

Tagalog (Tg)¹ is one of the An languages with a rich inventory of affixes that are added to roots to form the words of the language. But there is something unusual and perhaps even surprising about these affixes —namely, there are only four syllables plus a handful of inflectional affixes that comprise the vast majority, easily 90%, of them. These are /pa/ which may be closed by /g/ or which may cause the initial of the root to which it is added to be substituted for by a nasal. That makes three of the syllables found in Tag affixes, and then there is only one other syllable found in more than one or two affixes: /ka/. This limitation of affix shape characterizes the affixational system of most of the An languages that have a rich inventory of affixes, especially the languages that have a conservative morphological system, and if we reconstruct proto-Austronesian (PAn) morphology based on a comparison of these languages, we would arrive at a PAn set of affixes that share this unusual typological feature. Nevertheless, there are enough affixes of a somewhat larger variety of phonemic content attested over the range of An languages, that it is improbable that the PAn affixes were as limited in shape as the Tagalog affixes. In other words, the morphological systems attested give us ample reason to believe that PAn had affixations comprised of a number of different syllable types.

It is my purpose in this paper to show that this indeed was the case and to give a list of likely candidates for affixes in the proto-language. This is a first step in the enterprise of reconstructing the morphology of PAn, and inasmuch as syllables that occur in the morphology of the An languages have become limited to little more than four sequences plus the handful of inflectional affixes, we will be in a position to investigate why this has happened and how.

In this paper my methodology will be, not to look at affixes productive in the attested languages, but rather to look at petrified affixes that comprise the first syllable of disyllabic roots followed by syllables that originated as monosyllabic roots. First, we have to discover the monosyllabic roots. As is well known and has often been discussed, a comparison of the An languages reveals a sizeable inventory of disyllabic forms of

¹ The following list gives the language abbreviations in this paper:

An	Austronesian	Tg	Tagalog	Ttb	Tontemboan
PAn	Proto-Austronesian	Cb	Cebuano	ND	Ngaju Dayak
Kav	Kavalan	MI	Malay	Fi	Fijian
Am	Amis	Bu	Bugis		

comparable meanings that resemble each other and correspond regularly in their second syllable, but not the first syllable. Blust (88) called these elements ‘roots’ even though not all of them are reflected directly by monosyllabic roots in the daughter languages. Indeed, some of these monosyllabic elements can in fact be reconstructed as morphemes in PAn or in a lower order proto-language, but some cannot. The task thus is to discover the origin of each of the two syllables in these disyllabic reconstructions, and in so doing we may develop an inventory of petrified affixes that are strong candidates for status as PAn or Proto-Malayo-Polynesian (PMP) affixes.

The first step is make an inventory of forms that have a likelihood of being monosyllabic roots in PAn or PMP. This is in fact easily done, for monosyllabic roots have left reflexes in the An languages that unequivocally derive from monosyllabic roots. These reflexes can be identified by recognizing processes of disyllabization that have operated over the range of the An languages from PAn times to the present.² In short there has been a phonological drive to disyllabization, and by examining the results of this drive, we can identify PAn monosyllabic roots.

There are three processes that affected monosyllabic roots to disyllabize them: (1) adding a prothetic vowel or stretching the nucleus, (2) doubling the root or reduplicating it, or (3) petrifying an affixed form. The root *cek ‘crammed in, squeezed into recess’ shows all three processes (and is also widely retained as a monosyllabic root): (1) Sel *osso?* ‘crowded together with little room’. This form derives from *ecek with a prothetic vowel added to the monosyllabic root. Cb disyllabizes the root by stretching the vowel nucleus. Subsequently a glottal stop is inserted between the two morae : *su?úk* ‘remote corner, being deep inside’ (2) Am and Ml disyllabize this root by doubling it: Am *cekcek* ‘stab’ Ml *sesak* ‘crammed full, packed tight’. Various languages disyllabize the monosyllabic root by adding an affix. Here are three examples of well-attested affixes added to *cek to form disyllabic roots in the current languages. The semantic force of the affix is clear from the meaning that the attested root has: Am *pacek* ‘nail’ (< *cek plus a petrified occurrence of the causative affix *pa-), Ttb *isek* (< *cek plus a petrified occurrence of the instrumental passive affix *is-

² These processes do not currently operate in all An languages. Rather, they operated in PAn and have been maintained in some An languages to the present day. In other An languages these processes are not currently operative and could well have ceased to operate at an early stage of their histories.

