

# TOWARD RECONSTRUCTION OF DEMONSTRATIVES IN PROTO-AUSTRONESIAN

Joseph C. Finney  
Monterey, CA  
<jcfinney@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.<sup>1</sup> 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.<sup>2</sup>

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 casemark (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

---

<sup>1</sup> 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.

<sup>2</sup> 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.

---

Mark Alves, Paul Sidwell & David Gil, eds. *SEALSVIII: papers from the 8th meeting of the Southeast Asian Linguistics Society* (1998). Canberra, Pacific Linguistics, 2007, pp.79-92.

© 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 casemarkers *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:

<b>PAN</b>	ni	di	a	cu	na	da
<b>PMP</b>	ni	ri	a	tu	na	ra

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-r-r (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* [re] 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 casemarker *i* is a homonym of the casemarker *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 casemarker or both) come to be inserted at its left. An example is that Hawaiian has a set of demonstratives *tee-ia*, *tee-naa*, *tee-laa*, 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 casemarker 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 casemarker *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*.

#### 4 Demonstratives in Central MP

The most thorough study of a Central MP language is that of Kambera by Klammer (1994), in which the demonstratives are given on page 59. The forms are:

ni        nai        na        nu

These include the first (*ni*) and fifth (*na*) of our basic six. Semantically, Kambera's *ni* is at me, near me; *nai* is near me but farther than *ni*; *na* is at you, near you; and *nu* is far from both me and you. (I prefer to word them this way and not use the common awkward wording, 'near the speaker', 'near the addressee'.)

Kambera (and perhaps other Central MP languages) has a system for number that I have not seen elsewhere. Each of these demonstratives is followed, in the same word, with a third person pronoun: *na*, singular, or *da*, plural. A sentence given by Klammer is:

'Nu-na	atau	'ni-na ?
That one	or	this one ?

We need not be amazed that the 3ps pronoun element added (*na*) is of the same ancestry as its homolog which is one of the basic demonstratives, and so it occurs twice as '*na-na* for one of the distances. The Kambera set (ignoring '*nai*' which as bisyllabic might be a compound) suggests the possibility of an ancestral paradigm that could be as follows:

<b>Visible:</b>	ni	na	nu
<b>Invisible:</b>	di	da	du

It is of interest that of the two in Kambera that are not among our basic six, *nai* and *nu*, each, though uncommon in the AN family, has enough presence there to suggest the possibility of other PAN demonstratives, perhaps \**Cay* / *Cai* or *Cu*.

#### 5 Question: is there a PAN demonstrative 'ay'?

In Li's study of Tanan Rukai in Formosa (his Ph.D. dissertation under Starosta) (Li, 1973, 87) the casemarked demonstrative for the most proximal distance is *kay* or *kayvay* in the Nominative and *ki-kay* in the non-Nominative. In Tagalog, the so-called *sa* case (Case 3) marker for personal nouns is *kay* (singular) and *ki-na* (non-singular). Malagasy has a demonstrative *iai* which seems to be formed irregularly: its ancestry and its formation within the language are problematical. Why doesn't an epenthetic /z/ develop there as it did in *i(y)a* > *izy*? The possibility of other PAN \**Cu* demonstratives in a paradigm with Kambera's *nu* is a hypothesis to be considered, too.

#### 6 Questions about 'ta' and 'ca'

As Mayrinax shows *-ca* and Tsou shows *-ta*, it would appear that both forms may have been demonstratives in PAN. The two forms, of course, must merge in Malayo-Polynesian, and also in Bunun and in East Formosan, which includes Amis, Siraya, and the Kavalan and Babsay-Trobiawan group. What evidence there is in languages that keep /c/ and /t/ separate, suggests that *ta* is nearer, more proximate, and more visible than *ca*. I am not aware of any language that separates /t/ and /c/ and has both *ta* and *ca* demonstratives.

