

NOTES ON THE SEMANTICS OF PROTO-AUSTRONESIAN *-an 'locative'

Robert Blust
University of Hawaii

Many of the attested Austronesian languages of Taiwan, the Philippines and some other areas have complex systems of affixation which serve both case-marking (or 'focus-marking') and nominalizing functions. In their case-marking functions these affixes have prototypical uses which permit a fairly ready semantic characterization. However, the range of functions performed by 'focus-marking' affixes is surprisingly broad, and close attention to the comparative evidence shows that in the common ancestor of most attested focus languages some uses must have diverged sharply from those considered typical of their category. This paper examines some uses of Proto-Austronesian *-an 'locative' which cannot readily be classified as locative under any universal definition of this term. Lastly, it raises objections to the claim that the affixes in question had exclusively nominalizing functions in Proto-Austronesian, and implies that the line between inflection and derivation is less clear-cut in these languages than in languages such as English.

1. Introduction

The 'focus' systems of Austronesian languages such as Tagalog have been generally known to linguists at least since the work of Bloomfield (1917). Early interpretations of these systems found them highly exotic, as they appeared from an Indo-European perspective to offer multiple possibilities of passivization.

Wolff (1971) compared the focus-marking morphology of modern Formosan and Philippine languages, and on the basis of systematic correspondences of phonemic form and grammatical function he inferred a similar system for their immediate common ancestor, Proto-Austronesian (PAN). The core of Wolff's reconstructed focus system includes the following affixes: 1. *-um- 'Actor Focus (AF)', 2. *-en 'Direct Passive', 3. *-an 'Local Passive', 4. *-i- 'Instrumental Passive'. The perfective marker *-in- could co-occur with any of these voice markers in combinations that need not concern

* Acknowledgement: I am indebted to Byron W. Bender and Marian Klamer for helpful comments on an earlier version of this manuscript. They are in no way responsible for any shortcomings that remain.

us here, beyond noting that *-en is expressed with a zero allomorph in stems infixed with *-in-. For purposes of the following discussion I will adopt a slightly different terminology, using ‘Patient Focus’ (PF) for *-en, ‘Locative Focus’ (LF) for *-an, and ‘Instrumental Focus’ (IF) for *i-, which we now know was *Si- in PAN, but *i- in Proto-Malayo-Polynesian, the hypothetical ancestor of all non-Formosan AN languages (Blust 1995). An introduction to the operation of focus systems will be useful before going further. Given the central purpose of this paper such an introduction will necessarily be brief, informal and incomplete.

Focus affixes are attached to the verb stem, and signal the particular relationship which the focused NP bears to the verb. The focused NP is marked by a particle which typically distinguishes common from personal nouns (in Tagalog /ang/ vs. /si/); non-focused NPs are marked by a different particle: in Tagalog these are /ni/ for personal nouns, /sa/ for locative nouns and /nang/ (conventionally written /ng/) for most others. Consider the following sentences:

- (a) *B-um-ilí ang lalake ng tinapay sa tindahan*
 buy-AF FOCUS man PATIENT bread LOCATION store
 ‘THE MAN is buying/bought some bread at the store’
- (b) *B-um-ilí si Maria ng tinapay sa tindahan*
 buy-AF FOCUS Mary PATIENT bread LOCATION store
 ‘MARY is buying/bought some bread at the store’
- (c) *Bilh-ín ni Maria ang tinapay sa tindahan*
 buy-PF AGENT Mary FOCUS bread LOCATION store
 ‘Mary is buying THE BREAD at the store’
- (d) *B-in-ilí ni Maria ang tinapay sa tindahan*
 buy-PF/PERFECTIVE AGENT Mary FOCUS bread LOCATION store
 ‘Mary bought THE BREAD at the store’
- (e) *B-in-ilh-án ni Maria ng tinapay ang tindahan*
 buy-perfective-LF AGENT Mary PATIENT bread FOCUS store
 ‘Mary bought some bread at THE STORE’

To summarize: in (a) *ang lalake* ‘the man’ is the focused NP, and hence the actor (signalled by *-um-*); in (b) *si Maria* ‘Mary’ plays a similar role. However, in (c) and (d) it is *ang tinapay* ‘the bread’ which is focused, and hence the patient (signalled by

suffixed *-in* in the non-perfective and by infix *-in-* without an overt focus affix in the perfective), while in (e) it is *ang tindakan* 'the store' which is focused, hence the location of the action of buying (signalled by *-an*). The choice of focus appears to be governed by discourse considerations: focused NPs are old information (hence definite), and the focus apparatus serves as a means of tracking reference above the level of the sentence.