) ‘press down to make compact’ Cb *úruk* ‘drive stakes into the ground’ (< *um- ‘active dynamic verb affix’ added to *cek).³

Let us term these second syllables of disyllabic roots ‘potential monosyllabic roots’ if they recur with similar meanings with other penultimate syllables and correspond with similar meanings over several languages. It turns out that many of the potential monosyllabic roots are attested in one or more of the current languages in a form that indicates unequivocally they were disyllabized by one of the three processes listed above.⁴ But this is not the case of all such roots. It turns out that some cannot have been monosyllabic roots and were created at a time after the proto-language. We will come back to this point.

Let us turn our attention to the penultimate syllable of disyllabic roots. It turns out that not all these penultimate syllables originate as prefixes in the cases where the final syllable derives from a monosyllabic root. That is, not everything that is the first syllable of an attested disyllabic root is necessarily a petrified prefix when it is followed by a syllable that can be reconstructed as a monosyllabic root. An example is the root *pit, proven to be a monosyllabic root by Cb *pi?út* ‘narrow’. (This root developed in Cb by stretching the nucleus of the monosyllabic root.) Hence the following forms contain a final syllable inherited from PAN even though the first syllables of these forms are not connectable with each other or with anything else:

Am *?alapit* ‘tweezers’ Kav *sipit* ‘pinch between the fingers and twist’ Rat *sipi?* ‘pinch off’ Tg *lápít* ‘near’ Bu *seppi* ‘be pressed between’ Ml *jepit* ‘nip’ Ml *sempit* ‘narrow’ Ml *sumpit* ‘chopsticks’, Fi *ravita* ‘??’, etc.

³ Stress on the penult in this root indicates that the origin of the first syllable is the prefix *um- rather than a prothetic vowel. The loss of the *m of this prefix is normal over the entire An language range – in fact this affix rarely retains the *m when added to monosyllabic roots. The semantics strongly support the hypothesis that the first syllable of this Cb root derives from *um-: the monosyllabic root had a stative meaning ‘crammed, packed tight’. When *um- is added to this, the result would mean ‘to pack tight’. This became specialized in Cb to mean ‘drive stakes’ Further support is given by the Uma cognate *uncuk* ‘push through a crowd’, which preserves the nasal coda of the prefix.

⁴ Not all reflexes that seem to be the result of disyllabization must necessarily be that. For example, *enem ‘six’ has reflexes that could derive from *nem and others that could derive from *enem. It is not necessarily the case that *nem was the original shape and forms that reflect *enem developed the first syllable in the process of disyllabization by the addition of a prothetic /e/. Since this word in typical occurrences is pretonic (forms a phonological word with the form that follows it) and loss of the syllable three or more to the left of the stress is a normal process across the range of the An languages, it is just as possible that *enem is the original form and /nem/ developed when *e was weakened and then lost, being more than two syllables to the left of the stress. The shortened form was subsequently generalized. Further, in some cases an explanation in terms of analogical development (sound symbolism) is also possible (see below).