## 7 Demonstratives in some Western MP tongues

Tagalog shows the second, third, and fourth of the basic set of demonstratives as:

(i)re (iy)a(n) (i)tu

though the locative (*sa*) case has not *re* but *ne*, the first, leftmost of our basic six, and adds, as its most distal, one that is not in our list of six:

i(y)o(n)

but in locative (*sa*) case:

du'on

So we have two models for the far-distant non-basic demonstrative in Tagalog; both seem to have labialization, but the *i*-case suggests that it occurs in the final syllable, while the *d*-case form suggests that labialization began in the left syllable and spread to the right. A labial element /u/ or /w/ seems also to be shown in the Wolio demonstrative, Anceaux (1987) as follows: *iwe* 'there'. We'll return to Tagalog and Wolio's labials later.

Cebuano Visayan (Bunye and Yap 1971) shows four distances (five, with problematic '*ha*'), whose order is *di-ni-na-tu* (with left syllable *di-* and *nga-*). Changing their spelling from '*nh*' to '*n*' and from '*dt*' to '*t*', the distances are:

-ri < -di	near me (not found with ri- prefix)
-a	near me (not found with nga- prefix)
-ni	near both you and me
-na	near you
-ha	far from me or from me and you (not with nga-)
-tu < cu	very far from both me and you

But we'll revisit its *dtu* later.

## 8 Demonstratives in some Formosan languages

For Mayrinax Atayal, a fairly conservative Formosan language, Huang (1994: 133) shows the demonstrative pronouns as follows:

<u>Distance</u>	<u>Visibility</u>	<u>Form</u>
nigh to me	visible	ha-ni
nigh to you	visible	ya-ni
away from both	(in)visible	ha-ca
	invisible	ya-ni

The proximal *-ni*, of course, is one of the most commonly occurring demonstratives in Austronesian languages. The *-ca*, however, is not. This paradigm shows that the two dimensions, distance and visibility, are independent of each other. But there is no simple

relation between the morphology and the dimensions. Notice that *yani* means both ‘visible and near you’ and ‘invisible and not near you nor me’. Perhaps a way to describe its territory is that it covers the intermediate ground, i.e., neither invisible far away nor visible close.

For Siriya (Adelaar, 1997: 373, which merges /t/ and /c/), the demonstratives are simple (only two, and with no morphological distinction between singular and plural):

<u>Proximal</u>	<u>Distal</u>
(this, these)	(that, those)
a-ta	a-na

For Tsou, Tung (1964) also shows that distance and visibility are independent, though Tsou’s paradigm bears no resemblance to Mayrinax’s:

<u>Visible</u>		<u>Not visible</u>	
Near:	e	Nearby:	co < *cu
Mid-distance:	si	Seen by me before:	o < *u
Far away:	ta	Never seen by me before:	na

Linguists have called these Tsou forms casemarkers. In fact, they were case-marked demonstratives, but for most speakers the casemarking syllable *i-* has disappeared and only the demonstrative remains. These are the nominative [‘second conjunctive’]. For the other case or cases, the ancestry is less clear.

For Nataoran Amis (Chen, 1987), note that ancestral *na* and *ni* have merged. That’s my interpretation of Chen’s showing that in at least some constructions it is clear that *na* is an alternate way of saying *ni* a (followed by a noun), where *a* is called a ligature. If the manifest *na* is at least partly of *ni* ancestry, it is understandable that *na* can serve as a proximal demonstrative ‘this’. The distal is *ra* < \**da*, and it serves also as 3sg pronoun. The *i-a* form serves as article ‘the’. The fourth demonstrative form that occurs is *iri* < \**i-di*; it has been specialized as the right framing demonstrative and is no longer regarded as a demonstrative. ‘Framing’ is a term first used by Ed Keenan and Ralalaoherivony (1998) for Malagasy, where an NP often is framed between two demonstratives, left and right, but Malagasy has a constraint that the left and the right must be the same.

Note also that Nataoran Amis has an interrogative pronoun *icoa* < \**i-sua*. It is called impersonal, and occurs only in ‘neutral’ case (predicate nominative), in these forms: *icoa* ‘where?’, *hacoa* ‘how much? how many?’, and *o icoaan* ‘which one?’. Each occurrence of /o/ is ancestral \**u*.