Focus systems are found in a more-or-less solid block among the AN languages of Taiwan, the Philippines and northern portions of Borneo and Sulawesi. However, they also appear in scattered locations elsewhere, most notably in Chamorro of the Marianas Islands (western Micronesia), a language with no close subgrouping ties, and in Malagasy, a language whose closest relatives in southeastern Borneo have lost almost all traces of their earlier focus morphology.

Many languages which do not have a 'focus' syntax nonetheless contain either formal or functional traces of such a system. Formal traces of an earlier focus system appear in morphemes which have lost a tightly integrated grammatical function and appear instead as derivational affixes used in nominalization processes, or as relatively unintegrated particles which have little or no grammatical function. To cite two representative examples, Kelabit of northern Sarawak has a reflex of *-an 'LF', but it no longer functions to mark location as subject, and Malay has a reflex of *si, but it no longer functions to mark a personal subject as actor.

Kelabit has evolved a syntax with a simple active-passive contrast, the active marked by *ŋ-*, with a variety of surface realizations determined by stem-initial segment, and the passive by *-en* (in the perfective by *-in-*, also with several surface realizations). During fieldwork in the 1970's I recorded a few instances of stems suffixed with *-an/* which may function verbally: *bala* 'news, fame', *mala* 'to say, tell' vs. *bela-an* 'saying, telling', *kedha* 'able to withstand pain', *ŋedha* 'to suffer' vs. *kedaan* 'suffering'. However, in most cases *-an* forms nouns of location: *dalan* 'road, path', *nalan* 'to walk' vs. *delan-an* 'path made by repeated walking over the same course'; *guta* 'wading across a river' vs. *geta-an* 'fording place'; *irup* 'what is drunk; way of drinking', *m-irup* 'to drink' vs. *rup-an* 'watering hole for animals in the jungle'.

Like Kelabit, Malay has evolved a syntax with a simple active-passive contrast. Like Tagalog, Malay uses *si* as a marker of personal names, as in *si Ahmad* 'Ahmad'. However, unlike the situation in Tagalog, Malay *si* has no syntactic function. In particular, it cannot be used to distinguish actors from patients: *si Ahmad ber-kata apa ke-pada kamu?* 'what did Ahmad say to you?'; *kamu ber-kata apa ke-pada si Ahmad?* 'what did you say to Ahmad?'. Without cognate forms from other languages the origin of such a particle would be quite obscure, as its function is minimal (Macdonald and Soenjono 1967:125 suggest that it carries a mildly denigrating connotation, especially

when used with descriptive terms to form nicknames, as with *si gemuk* ‘Fatso’, or *si kurus* ‘Bones’).

As devil’s advocate, one could argue that the derivational function of affixes such as *-an or the grammatically unintegrated use of particles such as Malay *si* represents the original situation in AN, and that languages such as Tagalog have integrated these less tightly bound morphemes into a coherent system of verbal case-marking. In effect this is what Starosta, Pawley and Reid (1982) have done: because agents (the actors of non-AF, or passive verbs) and possessors carry the same morphological marking in many AN languages, non-AF verbs can be interpreted as nouns: *kapatid ni Juan* ‘John’s sibling’; *b-in-ili ni Juan ang tinapay* ‘John bought the bread’ or ‘John’s buying/what John bought was the bread’. Where agency or genitive is marked by an overt morpheme rather than solely by structural relations that morpheme can be glossed either ‘of’ or ‘by’, as with Tagalog *ni* or its cognates in most other AN languages. Starosta, Pawley and Reid have exploited this ambivalence to argue that focus systems are historically secondary, and that the functions of the focus affixes in PAN were exclusively nominalizing. I will not enter into details here, but it is clear from several lines of evidence that languages such as Tagalog represent the original grammatical type, and that cognate morphemes in languages such as Kelabit or Malay represent the remains of what was once a more highly integrated morphological system which functioned to signal case-marking and discourse reference. Moreover, it is clear from virtually all extant focus systems that a given affixed stem could function either verbally or nominally, dependent upon context. Simple application of the Comparative Method, and argument from parsimony favors the interpretation that PAN also had a focus system in which the focus affixes functioned either verbally or nominally in given contexts.