These forms do not derive directly from a PAN disyllabic root. All of them were probably formed in post-PAN times. Possibly some were formed by the addition of affixes that became petrified, but the majority were probably created by sound symbolism —they were created by the analogical processes whereby an existing form is changed to make it conform phonologically to a word with a similar meaning). For example, ML *jepit* derives from an earlier **kepit*, which is no longer attested in ML, but which is reflected widely over the MP area and probably derives from **pit* plus a weakened form of the well-reconstructed prefix **ka-*. In this case the earlier **kepit* became changed to *jepit* under the influence of other words beginning with an affricate referring to pinching and the like: *cubit*, *cebit*, *cepit* ‘pinch’, *jemput* ‘pick up in the fingers’. **pit* was clearly a monosyllabic root, and other forms were derived from it by analogical processes or by the disyllabization processes described above. Analogical processes may also lead to the creation of second syllables occurring in disyllabic roots with similar semantic meaning — i.e., not every potential monosyllabic root in fact derives from an earlier proto-morpheme. Some of these developed in times after the proto-language. A good example of how these monosyllabic elements may develop by analogical processes is the syllable **Vet*. No root [Vet] can be reconstructed for PAN. The starting point for this syllable was **keVet* ‘cut off’, well attested from northern Taiwan to the Solomons. Other forms ending in a reflex of **Vet* were developed on the basis of **keVet*. For example, Cb *punggut* ‘cut off head’ developed on the influence of the many forms beginning in /*pung/* that have a meaning of ‘sever, cut off’ e.g. *púngal* ‘break off s.t. firmly attached’, *pungdul* ‘blunt ended’, *punggak* ‘knock off from attachment’, *pungkul* ‘amputated, amputee’. There can be no doubt that this is the origin of Cb *punggut* even though no reflex of **keVet* is found in Cb, for **keVet* has reflexes in other languages of the Philippines, e.g. Bontoc *kelet* ‘cut hair’. This process is widespread over the range of the AN languages.

These analogical processes may affect any part of the word. Very commonly the end part of a word is altered to make it conform to another of similar meaning. An example is Cb *pukáng* ‘a child’s word for female genitalia’. This form arose as a euphemism for *púki?* ‘vagina’. It was formed under the influence of other words ending in /*kang/* that refer to the crotch or the spreading of legs, e.g. Cb *kangkang* ‘spread apart (of two or more long things joined at one end)’.

First syllables of disyllabic roots may have come into being just as well as final syllables. This process may take place when the second syllable of an existing root derives from a monosyllabic root. We might term the first syllables of such forms ‘quasi-prefixes’.⁵ An example of the creation of such a quasi-prefix is the syllable /le-/. This sequence is found in a dozen or more forms that occur in languages of an early subgroup. For example, ^{††}lekaN ‘become separated’ reconstructed for proto-Hesperonesian, [†]lekaq ‘split open’ reconstructed for PMP and *lepac ‘get free’ reconstructed for PAN. There is also a form [letas] ‘separate’, which as far as I know, does not have cognates outside of the Philippines and probably was formed on analogy to *lepac. One of these forms (possibly *lepac) provided the model whereby the others replaced the first syllable of roots ending in *kaN, *kaq, and *tac with *le-. The appendix shows these forms.

In short, both first syllables and second syllables of disyllabic roots may come into being through analogical creation. Our task is to identify what is inherited and what was created. We assume the roots that came into being by the processes of disyllabization discussed above to be inherited. The only difficulty presented is the third process listed, disyllabization by means of the petrification of an affix. In those cases we need to know which of the first syllables of the disyllabic roots reflect petrified affixes, for as was pointed out above, not all first syllables of disyllabic roots are prefixes inherited from a proto-language. They may develop by analogy,

The very clear cases of petrified affixes are those affixes that have reflexes currently still productive in the morphologically conservative An languages located on Taiwan and in the Philippines and elsewhere and that can be reconstructed for PAN on the basis of evidence from those languages. They are as follows: *is- (instrumental passive prefix), *pa-, *ka-, *um-⁶, all four of which are widely attested as petrified prefixes with monosyllabic roots, and *ma-, *-in-, *-an, and *-i, petrified with only a small number of

⁵ They are like the sequence /fl/ in English *flip, flap, fly, flutter, flail*, etc. that has a sound symbolic force and serves as the basis for the creation of new forms but is not a morpheme.