For Rukai, related dialects described by Li (1973), by Starosta, and by Zeitoun, *da* is the most distal demonstrative; it requires invisibility. In at least one Rukai language, Tanan, the demonstrative for an intermediate distance with visibility has interchangeable forms *na* and *nia*, showing that here as in Amis, the *ni* and *na* forms have merged, and perhaps for the same reason. The most proximal demonstrative determiner in Tanan Rukai is *kay*, as we have noted. A table modified from Li (1973:87) is:

<u>Proximal</u> ‘this’	<u>Intermediate</u>	<u>Distal</u> ‘that’
-NM      ki-kay	i-ni-a, ina, na	i-Da-a
+NM      kay(vay)	ku-a-ni	ku-a-Da

### 9 An intriguing discovery: A PAN demonstrative in ‘-wa’ or ‘-ua’: The logical steps in establishing it

Let’s begin, as I did, with noticing some peculiarities in one or two MP languages; and then examine more languages, one at a time, till we find a set of hypotheses and a diachronic scenario that comes close to accounting for them.

These observations are consistent with the fact that in some languages such as Tagalog the *-an* form is not the most distal, but is accompanied by a more distal form *-un*. Thus in Tagalog, corresponding to *iyán* and *diyan* we find the corresponding more distal forms *iyón* and *doon* [du’ón]. Schachter and Otnes (1972: 91) show this table of ‘deictic pronouns’:

<u>Ang Form</u>	<u>Ng Form</u>	<u>Sa Form</u>
ire /’ireh/	nire /nireh/	dine /di:neh/
ito /’itoh/	nito /nitoh/	dito /di:toh/
iyán /((’i)yan/	niyan /nyan/	diyan /dyan/
iyón /((’i)yon/	niyon /nyon/	doon /du’ón/
	noon /nu’ón/	

Note that the top line, for the most proximal ‘this’, mingles forms from the first two (*-ri* and *-ni*) of the basic six demonstratives. Even so, that combined distance form is fading out of the language. Some speakers, especially in Manila, no longer use it at all. So the generally neutral distance form, *-to* < *\*cu*, has had to move into the proximal ‘this’ distance. The a distance form (*iyán*, *niyan*, *diyan*) holds the mid distance; it’s a ‘that’ but not very far. And both forms that in general in AN show greater distance than a (*na* and *da*) are also missing. They lost out to a demonstrative from outside the basic set. Its varying forms show a labial vowel < *\*u* in the left or right syllable or in both. Whatever its origin, it shows regularities, and must have undergone some changes not easily understood.

For the second language, see the demonstratives in Wolio, a language of Southeast Sulawesi [Celebes], as given in Anceau’s (1987) Wolio Dictionary:

iwei	here
iwe	there
itu	that, yonder, over there

And some compound uses

iwe itu	there, yonder
iwei itu	over there, yonder

So far, not very promising. In the two languages the demonstratives have shown some odd forms. At this point, we have a hunch that PMP may have had a demonstrative for great distance with *-u-* or *-w-* in it. But still only a hunch.



Now, a third language, Malagasy, a language with a great number of demonstratives (seven, or as many as eleven, depending on who counts them). Malagasy (Keenan, Pearson, Randriamasimanana) has a demonstrative *iru*, or, as I prefer to write, *i-ru*. Still not very promising. Another source for Malagasy (Randriamasimanana, a linguist who is a native speaker of that language) shows what seems to be another version of the one we have just seen. It is *irua*, or as I'll write it, *i-rua*.

But three syllables are too many by one for a casemarked demonstrative. Has the first or the last syllable been added later? If the casemarked demonstrative is *i-ru*, the casemarking is the normal nominative and locative, and the demonstrative syllable must be *ru*. That doesn't fit the Wolio. If the casemarked demonstrative is *ru-a*, the casemarking is the locative /d/ as in the Tagalog, and the Wolio form suggests that the PMP form ancestral to the Malagasy and Tagalog forms may have been *ri-we*, while the Wolio came from an alternate PMP form *i-we*, using a different locative preposition. Both hypotheses of ancestral prepositions are consistent with evidence from other attested languages.

