand'" (1975, 682). Bare'e has riwu 'ten thousand'. I consider both the Bare'e and the South Sulawesi forms as borrowings from Malay, adapted to the local

phonology (*b* > *w* or *ø*). The semantic changes are due to the fact that the languages had already an "inherited" word for 'thousand'. The borrowed word was then used for the higher unity for which the language had no word before.

Two Formosan languages have a numeral which may be cognate to **γibu*: Kav *γasi:buʔ* and Ketagalan *latsibu*. Here it is possible to see a metathesis from **t'a-γibu*. In both languages the meaning of the word is however not 'thousand' but 'hundred' (Ferrell 1969, 418). The original meaning of the word may have been 'an extraordinary high indefinite number'. An argument for this is that in Sad the word *sa-riu* in addition to 'hundred thousand' also means 'any uncountable large quantity' (Mills 1975, 682). When the need for higher numerals than before developed, the meaning became 'hundred' in northern Formosa, and 'thousand' in the MJ subgroup and southern Sulawesi, trading societies with a higher culture. From the trading Malayans *ribu* 'thousand' has been borrowed into many languages where *r* is not the regular reflex of PAN **γ*, and as we have seen, sometimes with change of the numeral value.

## Interrogative

The interrogative 'how many' also belongs to the numeral category and has often the same affixes as the numerals. Dempwolff constructed it as PAN **pig'a*, by Dyen written **pija*. Tsuchida has added a final $*H_2$ to this because of TakituduhBun *piah*, Ami *pina*, Sais *pizaʔ*, Akl *pilah* 'how manyʔ' (Tsuchida 1976, 135 and 224).

It has a very broad distribution both in and outside Formosa with the meaning 'how many'. In some languages it also means 'some, a few', and in some languages, especially in Toraja, the meaning has changed into 'when?'.

## Evolution of the numerals

We have seen that the same words for 'two, three, four' are used in virtually all AN languages. PAN **t'a* 'one' is lacking only in a few languages, and in some it is used only in counting. Still it seems clear that **t'a* must have belonged to the original vocabulary of PAN, and that in some languages it has been supplanted by words meaning 'alone' etc.

From 'five' on different words begin to occur in Formosan, and from 'seven' on in the western languages. The words for 'seven' from the wordbase 'to point' and the subtractive words for 'eight' and 'nine' are probably relatively new words having supplanted the common words $*pit_1u$, **ualu* and **t'iua*, found also in eastern AN. Within the MJ subgroup Javanese and Madurese have cognates to **ualu*, and both these and Sundanese also to $*pit_1u$. In the period of common MJ the subgroup has thus had both these words. It is not so with the word for 'nine'. But as the local words for 'nine' are formed on the same principle as those for 'eight', I assume that they date from the same period. Similar words in other western languages are probably also local innovations, if they are not borrowed from Malay.

Not so in Formosan. We have seen that from 'five' on, the higher the number the fewer the cognates of the common AN words, and the more variation of words used in the different languages. At any rate some of these must be local creations.

A special category of numerals contains those formed with *ma-* with doubling effect on the meaning. They may depend on a special way of using the fingers in counting. Carl Meinhof has described how such doubling of the numerals in some Bantu languages is shown with the hands. When counting 'six', the thumb and the little finger of both hands are bowed under the hands; thus three fingers of each hand are shown. For 'eight' only the thumbs are bowed beneath, and four fingers of each hand are shown (Meinhof 1906, 58-59. See also Gonda 1975, 443). It will be interesting if the same use of the fingers is found in the Formosan societies which have this system (Atayalic, Thao, and Saisiat).

For 'seven' one little finger is released, and for 'nine' one thumb put forth. Atayalic uses the prefix *m(a)-* before 'seven' and 'nine' too, but without any doubling of the meaning. This may be due to such showing of the fingers when counting.

Sedeq, Tsouic and Bunun have formed 'ten' with **ma-* + **t'aɬ*. It is possible that this **t'aɬ* is an old word for 'five', or that with **ma-* it expresses the two complete hands.

I have considered the Formosan languages to represent the first offshoot from the main AN family (Dahl 1973, 125/1976, 126). The study of the numerals offers further support for this hypothesis.