⁶ A comment must be made on the reconstruction of the prefixes *is- and *um-. Because we assume a canonical structure of the PAN word to have no clusters —i.e. no syllabic codas except at the end of a word, these affixes must have undergone alternation to accommodate this structure. The prefix *is- either lost the *s, added a paragogic *e, or metathesized to *si-. The affix *um- remained a prefix with vowel initial roots and with monosyllabic roots and became an infix with C-initial disyllabic roots. It is most commonly attested in the currently spoken languages as an infix. With monosyllabic roots, where *um- remained a prefix, it altered to drop the *m or to add a paragogic vowel *e after the *m to maintain the cononical syllabic structure.

roots. There are other affixes that can be reconstructed, but I have not identified any reconstructed disyllabic roots that contain reflexes of them. To these we should also add the processes of doubling of the entire monosyllabic root and reduplication of the initial CV of the monosyllabic root. These processes in the daughter languages in some cases refer to repeated, iterated or plural actions, and this indeed may have been the original function of these processes of doubling and reduplication, but in most cases there seems to be no other function than to enable the disyllabization of the root – i.e., the process has been entirely bleached of meaning. We will say no more about these processes.

It is not necessarily the case that each and every seeming reflex of a first syllable that is assumed to be a prefix is indeed the reflex of a prefix. For example, †palag ‘palm of the hand’ contains a well-established monosyllabic root *lag ‘lie flat’, but the first syllable *pa- is not semantically connected with any of the well-known prefixes *pa- that can be reconstructed. †Palag is likely to have been created analogically, precisely how is unknown.

The appendix presents the examples of reflexes of the productive affixes that occur petrified with monosyllabic roots. Some of these examples can be reconstructed for PAN, some only for PMP and many of them for later proto-languages (examples numbered 4, 5, 5a, 6, and 6a in the appendix).

Now to discuss the affixes that are not so readily reconstructible on the basis of productive affixation in the morphologically conservative languages of the Philippines and Taiwan. The evidence for each of these must be examined individually for the likelihood of being the reflection of an affix of a proto-language.

First is the process of nasalization (substitution of a nasal for the initial syllable of the root). This process is productive in the Hesperonesian languages in combination with a prefix shaped /pa-/ or this prefix plus the verbal inflectional affixes. In the languages of Taiwan, eastern Indonesia, and Oceania this *pa-* plus nasalization is petrified, an indication that this process existed in PAN but was lost in the areas outside of the Hesperonesian languages (and in some of them it has become non-productive as well). This affix is attested in reflexes of †paNan, given under 5a in the appendix. There is also a process of nasalization, not connected with an affix *pa-*. This process is attested with some fifteen or so of roots (disyllabic and monosyllabic). There is no discernable semantic feature that characterizes nasalization (there is no common feature whereby the nasalized roots differ from the roots

with no nasal initial), but in this case we are clearly dealing with an inherited morphemic or morphophonemic process and not with an analogical process of creation. Examples are given in the appendix under #8.

Except for the process of nasalization, we will hypothesize that the first syllable is a prefix if it has a discernable effect on the root, just as do the petrified prefixes that have productive cognates, discussed above. Similarly, where cognates of these syllables thought to be petrified prefixes occur productively in one or a few An languages the hypothesis is strongly supported that these first-syllables are indeed prefixes. If no meaning for the first syllable is discernable, we assume that the first syllable developed through analogical creation (as is the case of most of the sequences that made up the penults of roots ending in reflexes of *pit ‘pinch’ cited above).

In short, in the case of attested forms where the first syllable does not have any discernable semantic connection with the reconstructed function of a prefix, it is very possible that the disyllabic root was created analogically. The notion of ‘having a discernable effect on the root’ is not an unequivocal matter, for the reconstructed prefixes have grammatical meanings, and the determination of the meaning of a reconstructed morpheme as a transitive verb, or an adjective, or as a noun, or as a stative, or with some other semantic property reflective of its grammatical status is a matter of weighing probabilities. The reason for this is that cognate unaffixed roots may occur in different categories in the attested languages — in fact, it is normal for roots to be used in several categories (the same root may be a noun and a verb or an adjective and a noun, etc.) That means that the roots could have changed categories in the course of their history. Thus cognate roots that evince a given petrified affix in several languages may fall into quite different grammatical categories. That is not to say that it is impossible to determine what the function of a petrified affix was, or if indeed they had a function at all and were not created by analogy. In fact the totality of the attestations normally gives indications of the semantics of the reconstructed disyllabic root, and the function of the petrified prefix is discernable.