The next step is a lucky accident. At the March 1998 meeting of the Austronesian Formal Linguistics Association (AFLA) I happened to encounter a paper on the language of Madura by William D. Davies of the University of Iowa. In it are the demonstrative adjectives (determiners):

jiya	this
juwa	that

In a swift oral consultation, Bob Blust said that he had never seen AN demonstratives of that form. He identified the initial consonant /j/ as PAN \*Z.

The paired words appear to belong to a paradigm. If so, the left syllable should be a casemarker and the right syllable a demonstrative showing distance. But there's a problem. With initial PAN \*d-, a casemarking syllable \*di occurs widely, and a variant \*du is plausible. But no PAN PCM \*Zi nor \*Zu is reconstructable from other AN languages.

The final piece of evidence that solved the problem came when I looked again at Ferrell's (1981) Paiwan Dictionary. There lay an entry that I had long forgotten: a locational noun *zua* with some uses as a verb and as a demonstrative pronoun. Among its 37 entries are:

i zua	there; over there
i-zua	there is; there exists
sa-zua-u	go there!
a-zua	that, those (also contracted to: a-za)
pi-zua	to put there
pa-ka-zua-n	(1) place; (2) route, itinerary
ma-i-ta-zua	thus, in that way; (Western Paiwan) gratis
sa-zua-in	destination
zua-zua-n	(Western Paiwan) furthest

Here I use bold face for what I think is the most ancient use, 'far away place', and for the demonstrative use which it has in MP languages.

Now it becomes possible to solve the problems of the *-u-* and *-w-* demonstratives not only in Madurese but in other languages.. The word *\*zu(w)a* must have occurred as such both in Proto-Paiwan and in Proto-MP. In both, it must have had, as one of its meanings, the far distant demonstrative ‘that’. A hypothesis to consider is that *\*zu(w)a* ‘that, far distant’ was inherited from PAN or a lower-level ancestor common to Proto-Paiwan and Proto MP. But that’s not possible, at least not for PAN, because the ancestor of Madurese *j-* is PAN *\*Z*, and the ancestor of Paiwan’s *z-* is not. Notice that the word has not been reported in other branches of AN.

Another hypothesis might be that Proto-Paiwan and Proto-MP shared a common ancestor below PAN, which innovated this word after all the relevant phonological changes had taken place. That’s not the most plausible choice, either.

Notice the range of semantics in Paiwan for *zua*. In accordance with general linguistic principles the semantic ‘a distant location’ is almost certainly the earliest of the attested meanings, and the demonstrative ‘that (far away)’ is a later meaning evolved by grammaticalization. So the reasonable conclusion is that the word *zu(w)a* was borrowed from Proto-Paiwan into Proto-MP as a demonstrative pronoun (and possibly with some other uses). So the word was not a member of the class of PAN words we are reconstructing, the class of true demonstratives, each of which was a single syllable, and, indeed, a single mora. In *zu(w)a*, just as in *i(y)a*, the semivowel was nonphonemic (unless the language had a constraint that every syllable must begin with a non-zero consonant).

Notice that the two words in Madura, though made to look alike with initial /j/, are not morphologically or syntactically analogous. The casemarked demonstrative *j-i(y)-a* is made of at least two morphemes: (1) a PAN casemarker consisting of /i/ or possibly /Ci/ (in earliest times either a preposition or a topic marker), and a demonstrative *a*. But *zu(w)a*, as befits a noun, is a single morpheme of two syllables.