The numerals from 'one' to 'hundred' are so common both in west and east that they must have developed in the main centre of the AN family. But only those from 'one' to 'four' are found in all Formosan languages (only 1 to 3 in Luilang). Before the migration to Formosa began, the AN society thus seems to have had so little need of numbers that counting on only one hand was sufficient.

The differences in the presence of the common numerals from 'five' to 'ten' in the Formosan languages may be explained by several migrations during the time of development of these numerals in the AN centre, or by more or less vivid communication with this centre from the different Formosan groups, or by both. We have seen that at any rate **siua* 'nine' has been borrowed into the Formosan languages which have the word.

This seems to indicate rapid cultural growth in the AN centre during and shortly after the migrations to Formosa, a growth with increased need for numerals. The reason may have been navigation and trade, precursory to the great spread of the Austronesians.

Blust has found that

> All non-Formosan AN languages have undergone certain changes in the original pronominal system not known to be shared by any AN language of Formosa. In view of the fact that the non-Formosan AN languages in aggregate correspond closely to the collection of languages to which the term 'Malayo-Polynesian' was originally applied

by Humboldt, it seems appropriate to revive this term as the designation of the non-Formosan group (Blust 1978, 219).

The different development of the numerals in and outside Formosa gives support to Blust's consideration of all non-Formosan (MP) languages as a single first-order subgroup (Blust ibid., 220).

Dyen assumes that the Formosan languages together with the western AN languages form one first-order subgroup which he names Hesperonesian. He contrasts this with a first-order subgroup in the east which he calls Heonesian. He bases his findings on lexicostatistics (Dyen 1965a). The present study as well as Blust's examination of the pronouns refute Dyen's hypothesis. Such subdivision cannot be based on the evidence of vocabulary alone. Grammar and phonetic development are at least as important considerations.

The word for 'nine' seems to be not a PAN but a PMP word, formed as **siua* after the change PAN **t'* > PMP **s*, and borrowed into some Formosan languages (perhaps also into SE Barito).

The question of whether the Formosan languages form one or more first-order subgroups cannot yet be settled. It is clear from several facts that the Atayalic and Tsouic subgroups exist, as first-order or as secondary under a first-order Formosan subgroup. But Ferrells joining of all the other Formosan languages into one Paiwanic subgroup seems dubious (Ferrel 1969). The northwestern languages Pazeh and Saisiat together with the surrounding more or less extinct languages have a development of the numerals different from all the others. If we suppose more migrations into Formosa, to judge only from the numerals, these languages could represent the first wave. Paiwan, Rukai, Puyuma and Ami (with cognates to **puluq*) could then be a last wave, with the other languages in between. But this last inference is not reliable, since we have seen that **siua* 'nine' has been borrowed into these languages; **puluq* may equally be a loanword. Alternatively they may have had later communication with the MP languages. Rukai has *porók* 'ten', but *ma-pusál* 'twenty'. Is **puS$_1$aɬ* an older word for 'ten', having acquired the meaning 'twenty' from the doubling **ma-* ?

It is not possible to base a subgrouping hypothesis on the development of the numerals only. A thorough examination of phonetic development, grammatical structure, and vocabulary is needed to decide the issue. The results of my study thus only make Ferrell's subgrouping questionable. A final answer lies in the future.

## Derivatives

We have already met one derivative of the numerals with such wide distribution that it must be old. It is the reduplication of the first consonant + *a*.

We have seen that it is very common in Philippine languages. Blust has shown that two series, simple and with reduplication, exist in Ivatan and also in Balinese (Blust 1974, 136). Dyen has found that in Chamorro the reduplicated forms are used only to designate animates. In Ami, Kanakanabu and Saaroa they are restricted to persons (Dyen 1975, 4). I have found the same in Bunun.

In the Sakalava dialect of Malagasy only *rue* 'two' and *telu* 'three' are reduplicated, but not with *a*, the whole first syllable is repeated. They are used only as attributes to personal pronouns designating persons, e.g. *zahay ru-rue* 'we (excl.) two', *ruze te-telu* 'they (being) three'. The restriction to persons must then be very old.