My purpose in the following discussion is to draw attention to some of the more striking non-canonical uses of the ‘locative focus’ marker in early AN proto-languages (and in many of their descendants). By ‘non-canonical’ I mean uses that cannot under any rule-governed definition be construed as locative, and which in some cases appear far more transparently to involve other case relations (such as instrumental). While the focus of my attention will be on PAN *-an ‘LF’ and its reflexes, many of the observations that I will make can be applied to the other oblique focus affixes as well.

2. PAN *-an

Granted that it is descriptively convenient to characterize *-an and its reflexes in modern languages as signalling a locative relationship between focused NP and the verb, it is important to realize that this characterization selects a frequently recurrent type of relationship to represent a semantically much broader and more diffuse category. In the following section I will try to sketch some idea of how broad and diffuse this category really is.

2.1. *-an as source. Some Formosan aboriginal languages reflect *RiNaS, and others *RiNaS-an for various native pheasant species which within the ethnographic present were valued for their long tail feathers:

- (1) Bunun /Linan/,
Saisiyat (Taai) /Liliran/,
Thao /lhizash-an/ 'pheasant'

Both Kavalan and Thao, two languages which probably have not had a common ancestor since the dispersal of the PAN speech community on Taiwan around 4000 BC reflect the shorter form in the meaning 'long tail feathers of a bird': Kavalan *linan* 'long and beautiful tail feather' (Tsuchida 1994), Thao *lhizash* 'long feathers of a bird or fowl'. The simplest explanation of these cognate distributions is a hypothesis that PAN *RiNaS meant 'long tail feather' and that *RiNaS-an meant 'pheasant'. The essential correctness of the foregoing interpretation is corroborated by another comparison with two subparts:

- (2) a. PAN *waNiS 'tusk of a boar' :
Thao /wazish/ 'the Formosan wild boar: *sus scrofa taivanus* (Swinhoe)',
Bunun /vanis/ 'tooth, tusk',
Tsou /hisi/ 'tooth', /hisno tacvohi/ 'large tusks of a boar' (Szakos 1993),
Rukai (Budai) /valisi/ 'tooth, tusk',
Paiwan /aLis/ 'tooth, fang, tusk',
Puyuma /wali/ 'tooth, tusk of a boar',
Taokas /walis/ 'armlet' (Tsuchida 1982).
- (2) b. PAN *waNiS-an 'wild boar' :
Saisiyat /waliran/,
Rukai (Tona) /vaLisanə/ 'wild boar'

The PAN words for ‘pheasant’ and ‘wild boar’ are notable for their morphological parallelism: in each word the name of an animal is formed from the name of an economically valued product obtained from that animal, suffixed with *-an. In effect the pheasant was called ‘source of the long tail feathers’ (valued as head ornaments) and the wild boar ‘source of the tusks’ (valued as armbands). In these cases rather than signalling the place in which an action occurred, *-an signalled the ‘place’ from which an animal product was obtained.¹

2.2. *-an as instrument. A number of AN languages both within and outside Taiwan reflect PAN *CapeS, Proto-Malayo-Polynesian (PMP) *tahep ‘to winnow’. Reflexes of both *tahep and *tahep-an are found in a number of Philippine languages and in Kayan of central Borneo:

- (3) a. *tahep:
 Ilokano /taép/ ‘chaff, glume, husk’,
 Tagalog /tahíp/ ‘up and down movement of rice grains being winnowed on a flat basket’,
 Cebuano /tahúp/ ‘chaff of cereals; separate the husk from husked grains’,
 Kayan /tap/ ‘to winnow a tray of grain’
- (3) b. *tahep-an:
 Ilokano /taep-án/ ‘to winnow’,
 Tagalog /táhip-an/ ‘the flat basket used in winnowing’,
 Cebuano /taph-an-án/ ‘ground cereals to be winnowed; place where winnowing is done’,
 Kayan /tap-an/ ‘tray for winnowing rice’

In addition, a number of other languages in Borneo reflect *tahep-an with loss of the morpheme boundary in the meaning ‘winnowing basket’, and form the verb ‘to winnow’ in an innovative manner, through the use of homorganic nasal substitution:

- (3) c. Kenyah, Murik /tapan/ ‘winnowing basket’ : /napan/ ‘to winnow’

¹ Chen (1988), who is generally quite thorough, mentions neither pheasant feathers nor boar tusks as raw materials valued by the Formosan aborigines. However, Chai (1967:68) contains a photograph of a Tsou man adorned with a headdress that appears to contain the tail features of Swinhoe’s blue pheasant or a similar species, and the etymology of the Taokas word for ‘armlet’ speaks for itself.