First let us talk about the clear-cut cases. These are cases where the petrified prefix seems to be cognate with a prefix that is productive in one of the An languages. The first of these prefixes is the sequence *ba-. This apparently has the function of forming an

adjective. It could be connected with ND *ba-* ‘adjective forming prefix’, but note that ND *ba-* may well derive from **baV-* as does the cognate MI prefix *ber-*.⁷ Indeed, there is an example, **beVekec* ‘bind’ that derives from a prefix **ba-* plus a coda **V* (where the /a/ of the prefix has been weakened to /e/ in the antepenult), but the other examples may or may not have had a coda **V*.⁸ Examples are given under 12 in the appendix.

An alternative form of the sequence **ba-* is **be-*. This is a weakened form of **ba-* and arose under accentual conditions that have not been identified. One hypothesis is that the forms reflecting **be-* derive from an earlier **beVe-* (the alternant of **baV-* that arose after the addition of a paragogic vowel and subsequent weakening of the antepenult), which then lost the syllable /e/ by syncope, and subsequently lost **V* by CC simplification. This is not a definitive explanation, and there is probably another explanation more likely to be true. Inasmuch as **ka-* also occurs with a weakened alternant **ke-* (see below), an as-yet unidentified process that affected both **ba-* and **ka-* is more likely to be the basis for the development of **be-*. Another argument against the hypothesis that **be-* arose from **baV-* is that there is no apparent explanation for why some roots retained **V* and others lost it by CC simplification.

Another fairly clear-cut prefix is **ta-*. This prefix is productive in ND and is probably cognate with MI *ter-*. The considerations of whether this prefix originally was closed with **V* mirrors those of the prefix **ba-* discussed in the paragraph above. Nd *ta-* and MI *ter-* both may have the function of forming statives referring to the state that results from an action, to give a very general characterization. In any case the semantic reconstructions of the disyllabic roots that contain **ta-* have a meaning component in common with the Nd and MI forms with *ta-* and *ter-*. The roots with this prefix very largely refer to having come into a certain state as a result of an action: ‘be cut down, concealed, covered, strewn, piled up, removed, loosened’. Examples are given under #10 in the appendix.

⁷ I have not done the requisite research to determine if indeed there are any remnants of ND *bah-* (deriving from **baV-*) or there are forms in languages closely related to ND reflecting **baV-*, as there are in MI.

⁸ If the original form of this prefix was **baV-*, it must have developed two alternants to accommodate the canonical open syllabic structure of PAN: **ba-* (from CC simplification) and **baVe-*, with the development of a paragogic vowel.

Somewhat less clear cut are the prefixes *ca- and *da-, the semantics of which are indistinguishable from one another. Both prefixes seem to do the same thing to the root. That is, they form roots referring to intransitive actions – an action that is performed involuntarily ‘slip, fall, choke, be tight or confined’ The Fijian prefix *ra*, which is one of the alternative stative formers, may be cognate with the first syllable of forms in other languages reflecting the petrified prefix *da-. Otherwise, I know of no language in which reflexes of these prefixes are productive. They occur as petrified affixes in some of the Philippine languages I am familiar with. The roots *cakep and *dakep ‘catch’ are clearly transitive and seem to be exceptions. We may assume that they developed transitive meanings by the process of category changing referred to above. Examples are given in the appendix #11.