In Javanese too we find the repetition of the consonant with its vowel in *loro* 'two'. The development must have been: PAN **d$_3$uS$_1$a* > **ruá*, then with loss of syllabicity of the **u* > old Jav *rwa* > **ro*, thereafter reduplication of this > **ro-ro*, then regular dissimilation of consecutive *r*'s > modern Jav *loro*. Javanese has also *pa-pat*, reduplicated from the new form *pat* 'four'.

The reduplication of the numerals to designate numbers of persons or animates in general must be considered as PAN, but new formations after the old pattern also occur.

A very common formative of ordinals is the prefix **ka-*, e.g. Mny *ka-*, Iloko *ka-*, Mal *kə-*, Sangir *ka-* (Adriani 1893, 80), PMin **ka-* (Sneddon 1978, 104), Ledo *ka-*, Mak *(ma)ka-* (Matthes 1858, 57). Toba Batak has *pa-* or *pa-hi-* (v.d. Tuuk 1867, 131; 134; 191). Malagasy has a combination of both these prefixes: *fa-ha-*. Tagalog and other Philippine languages have another combination: *i-ka-*, which seems identical with Paiw *si-ka-*. Tag *ka-lu-luwá* 'soul' is probably formed with the same *ka-*, thus maning 'the second something'. The reduplication of *lu* implies vagueness. The soul belongs to a realm where vague expressions are prudent.

In other languages the ordinals have the same form as the cardinals, the syntactic rules are sufficient to distinguish between the two. In some languages cardinals precede the noun, whereas ordinals follow it. As in other fields of comparative grammar our knowledge is still too limited to get a full picture.

Gonda has shown that the word for 'first' is generally not an ordinal, but a word designating 'the foremost, that which is before, ahead, in time or place'. It is in opposition not to the second and the third, but to the last (Gonda 1975, 244-54. This article has also other interesting viewpoints). 'The first' is thus not a numeral, but an adjective, often in superlative, if the language has a superlative form. In AN languages too we find that 'the first' has generally an "irregular" form compared with the ordinals.

There are also other derivatives of the numerals, but their formatives vary very much. It is therefore not possible for the time being to determine PAN or PMP forms of them. I shall only mention some of them as an incentive to further research.

Iteratives, indicating so and so many times, have the prefix *haŋ-* in Ma'anyan, and in Ngadju Dayak *haŋ(-ka)-*, e.g. Mny, NgD *han-telo* 'three times', Mny *haŋ-epat* 'four times', NgD *haŋ-ka-lime* 'five times'. In 'once' the prefix has the vowel *i*, and in

NgD the consonant *s*-: Mny *hin-ra*, NgD *sin-da*. In Malagasy it is *in/iŋ*-, e.g. Mer *in-dray* 'once', *in-telu* 'three times', Sak *iŋ-éfatse* 'four times'. The protoform of this prefix must be **t'iŋ*-. The **i* is still present in Malagasy, and in Mny *hin-ra*, NgD *sin-da* it is also maintained because here it is in the penult and therefore accentuated. In the other forms it stands before the accentuated syllable, and there most vowels become *a*. Malagasy has not the same rule. Makassarese has *piŋ*-, e.g. *pin-tallu* 'three times', *piŋ-áppaʔ* 'four times' (Matthes 1858, 56). Chamorro uses *faha*- with the same function (Costenoble 1940, 263). But many languages use the cardinals followed by a word meaning 'times', e.g. Mal *sə-kali* 'once'.

Malagasy, Ma'anyan and Paiwan have forms meaning 'so and so many days', formed with Mlg *ha*- + -*ana*, Mny *ka*- + -*an*, Paiw *maka*- + -*ɮ*, e.g. Mlg *ha-telû-ana*, Mny *ka-telo-an*, Paiw *maka-t'əlu-ɮ* 'three days'. As it is found with nearly identical formatives both in Formosa and Madagascar, the form must be very old. The protoform seems to be **(ma)ka*- + -*(a)ɮ*. Chamorro has the suffix -*a* in the same function, probably not identical with this *-*ɮ*, e.g. *fat-a* 'four days', *limi-y-a* 'five days' (Costenoble 1940, 264).

