TOWARD RECONSTRUCTION OF DEMONSTRATIVES
IN PROTO-AUSTRONESIAN

Joseph C. Finney
Monterey, CA
<jfinney@redshift.com>

1 Overview
This is a study in diachronic syntax of the Austronesian (henceforth AN) languages. This is a preliminary study, based on study of published grammars and dictionaries of languages representing major branches of the AN language. The role of visibility in choice of demonstrative needs special study. In some languages (Mantauran Rukai) the differentiation of da (most distant) and na (intermediate) is one of visibility: if the object is not in sight it is da, and if it is in sight (and not here) it is na. Some other AN languages differentiate here-visible, here-not-visible, there-visible, and there-not-visible.2

Demonstrative adverbs (‘here’ and ‘there’) are semantically related to demonstratives (‘this’, ‘that’), and in some languages their lexical forms are systematically related, though in English, ‘this’ is not lexically related to ‘here’. The present study will refer only occasionally to demonstrative adverbs. It deals with demonstratives both as substantives (‘I saw that’, where ‘that’ is a whole DetP) and as what used to be called demonstrative adjectives or pronominal adjectives (‘I saw that chair’, where ‘that’ is the Determiner within the DetP). Nearly all AN languages have VO order, and so prepositions (and other casemarkers) come left of their Det phrases.

The conservative AN languages have casemarked demonstratives of CVCV form, in which the left syllable is a casemaker (in most cases, a former preposition) and the right syllable is a demonstrative showing distality (distance) comparable to ‘this’ and ‘that’ in English. Either consonant, or both, may be zero. While English has only two distances and Latin and Japanese have three, many AN languages have more than three, and we can list six or more monosyllabic CV demonstratives that occur widely in AN and show differing distalities. In some three-distance AN languages, including Hawaiian Polynesian and

---

1 The aboriginal languages of Formosa/Taiwan are called Formosan, as the term ‘Taiwanese’ refers to the South Chinese language that became the native language of most Chinese in Taiwan. Abbreviations used are: C and V for generalized consonant and vowel; and for language groups, IE for Indo-European, AN for Austronesian, MP for Malayo-Polynesian, and PN for Polynesian, all of which can be preceded by P for Proto. For purposes of easy comparison among languages, I respell the Hawaiian [t/k] phoneme as ‘t’. I spell the Malagasy [u] as ‘u’ and not ‘o’. Despite the tradition in Malagasy of spelling final /i/ confusingly as ‘y’, I spell it ‘i’. I spell Dyen’s PAN *C as ‘c’, so as not to confuse it with the generalized consonant. It’s an affricative like ‘ch’ in English or Spanish. In Malayo-Polynesian, and in some Formosan languages, this sound merges into PAN/PMP *t, and so, of course, the attested form of the demonstrative cu is tu in Tagalog and many other languages. At points where morpheme boundaries are relevant I have inserted hyphens that are not in the quoted sources.

2 Unlike Indo-European, AN has only one language substantially attested for a thousand years (Old Javanese); all the others became significantly attested only about 200 years ago. In many AN languages the first serious attempt to make a grammar and a dictionary was done by missionaries intending to translate the Bible and to preach.

© Joseph C. Finney
Tuvalu Polynesian, the intermediate distance (*na in both those languages) means ‘near you’. Interestingly, the same is true of Latin (*iste) and Japanese (*so-). In some of those languages, including Latin and Hawaiian, this ‘second-person’ demonstrative can be used scornfully.

In work not to be reported here, I have found that all the casemarkers attributable to PAN were ancestrally either prepositions or topic-markers in the earliest reconstructable phase of PAN. Many of the examples that we’ll see have casemarker *i (topic, nominative), *i (locative) or *di (locative). For convenience I’ll use the abbreviation PCM for ‘preposition(s) and/or (non-prepositional) casemarkers’. The PCMs used as independent words left of Det phrases are in origin the same as those occurring as the left syllables of case-marked demonstratives (though, of course, they can come to develop differently over time).