When the bisyllabic *zu(w)a* became fully a demonstrative in MP, it evidently posed a challenge in that it disrupted the system of AN demonstratives to which it did not conform. Let’s explore the hypothesis that the coping with such challenge can account for the irregular demonstratives that we have noted. First, what happened in the ancestor of Madurese? It must have lost all the other demonstratives, and had none left but *i(y)a* and *ju(w)a*. The latter was the distant ‘that’; and so the former, which has been a neutral demonstrative, took the territory of ‘this’. Note that Madura’s ‘that’ word is identical phonologically and morphologically with the word in Paiwan, but for the regular shift from /z/ to /j/. Then the language forced the two Madura words into a morphological paradigm by changing the initial consonant of ‘this’ from zero to /j/, i.e., the language’s definition of the demonstrative category came to include that the word begin with /j/, which became the element marking a demonstrative.

Among the languages we have seen, Madura was alone in changing the true ancestral demonstrative to fit the pattern of the invader *zua*. Apparently most languages rejected *zua* and let it go obsolete. All the other four MP languages in which we have found a *zua* descendent have changed it to fit into the paradigm of the true ancestral demonstratives. But the four tongues we are about to examine did it in four different ways.

In Wolio, the form *iwe* was reached by zeroing the initial *\*Z* and weakening the final /a/. The normal AN *\*i* vowel was inserted into the left syllable, making it a casemarking syllable, but the *u(w)* was reduced to /w/, so that the word kept only two syllables.

Tagalog is another in which the invader *zua* was reformed to fit (at least in part) the paradigm of the true ancestral AN demonstratives. As we see in the Tagalog table's bottom lines, in some syntactic cases the left syllable's vowel changed to the standard /i/ (and, if so, kept the nonphonemic /y/ glide), while in other syntactic cases the left vowel kept the invader's /u/ (or made it /o/). The initial consonant of the initial syllable of all syntactic cases shifted from *z* to the standard marker of the appropriate case. By backward assimilation, the vowel of the second syllable shifted from /a/ to a labial vowel. All this accounts for the forms seen in Tagalog's set of the most distal demonstratives, which without this reconstructed history are highly confusing and inconsistent.

The proposed scenario for Malagasy is also fascinating. First, the *zua* took the casemarking prefix and became *izua*. Then the consonant changed and so the word became *irua*, which in Malagasy's standard spelling is 'iroa'. One authority today lists it as *iroa*, while another lists it as *iro*; so apparently the language today is in the process of losing the final vowel and beginning to conform to the bisyllabic paradigm of the true ancestral demonstratives. In Malagasy, the scenario that we have established shows that (between the alternate forms *iru* and *irua*) it is *irua* and not *iru* that stays closer to its ancestral pattern (*\*zua*). We might say that through the sequence *zua* > *izua* > *irua* ('iroa') > *iru* ('iro') the word has succeeded in faking its ancestry. Now it comes to mimic the pattern of the Malagasy demonstrative *iri*, which descends normally from PAN *\*i-di*. That gives it an appearance of legitimacy. But the final step in that sequence is still in process. We are fortunate to have caught it before *irua* is totally replaced by *iru*.

Incidentally, Malagasy's *i-ri* may also descend from PAN *\*i-da*, with the same final lenition that produced *iti* from ancestral *\*i-ta*.)

One other MP language turned up with what must be a descendant of *\*zua*, and it is in the right position, as the most distal of the three demonstratives in the language. That language is Manobo (Elkins, 1968). The proximal is given as *kayi*, *he'i*, and *he'ini* (PAN *\*-ni*); the intermediate 'that, not far' is *ne'eya*, *he'eyan*, *heya* (PAN *\*-a*); and the distal 'that, far' is *haza*, *he'aza*. The reasonable interpretation is that the element *-za* is a lenition of *\*zua*. Again, the phonology was adjusted so as to fit the paradigm of the genuine PAN demonstratives, each of which is a monosyllable with a single short vowel (with a monosyllabic casemarking determiner on its left). So, the theory given here purports to show that four MP languages, in series of steps, altered the invader *zua* to conform with the paradigm of the ancestral PAN demonstratives. And each of the four did it in its own distinctive way.

So this is a nice detective story with a happy ending. Everything puzzling has been explained. The diachronic scenario works.