Another prefix found petrified with monosyllabic roots is *ti-. There are not many disyllabic roots with *ti- that contain this prefix, but it has a clearly discernable meaning and also it is found with that meaning in at least one An language (and very likely in many of them). In Cb this affix is not productive but it is found added to many roots and also as a petrified prefix in many other roots. A few examples are *tihulúg* ‘go downwards’ (*húlug* ‘lower s.t.) *tibu?agsa?* ‘move in various directions’ (*bu?agsa?* ‘fall scattered’) *tibugsu?* ‘fall headlong’ (*bugsu?* ‘drive into ground’) *tibukúl* ‘turn into a lump, clump’ (*bukúl* ‘lump, clump’) *tikubú?* ‘bend over’ (*kubú?* ‘stoop-shouldered’), etc. Some of the reconstructed forms with *ti- as the first syllables refer to moving into a certain position or state, very much like the force of the Cb prefix *ti-*: *tikaN ‘split apart’, *tikel ‘bend the body back’, *timbaw ‘rise to the top’, and others. These are exemplified in the appendix #12.

Finally, there is the prefix *Va-. The function of this affix is the least clear-cut. It is very possible, but not absolutely certain, that forms containing *Va- arose by petrification of a prefix. They may have arisen by analogical processes. The attestations are largely verbal roots, and in many cases the root with *Va- refers to a more intense or forceful action or severe state: ‘cut down’, ‘take by force’, ‘uproot’ ‘decay to powder’, ‘get full of holes’. Examples are given in the appendix #13. There is an infix reflecting *-aV- that occurs widely and refers an intense action or severe state – e.g. Thao *parakpak* ‘sound of popping’ Cb *sagunsun* ‘coming one after another in rapid succession’ (cf. *sunsun* ‘close to each other’) Ml *kerusut* ‘totally wrinkled’ (cf. *kusut* ‘rumped, tousled’). It is at this point a matter of speculation that this petrified *Va- is cognate with the *-aV- infix, and that it derives

from an earlier *aV- that became metathesized under pressure of the canonical syllabic shape with monosyllabic roots, but was infixes when added to disyllabic root.⁹

There is also a ND prefix *ha-*, which if it is directly inherited from Pan, would derive from *Va-. The function of ND *ha-* in Hardelandt's description (1858: 32-37) is indistinguishable from *ba-*, and most of the forms with *ha-* can as well be prefixed with *ba-* instead with no difference in meaning. *Ha-* like *ba-* is an adjective forming prefix, and like *ba-* can be added to nouns to form adjectives meaning 'having [noun]'. This indicates that the ND *ha-* if it is inherited from PAN, has undergone change in function and gives no information for the reconstruction of the meaning of PAN *Va- and also not of the function of this possible prefix when added to monosyllabic roots.

REFERENCES:

- Blust, Robert. 1988. 'Austronesian root theory an Essay on the limits of morphology'. *Studies in Language Companion Series* volume 19. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hardelandt, August. 1858. *Versuch einer Grammatik der dajackschen Sprache*. Amsterdam: Frederik Muller

⁹ This speculation deserves further investigation. There are numerous disyllabic roots consisting of /a-/ plus a monosyllabic root. It is possible that these forms in fact contain reflexes of a speculative prefix *aV- that instead of metathesizing, lost the *V as a result of CC simplification. It is possible also that the coda of the prefixes *ta- and *ba- that appears in the MI prefixes *ber-* and *ter-* and in many individual petrified attestations, arose from this speculative affix (contracting the two vowels—i.e., *ba + *aV yields *baV- and *ta + *aV yields *taV-). There are also the affixes *pa- and *paV, and if *baV- and *taV- have this origin, *pa- and *paV- were formed in the same way.

The preceding document was presented at the Tenth International Conference on Austronesian Linguistics (10-ICAL). To properly reference this work, please use the following format:

<LastName>, <FirstName>. 2006. <PaperTitle>. Paper presented at Tenth International Conference on Austronesian Linguistics. 17-20 January 2006. Puerto Princesa City, Palawan, Philippines.
<http://www.sil.org/asia/philippines/ical/papers.html>

For other papers that were presented at 10-ICAL, please visit <http://www.sil.org/asia/philippines/ical/papers.html>.