In AN as in IE, third-person pronouns typically develop from demonstratives. In AN as in IE, demonstratives are used either as substantives (constituting a whole Det phrase) or as Determiners heading a Det phrase (a use formerly called demonstrative adjective or pronominal adjective in English). The term ‘Determiner’ includes nonprepositional casemarkers as well as articles and non-substantive demonstratives. In AN as in IE, definite or specific articles can arise from further grammaticalization of demonstratives, often with lenition.

As a first approximation, there are at least six basic demonstratives that occur widely throughout AN and are candidates for being reconstructed as PAN. Roughly in order from proximal (this, here) to distal (that, there), they are:

\[
\begin{align*}
\text{PAN} & \quad \ni \quad \text{di} \quad \text{a} \quad \text{cu} \quad \text{na} \quad \text{da} \\
\text{PMP} & \quad \ni \quad \text{ri} \quad \text{a} \quad \text{tu} \quad \text{na} \quad \text{ra}
\end{align*}
\]

The left two and the right two may have been systematically related semantically and historically. The two in the center may have been inherently semantically neutral in regard to distance. It appears that in languages that have lost one or more on the distal demonstratives, the central demonstratives move distally; and in languages that have lost one or more of the proximal demonstratives, the central demonstratives move proximally in their usage. This is especially true of *tu. Note that [c] weakens to [t] not only in MP but also in some Formosan branches of AN.

The *di/ri* demonstrative is not found broadly in the Formosan languages. Its clearest occurrence in Formosa is in Nataoran Amis (Chen 1985), where it is fossilized as the right framing demonstrative regardless of distality. The variety of PAN *d occurring in the demonstratives is *di*, which in PMP is *d-ri- *i (i.e., [d] initial and [r] medial and final). Thus the *di* demonstrative appears in Tagalog as *re*, at least in the right syllable of casemarked demonstratives. In conservative MP languages, [e] is generally split from PAN *i, and [o] is generally split from PAN *u, though in major morphemes of two syllables AN final *-aw and *-ay become [-o] and [-e] respectively. Schachter & Otanes (1972: 93) say ‘the initial [d] of the *sa* forms of a deictic is frequently replaced by [r] when the deictic occurs in the middle of a phrase, particularly after a vowel.’ Later in this paper we’ll consider other possible PAN demonstratives, *ca, ta, and ti*, as well as an invader.

Perhaps the most typical language will have only three of the basic six. The order of distality varies somewhat. I’ve placed a left of *cu (tu)* because it’s used that way in Amis
Formosan as well as in some MP languages (Malagasy, Bikol). But in Tagalog, a is more distal than cu [tu], and in fact tu serves as the most proximal demonstrative of all for those Tagalog speakers who don’t use ri [rei] at all. In Sebuano (Cebuano) the order is ri (near me), ni (near you and me), na (near you), tu (away from us both or invisible to us both), though a substitutes for ri in a part of the paradigm. Some orders, however, allow no exceptions in any language. In every language that has both ni and na, na is more distal than ni. And in every language that has both di and da, da is more distal than di.

So it seems very likely that at a very early level, perhaps older than PAN itself, semantic differences were established both between *d- and *n- and between *a and *i. That implies that at some early time each of the forms *ni, *na, *di, and *da, may have been composed of two morphemes. Did one of the contrasts denote distance and the other contrast denote visibility? That’s the sort of hypothesis that must be explored. In some languages, such as Mantauran Rukai in Formosa, the most distal form, da, refers to objects that are not visible to me (and, in general, not visible to you, either).

In one group of MP languages, including Malagasy, there is a complex set of demonstratives, distinguished by both visibility and distance, and some demonstratives have been created by compounding. Though each casemarked demonstrative consists historically of two morphemes, we’ll find that in some languages the word comes to be treated as a single morpheme. One accidental fact that facilitates that outcome in MP is that the nominative or absolutive casemaker i is a homonym of the casemaker i which is locative, and which comes to add other oblique functions (genitive in some languages, accusative in others). So the combination of i plus demonstrative comes to be considered a demonstrative. So, other determiners (article or casemaker or both) come to be inserted at its left. An example is that Hawaiian has a set of demonstratives tee-ia, tee-naa, tee-la.a, whose right halves were ancestrally *i-a, *i-na, *i-ra. Such occurrences in other AN languages misled Dempwolff (1938) into constructing a PMP one-morpheme demonstrative *i(y)a. Most of us now construe the demonstrative ancestrally as simple *a.