These comparisons clearly support the reconstruction of a verb-base *tahep 'winnow' and a nominal derivative *tahep-an 'winnowing basket'. What is of interest in the present context is the use of *-an in the latter form, since *tahep-an seems clearly to be an instrument rather than a location in the sense in which that term is normally understood. The reflex of *tahep-an in Cebuano Bisayan (with double suffixation) shows the less marked, or more commonly expected semantics ('place where winnowing is done'), but this appears to be a secondary regularization of the semantics to conform with the usual semantic outcome of nominalizations with -an.

Universal semantic roles in case grammar are not explicitly defined, but it is at least intuitively unexpected that a winnowing basket would be conceived as the (stationary) place where winnowing is done rather than as the instrument with which winnowing is done, since the instrument can be moved from place to place. To date I have found no clear parallels for this usage, but its value lies in its distribution over a number of languages.

2.3. *-an as state. In the two previous examples PAN *-an or its reflex in various modern languages was shown sometimes to mark source and sometimes instrument rather than serving its expected function of marking location. Apart from these apparent cross-overs of case-marking function *-an also appears in a few cases to usurp the function of PAN *ma- 'stative'.

Word classes in PAN, PMP and many attested AN languages include verbs, nouns and various particles, but no distinct class of adjectives. Rather, verbs are distinguished as dynamic (eat, walk, swim, dream) or stative (afraid, sleep, living/alive). The latter were originally marked with the prefix *ma-, which is retained in many modern languages. The Long Semado dialect of Lun Dayeh in northern Sarawak illustrate this:

- (4) dynamic verbs:
 /ŋe-lak/ 'to cook (rice)',
 /ŋe-buen/ 'to smell something', /ŋ-abet/ 'to bite'
- (5) stative verbs:
 /me-buen/ 'smelly, malodorous',
 /m-ulun/ 'living, alive',
 /me-berat/ 'heavy',
 /me-birar/ 'yellow',
 /me-kapal/ 'thick, as a board'

PAN *tiaN 'abdomen', PMP *tian is widely reflected throughout island Southeast Asia

and the Pacific (6a). Next to PMP *tian we also find *tian-an ‘pregnant’ (6b):

- (6) a. PAN *tiaN, PMP *tian:
 Thao /tiaz/ ‘abdomen above the navel’
 Puyuma /tial/ ‘abdomen’
 Ilokano /tián/ ‘abdomen, belly; womb, uterus’
 Tagalog /tiyán/ ‘abdomen’
 Sasak /tian/ ‘belly’; /be-tian/ ‘pregnant’ (lit. ‘having a belly’)
 Banggai /tia/ ‘womb; belly’
 Palauan /díil/ ‘abdomen; womb’
 Elat /tian/ ‘belly’
 Roviana /tia-/ ‘abdomen’
 Lau /i:a/ ‘belly’
 Sa’a /ie-/ ‘belly, stomach, bowels, womb’
- (6) b. PMP *tian-an:
 Banggai /tian-an/ ‘pregnant’
 Makasarese /tian-an/ ‘pregnant’
 Lau /i:an-a/ ‘pregnant’
 ‘Are’are /iana/ ‘pregnant’
 Mota /tiana/ ‘pregnant’

In some languages, as Banggai and Lau it can be argued that the original morpheme boundary has shifted one segment to the left (hence /tia-nan/, /i:a-na/), but it is clear from the overall comparative picture that the PMP word for ‘pregnant’ consisted of the word for ‘abdomen; womb’ suffixed with *-an. Contrary to what we might expect from the general pattern for lexical items describing a state or condition, then, the PMP word for ‘pregnant’ was not marked with the stative prefix *ma-, but rather with the ‘locative’ suffix *-an, and its conceptual structure was something like ‘in the womb’.