Distributives meaning 'so and so many each time' are formed in Malagasy with the prefix *tsi*- and doubling of the numeral, e.g. *tsi-telu-telu* 'three and three, three each time'. The protoform of the prefix is probably PAN **təγ/taγ*, compare Mal *tər*- in other functions. Ma'anyan uses the prefix *ma*- + nasal substitution in the same function.

The prefix **ma*- with doubling effect is perhaps present in western languages too. I have not found it with numerals (if Mny *ma*- above is not the same); not at least in its simple form, but as **ma*- + *γ-. In Tagalog Bloomfield names these forms 'dual collectives with *mag*-, (Bloomfield 1917, 242). Malagasy has the prefix *mi*- in the same function, and both are regular modern forms of a PAN (or PMP) **maγ*-. If this is a compound prefix, it may include the doubling **ma*-.

Its function is essentially the same in both languages. Prefixed to a kinship term it indicates two persons in the relationship expressed by the wordbase, e.g. Mlg *mi-ánaka* 'one of the parents with one child', *ánaka* 'child'. Tag *aŋ mag-áli* 'one aunt and one niece or nephew', *áli* 'aunt'. Mlg *Ra-petera mi-vadi* 'Peter and his wife', *vadi* 'spouse'. Tag *aŋ mag-asáwa ŋ Pedro at Hwana* 'Pedro and his wife Juana', *asáwa* 'spouse'. Mlg *izahay mi-anadahi* 'we (excl.) two being brother and sister', *anadahi* 'brother of a sister'. Tag *aŋ mag-áte ŋ si Marya at si Maryano* 'Mariano and his oldest sister Maria', *áte* 'elder sister'.

If there are more than two in the same relation, the number must be added in Malagasy: *isika telu mi-ánaka* 'we (incl.) three in the relation parent and child', i.e. whether both parents and one child, or one of the parents and two children. *Ra-petera éfatra mi-anadahi* 'Peter with three siblings', four altogether, both sexes represented. In Tagalog the wordbase is reduplicated if there are more than two, e.g. *aŋ mag-a-amá* 'the group of a father with two or more of his children', *amá* 'father'. Doubling of the wordbase may be used in Tagalog to indicate plural. Here reduplication changes the dual into plural.

In Tagalog *mag-anák* is an exception, in that it does not mean only two persons, but the whole family of parents and children. In Malagasy the same is obtained by adding *avi* 'every': *izi mi-anak-avi* 'he/she with the whole family' (Dahl 1972, 357-58).

The identification of this **ma*-γ- with the **ma*- forming dual of numerals in some Formosan languages is hypothetic, but the fact that its simplest form indicates duality makes the possibility a very real one.

## ABREVIATIONS

| | | | |
|---|---|---|---|
| Akl | Aklanon | Paz | Pazeh |
| AN | Austronesian | PBis | Proto-Bisayan |
| Atay | Atayal | PForm | Proto-Formosan |
| Bar | Bare'e | PMin | Proto-Minahasan |
| Bug | Buginese | PMJ | Proto-Malayo-Javanic |
| Bun | Bunun | | |
| Ci | Ciuli | PMlg | Proto-Malagasy |
| Fav | Favorlang | PMP | Proto-Malayo-Polynesian |
| Fi | Fijian | | |
| Hny | Hoanya | PPh | Proto-Philippine |
| Jav | Javanese | PSS | Proto-South-Sulawesi |
| Kan | Kanakanabu | | |
| Kav | Kavalan | Ruk | Rukai |
| Mad | Madurese | Sad | Sa'dan Toraja |
| Mak | Makassarese | Sais | Saisiat |
| Mal | Malay | Sak | Sakalava |
| Mer | Merina | Sang | Sangir |
| MJ | Malayo-Javanic | Sar | Saaroa |
| Mlg | Malagasy | Sed | Sedeq |
| Mny | Ma'anyan | Sq | Squliq |
| NgD | Ngaju Dayak | Sund | Sundanese |
| Paiw | Paiwan | Tag | Tagalog |
| PAN | Proto-Austronesian | Taok | Taokas |
| | | TB | Toba Batak |
| Pap | Papora | | |

## FOOTNOTES

1. I am indebted to R.A. Blust for valuable criticism and stimulating suggestions which have led to improvements in the present article. I am however the solely responsible for views presented.