In AN languages as in languages in general, demonstratives and third-person pronouns can often be shown to have common ancestry. Note the Romance languages which have differentiated third person pronouns, demonstratives, and definite articles from forms of Latin ille, the most distal demonstrative. Within MP, the form *i-(y)a (nominative and locative casemaker plus intermediate distance *a) is the usual source of third person singular pronoun, though the 3ps comes from PAN *da in Amis and from PAN *na in Kambera. Dempwolff, whose scope of study was MP, reconstructed both a personal pronoun i(y)a, ‘he, she’, and a ‘demonstrative pronoun’ i(y)an ‘that’. (I have modernized the notation and translated from the German.) Because -an occurs as a postposed locative marker in some Formosan languages, it seems likely that the -n forms may have earlier meant ‘there’ before meaning ‘that’. It is also worthy of note that while in two of his citations Dempwolff translates the -an form as ‘that’, in the third one he says ‘this’. That’s consistent with the intermediate position of a in distality.

2 Criteria
A hard question is: in how many languages, and in which language groups must we find a demonstrative (or any other feature) in order to justify our attributing it to PAN? The answer must depend in part on the identification of the primary branches of PAN. If we know for sure that both of two groups are primary branches of PAN, the reasonable
attribution of the element to the proto-language of each of the two groups (not by loan) is all that is needed to attribute it to PAN. But obviously, the more primary groups it occurs in, the more comfortable we feel in the attribution to PAN.

Although the MP group includes nearly all the AN languages (98 or 99 per cent of them), the conventional conclusion from Robert Blust’s thorough and systematic work has been that at best, MP is one of several primary branches, and perhaps not even that. The aboriginal languages of Formosa have long been considered by most scholars to comprise all three of the primary divisions of AN, and MP may be a subdivision of one of them. Recently, however, in a paper given at the 8th International Conference on Austronesian Languages, Blust has proposed that the Formosan languages comprise nine primary divisions of AN, with MP the tenth.

Blust’s nine Formosan branches are: 1, Atayalic; 2, East Formosan (Amis, Siraya and others); 3, Puyuma; 4, Paiwan; 5, Rukai; 6, Tsouic; 7, Bunun; 8, Western Plains; 9, Northwest Formosan (Saisiyat and Kulong-Pazeh). If, indeed, PAN has ten primary branches, it would be a good idea, in diachronic studies such as this, to examine a language in each branch, although, in theory, finding an element in two branches should be enough to reconstruct it in PAN. One principle that has been proposed is to require the element to be found both in an MP language and in a Formosan language, because Formosa is small enough that borrowing of an element from any branch to any other branch there can be suspected.

3 Demonstratives in Eastern MP

Eastern MP consists of the (many) Oceanic languages and the (few) SHWNG. With some irregularities, Hawaiian shows the first, third, fifth and sixth of the basic demonstrative set as:

(tee)nei  (*tee)ia  (tee)naa  (tee)laa

These are from PPN casemarked demonstratives:

*e-ni  *i-a  *e-na  *e-ra

These PPN forms are also reconstructed for Proto-Eastern Oceanic (though all with the ancestral casemaker $i$). Forms from other Eastern MP languages are from within this set, though often with more lenition or with fewer of the four forms. Proto-Polynesian was a highly conservative language, despite its phonological mergers and its wobbling between Accusative and Ergative syntax.

The Hawaiian Dictionary (Puku’i and Elbert 1971) indicates that ‘nei’ and ‘ia’ are virtually synonymous (both defined ‘this’), though nei never takes the tee prefix, which is a specific or definite article. Though I’ll not discuss it here, the tee element also seems to be ancestrally two morphemes, *ti-a.