Somewhat similar, but different in detail is a cognate set of more restricted distribution which is found in several of the languages of western Indonesia. Next to PMP *bulu ‘body hair; fur; feather; down; floss on plant stems; color; type, kind’ there is evidence for *bulu-an ‘hairy; kind of hairy fruit, rambutan : *Nephelium lappaceum*’:

- (7) PWMP *bulu-an:
 Sundanese /bulu-an/ ‘hairy, shaggy, feathered’
 Old Javanese /wulw-an/ ‘kind of hairy fruit and its tree: *Nephelium rambutan*’

Sasak /bulu-an/ 'hairy; a hairy fruit, the rambutan, *Nephelium lappaceum*'

Tae' /bulu-an/ 'a fruit, the rambutan, *Nephelium lappaceum*'

Wolio /bulu-a/ '(head of) hair; kind of fruit tree, rambutan, *Nephelium lappaceum*'

What is noteworthy about this comparison is that it parallels *tian-an in its use of *-an to signal a state or condition ('hairy'), but parallels *RiNaS-an and *waNiS-an in using *-an to form the name of a natural species (as does its Malay equivalent /rambutan/, from /rambut/ 'hair'). However, unlike the Formosan animal terms which are based on the names of culturally valued parts of the animal (long tail feathers, tusks), the *bulu-an was not a plant valued for its hairy skin. On the contrary, it is a succulent fruit valued for its edible flesh (similar to that of a lychee); the hair-like appendages covering its skin are simply a visually striking means of identification. Another example of this type is the name of the well-known durian (< *duRi 'thorn' + -an).

2.4. Unclassified. One other well-known idiosyncratic use of *-an is seen in the word for 'gold'. A number of languages reflect PAN *bulaw 'golden colored', while some of these and others also support PAN *bulaw-an 'gold':

(8) a. *bulaw:

Kavalan /bulaw/ 'golden color of ripening rice'

(loan from an unknown source)

Ilokano /boláw/ 'cock with dark brownish-yellow or drab plumage'

Tagalog /buláw/ 'reddish; reddish-gold; golden orange'

(said of young pigs and roosters)

Maranao /bolao/ 'blond, as hair'

Ngaju Dayak /bulaw/ 'gold'

(8) b. *bulaw-an:

Kavalan /berawan/,

Itbayaten /voxawan/,

Maranao /bolao-an/,

Bare'e /wuyawa/ 'gold'

Tae' /bulaan/ 'gold; noble, precious, lofty, holy'

Paulohi /hulawane/,

Numfor /brawen/ 'gold'

The morphological relationship between these forms appears to be beyond dispute.

What is curious is the manner in which *-an is used to form a concrete noun from an unaffixed stem which was either an abstract noun or an attributive. This usage has no clear parallels among the other cases cited here.

3. Verbalization vs nominalization?

As noted briefly above, Starosta, Pawley and Reid (1982) have proposed that the focus-marking affixes of Austronesian languages originally had exclusively nominalizing functions. The attested verbal functions of these affixes in scores of languages throughout Taiwan, the Philippines, northern portions of Borneo and Sulawesi, and in such outlying areas as Chamorro of western Micronesia and Malagasy, are said to have arisen through syntactic analogy with equational constructions. Although proportional analogy is an established part of the theory of language change, and has demonstrated its value in the solution of many particular problems in both phonology and morphology, syntactic analogy is far less commonly recognized. The attempt to use such a concept in accounting for the evolution of focus in Austronesian languages is thus simultaneously an attempt to explain a particular set of observations and to elaborate *de novo* the theoretical machinery essential to that explanation.²

REFERENCES

- ANDERSON, J.M. 1992. 'Linguistic representation: structural analogy and stratification.' In: Werner Winter (ed). *Trends in Linguistics Studies and Monographs* 67. Berlin: Mouton de Gruyter.
- ANTTILA, R. 1977. 'Analogy.' In: Werner Winter (ed). *Trends in linguistics State-of-the-Art Reports*. The Hague: Mouton.

² Lehmann (1992:228ff) maintains that for analogy to take place 'some linguistic set is necessary', and this may be phonological, morphological, syntactic or semantic. Most other writers on analogy, however, restrict their discussions to phonology and morphology. Anttila (1977) treats analogy in considerable detail, but does not apply the concept to syntax, Skousen (1989,92) argues that language behavior can be predicted with far greater success by analogical models than by structural models, but he makes no direct reference to syntax, and Anderson (1992) speaks of 'structural analogy', but appeals to it as a principle mediating levels of representation in linguistic descriptions rather than as an extension of proportional analogy from the well-attested domains of phonology or morphology. Finally, in a detailed discussion of principles of diachronic syntax Lightfoot (1979) is silent on analogy as a possible mechanism of syntactic change.