2. Dempwolff has constructed this word as a trisyllabic PAN **bi*[*t*]*uhən*, and Dyen has rewritten it with **q* (1953, 11). I have constructed it as a disyllabic wordbase: **bi(N)t*$_1$*uq*. In his review of Dahl 1973 Blust found that my construction is unmotivated, because the cognates in AN languages reflect a trisyllabic root (1976, 234). This is correct, but in Formosan languages we find it both with prenasalization and with the vowel *a* in the last syllable: Paz *bintun*, Takituduh Bunun *bintoqán*, Paiw *vit'uqan*, Sais *bintöʔän* 'star'. Outside Formosa the vowel of the last syllable regularly reflects *ə, as in Pazeh above, and also in Puy *vituqə́n* id. But because both *-*ən* and *-*an* exist, I find it natural to consider these as the old PAN suf-

fixes, and to construct *$bi(N)t_1uq$ as wordbase.

3. A reduction of Dyen's and Tsuchida's *S*'s and *x*'s into two protophonemes *$S_1$ and *$S_2$ will be exhibited in a forthcoming publication on old phonetic changes in Austronesian.

4. In private correspondence Blust has remarked: "NgD *leŋe* is most simply derival from **ləŋən*, with -ŋ reinterpreted as 3rd sg. possessor." This is certainly a better explanation than my hypothesis of dissimilation.

5. Blust has informed me that a quinary system is also found in Ilongot (northern Philippines), Ngadha etc. (Flores), and in various languages of Melanesia, e.g. Nauna (Admiralty Islands), and Manam (New Guinea).

6. Sundanese has both *pitu* and *tud'uh*, and today the latter seems to be most frequent. It is possible that this is a loan from Malay, but *pitu* may be a loan from Javanese. Which of them is borrowed is an open question, but because *pitu* is the old AN word, and is going out of use, this may be archaic and *tud'uh* a loan.

7. The syllabic and non syllabic character of PAN vowels will be treated in a forthcoming publication on old phonetic and phonemic changes in Austronesian.

## REFERENCES


Adriani, Nicolaus 1893. *Sangireesche spraakkunst*. Leiden.

----------. 1928. *Bare'e-Nederlandsch woordenboek*. Leiden.

---------- and Alb. C. Kruijt 1914. *De Bare'e-sprekende Toradja's van Midden-Celebes*, vol. 3 by Adriani. Batavia.

Bloomfield, Leonard 1917. Tagalog texts with grammatical analysis. *University of Illinois Studies in Language and Literature*, 3.2-4. Urbana, University of Illinois.

Blust, Robert A. 1970. Proto-Austronesian addenda. *Oceanic Linguistics* 9:104-62.

----------. 1974a. The Proto-Austronesian word for "two": a second look. *Oceanic Linguistics* 13:123-61.

----------. 1974b. The Proto-North Sarawak vowel deletion hypothesis. Paper presented to the First International Conference on Comparative Austronesian Linguistics. Honolulu.

----------. 1976. Review of Dahl, *Proto-Austronesian*. *Language* 52:221-37.

----------. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics* 9.2:1-15. Honolulu, Department of Linguistics, University of Hawaii.

----------. 1978. Eastern Malayo-Polynesian: a subgrouping argument. Proceedings, Second International Conference on Austronesian Linguistics, Fascicle 1:181-234. *Pacific Linguistics* C61. Canberra, Australian National University.

Breukink, J. 1906. *Bijdragen tot eene gorontalische spraakkunst*. 's-Gravenhage.

Charles, Mathew. 1974. Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. *Oceanic Linguistics* 13:457—509.

Costenoble, H. 1940. *Die Chamoro Sprache*. 's-Gravenhage.

Dahl, Otto Chr. 1951. *Malgache et maanjan*. Oslo.

----------. 1972. Les noms de parenté en malgache. *Langues et techniques, nature et société*, tome I (En hommage à André G. Haudricourt):353-58. Paris.

----------. 1973/76. *Proto-Austronesian*. 2nd ed. 1976. Lund.