In summary, the scenario is as follows: In Proto-Paiwan was an old noun *zua* meaning distant location, distant place; and it was beginning to be used somewhat like a demonstrative. In PMP its use became clearly that of a demonstrative. Speakers had to reconcile its use with that of the normal PAN demonstratives. Each member of that set consisted of a right syllable that was one of 6 or 8 distance demonstratives (*ni*, *di*, *a*, *tu* < *cu*, *na*, *da*; plus probably *ta* < PAN *\*ta*, *\*ca*, and possibly *ti*), and a left syllable that was a casemarker (commonly nominative *i* or locative *i* or *di*). The Madura language responded by altering the morphology of the single remaining PAN true demonstrative to conform to that of the *zua* derivative. But Tagalog, Wolio, Manobo, and Malagasy, each in a different

way, altered the morphology of the *zua* derivative to conform to that of the PAN true demonstratives.

### 10 Marking of plurality in the demonstratives

In some AN languages the plural demonstrative is formed by inserting the syllable *-re-* between the casemarking syllable and the demonstrative syllable. Thus in Malagasy a singular casemarked demonstrative is *i-tu* (routinely misspelled *i-to*) and its plural is *i-re-tu* (*i-re-to*).

Independent 3pl pronouns in Rukai (Li, op. cit. 83) are as follows. Note that the intermediate demonstrative is used for definite and the distal demonstrative for indefinite in these pronouns:

<b>def</b>	ku-l-i ni	l-i-ni-a	zero	-l-i-ni
<b>indef</b>	ku-l-i-Da	l-i-Da-a	zero	-l-i-Da

1. Some languages don't mark plurality in the demonstratives.
2. Some languages mark plurality by inserting left or right of the demonstrative, a plural marker. This may or may not be related to the 3pl personal pronoun.
3. Some mark plurality for demonstrative pronouns as well as for common nouns by inserting a collective noun (or a former collective noun).
4. One Central MP language overtly marks number for both singular and plural by adding on the right the third person singular or plural pronoun.

### 11 Vowels 'i' for high and 'a' for far and the possibility of ancient 'u'

As we have noted, *di* is proximal and *da* distal in every language that has both. Likewise, *ni* is proximal and *na* distal in every language that has both. Does that principle hold for all consonants?

In Bunun, a Formosan language (Jeng, 1977), the principle holds for *-ti* and *-ta*. Thus (p 244) we see *?iti?* glossed 'this place' in a sentence 'I touch this place'. As determiners, the demonstratives come after the nouns in Bunun. On page 154 we see the phrase *ludun ta?* glossed 'in the mountain there'. The book has many examples of determiner *di* 'this' after nouns (busul di 'this rifle' page 154); I didn't find examples of *\*da*. Interestingly, Bunun (Jeng, op.cit.) shows the proximal *i*, distal *a* also with consonant *p*, which seldom occurs in monomoraic function words in the AN family. Examples, page 131, are for the 3<sup>rd</sup> person singular pronoun, animate, six choices, three of which are:

- c) ?aipa? (far)
- d) ?aiپی? (near)
- e) ?aipu? (far)

That example is intriguing for two reasons. First, because it shows the proximal-distal distinction by vowel with morphemes other than the *d*, *n*, and *t*. That strengthens the hypothesis that in a very ancient level, possibly earlier than PAN, each of those demonstrative syllables was built of two morphemes, with the vowel marking the distance. Thus, possibly, *\*ni* was earlier *nV + i*.

The other intriguing point is that a third vowel, u, occurs in the demonstrative paradigm and denotes far distance. That's also what it denotes in Klamer's Kambera in the Central MP area. It occurs as an interrogative pronoun in some other languages. Such examples suggest the possibility that similar use of the third vowel may have occurred in PAN, though it survives now only in a very few uses in a very few languages.