Dempwolff, Otto. 1924-25. Die L-, R- und D-Laute in austronesischen Sprachen. *Zeitschrift für Eingeborenen-Sprachen* 15, 19-50, 116-38, 223-38, 273-319.

----------. 1934-38. *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3 vols. Berlin.

Dyen, Isidore. 1947a. The Malayo-Polynesian word for 'two'. *Language* 23:50-55.

----------. 1947b. The Tagalog reflexes of Malayo-Polynesian D. *Language* 23:227-38.

----------. 1951. Proto-Malayo-Polynesian *Z. *Language* 27:534-40.

----------. 1953. *The Proto-Malayo-Polynesian laryngeals*. Baltimore.

----------. 1962. Some new Proto-Malayopolynesian initial phonemes. *Journal of the American Oriental Society* 82:214-15.

----------. 1965a. *A lexicostatistical classification of the Austronesian languages*. Baltimore.

----------. 1965b. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285-305.

----------. 1971. The Austronesian languages and Proto-Austronesian. *Current Trends in Linguistics*, Thomas A. Sebeok, ed. Volume 8: Linguistics in Oceania: 5-54. The Hague.

----------. 1975. A reconstructional confirmation: the Proto-Austronesian word for "two". *Oceanic Linguistics* 14:1-11.

Egerod, Søren. 1978. An English-Rade vocabulary. *The Museum of Far Eastern Antiquities, Stockholm*, Bulletin 50:49-104.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups*. Taipei.

Gonda, J. 1975. Observations on ordinal numbers, pp. 244-54, Miscellanea about Indonesian numerals, pp. 441-46. In *Selected Studies*, vol. 5: Indonesian linguistics. Leiden.

Hardeland, August. 1858. *Versuch einer Grammatik der dajakschen Sprache*. Amsterdam.

Hudson, Alfred B. 1967. *The Barito isolects of Borneo*. Cornell University.

Li, Paul Jen-kuei. 1978. A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology, Academia Sinica*, vol. 49, part 2: 133-99.

Matthes, B.F. 1858. *Makassaarsche spraakkunst*. Amsterdam.

Meinhof, Carl 1906. *Grundzüge einer vergleichenden Grammatik der Bantusprachen*. Berlin.

Mills, Roger Frederick. 1975. *Proto South Sulawesi and Proto Austronesian phonology.* Ann Arbor, Michigan.

Nothofer, Bernd. 1975. The reconstruction of Proto-Malayo-Javanic. *Verhandelingen van het Koninklijk Instituut* 73. 's Gravenhage.

Ogawa, Naoyoshi and Erin Asai. 1935. *Taiwan takasagozoku densetsu-shû (Myths and traditions of the Formosan native tribes).* Taihoku.

Prentice, D.J. 1971. The Murut languages of Sabah. *Pacific Linguistics* C18. Canberra, Australian National University.

Ramos, Teresita V. 1971. *Tagalog structures.* Honolulu.

Reid, Lawrence A., ed. 1971. Philippine minor languages. *Oceanic Linguistics Special Publication* No. 8.

Sneddon, J.N. 1970. The languages of Minahasa, North Celebes. *Oceanic Linguistics* 9:11-36.

---------- 1978. Proto-Minahasan: phonology, morphology and wordlist. *Pacific Linguistics* B54. Canberra, Australian National University.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology.* Tokyo.

Tuuk, H.N. van der, 1867. *Tobasche spraakkunst.* Amsterdam.

Warneck, Joh. 1977. *Toba Batak-deutsches Wörterbuch.* New ed. Den Haag.

Wolff, John. 1974. Proto-Austronesian *r and *d. *Oceanic Linguistics* 13:77-121.

Zorc, R. David Paul. 1977. The Bisayan dialects of the Philippines: subgrouping and reconstruction. *Pacific Linguistics* C44. Canberra, Australian National University.

---------- 1978. Proto-Philippine word accent: Innovation or Proto-Hesperonesian retention? Proceedings, Second International Conference on Austronesian Linguistics, Fascicle 1:67-119. *Pacific Linguistics* C.61. Canberra, Australian National University.

* * *