## 12 Conclusions

1. Proto Austronesian had at least six, and perhaps as many as nine, demonstrative pronouns (also serving as demonstrative determiners). All show distance in some sense.
2. All these basic root demonstratives were single light syllables CV in which the C could be zero.
3. In their common and normal use, the root demonstratives have attached at their left a casemarking syllable, which is also CV, in which the C can be zero.
4. In the relatively conservative AN languages today, the (casemarked) demonstratives still have that CVCV structure. The potential number of such is the language's number of casemarkers times its number of root demonstratives.
5. Each conservative language uses a subset of the PAN root demonstratives, sometimes as few as two (English 'this' and 'that') but typically three or four, and a few languages use more than four of them.
6. In the AN demonstratives, visibility is one of the key semantics. In some languages two demonstratives may differ only in visibility. In some languages, visibility crosscuts distance. Languages also vary in the extent to which (if at all) the demonstratives also denote visibility by the second person and/or proximity to the second person.
7. Variations are discussed in the ways in which (if at all) the demonstratives are marked for plurality.
8. I have reconstructed the prehistory of a bisyllabic invader of two syllables which entered the semantic territory of demonstratives late in the PAN period. I have traced in four languages the ways in which the invader was altered step by step to conform to the pattern of the standard casemarked demonstratives (very different ways in each of the four). And I have found one language in which it was the standard demonstrative that was reshaped and reanalyzed to conform to the pattern of the invader.

## References

- Adelaar, K. Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36(2): 362-397.
- Anceaux, J. C. 1987. *Wolio Dictionary*. Dordrecht, Netherlands: Foris.
- Bunye, Maria V. R. and Elsa P. Yap. 1971. *Cebuano Visayan Dictionary*. Honolulu: U. Hawaii Pr.
- Chen, Teresa. 1985. *Verbal Constructions and Verbal Classification in Nataoran Amis*. Pacific Linguistics C-85. Australia National University Press,
- Davies, William D. 1998. Paper read at the annual meeting of the Austronesian Formal Linguistics Society, Honolulu.
- Dempwolff, Otto. 1938. *Vergleichende Lautlehre des Austronesischen Wortschatzes*.

- Finney, Joseph C. 1997. Casemarking and the \*ia Demonstratives. *Proceedings of the Seventh International Conference on Austronesian Linguistics*, Cecilia and Wim STOKHOF (Eds.). Amsterdam.
- Huang, Lillian M. 1994. *A Study of Mayrixan Syntax*. Taipei: Crane.
- Jeng, Heng-hsiung. 1977. *Topic and Focus in Bunun*. Taipei: Special Publication 72 of the Institute of History and Philology, Academia Sinica.
- Keenan, Ed and Balohisoa Ralalaoherivony. 1998. *Raising from NP in Malagasy*. Presented at the Fourth Annual Meeting of the Austronesian Formal Linguistics Association, UCLA 1997
- Klamer, Marian. 1994. *Kambera: A language of Eastern Indonesia*. [HIL dissertation 11]. The Hague: Holland Academic Graphics.
- Li, Paul Jen-kuei. 1973. *Rukai Structure*. Institute of History & Philology, Special Publication, No. 64. Taipei: Academia Sinica. (University of Hawaii at Manoa Ph.D. dissertation)
- Puku'i, Mary Kawena and Samuel Elbert. *Hawaiian Dictionary*. 1971, Honolulu: University of Hawai'i Press.
- Rich Schachter, Paul and Otnes. 1972. *Tagalog Reference Grammar*. Berkeley and Los Angeles: University of California Press.
- Starosta, Stanley. 1996. *Formosan clause structure: transitivity, ergativity, and case marking*. Symposium series of the Academia Sinica, Number 2: Chinese languages and linguistics IV, Academia Sinica. Taipei, Taiwan. 18-20 July 1994.
- Tung, T'ung-Ho. 1964. *A Descriptive Study of the Tsou Language, Formosa*. Taipei: Special Publication 48 of the Institute of History and Philology, Academia Sinica.
- Zeitoun, Elizabeth. 1995. *Issues on Formosan Linguistics*. Paris, France: Universite Denis Diderot Ph. D. dissertation (Paris 7).

