

# A GUIDE TO THE FILIPINO LINGUIST IN READING AND APPRECIATING ROBERT A. BLUST'S AUSTRONESIAN ETYMOLOGIES (1980)

R. DAVID ZORC  
School of Australian Linguistics  
Darwin Institute of Technology

## INTRODUCTION

While it is not generally customary to review an issue of a journal, the appearance of these additional etymologies in a single volume affords us the opportunity to acknowledge the monumental contribution that Blust has made over the last decade and a half and to take stock of the state-of-the-art (or, more accurately, science) in Proto Austronesian reconstruction. For the reader who may be unable to appreciate the significance of this book, I will take the liberty of offering an outline introduction to the main tenets of comparative Austronesian linguistics, drawing on illustrations from Blust, but also including material relevant to the Philippine context (Part 1). I will then embark on a general review of the book (Part 2). Since some readers may not wish to be involved in extensive complexities, these will be limited to a more detailed critique (Part 3).

## 1. BACKGROUND INFORMATION

(1) **LANGUAGES USED.** A scholar must include a wide range of languages sufficient to give a full enough picture of both the sound system (phonology) of the parent language and the distribution of its daughter languages. In the Philippines, for example, one could theoretically reconstruct Proto Philippine on the basis of a comparison of Ilokano and Tagalog, since they are historically and geographically separated, but this reconstruction would miss numerous important distinctions and give a very limited view of what the parent language had in its lexicon and morphology. In my review of Paz's PPH reconstruction (1981), I commended her treatment of 29 widespread Philippine languages (Zorc 1981). Blust's 'present study draws on approximately 200 of the 700 or more AN languages' (9) representing diverse subgroups of the family, and an increase of at least 142 languages since his initial major contribution (1970).

(2) **CORRESPONDENCE CLASSES.** The process of historical reconstruction involves the search for similarities in four areas: sound, form, function, and meaning, or, in the jargon of the field, the isolation of phonological, morphological, grammatical, and semantic correspondences. These, of necessity, involve a further four degrees of resemblance: (a) identical, (b) regular, (c) irregular, or (d) false correspondences.

(2a) **Identities** yield rather straightforward reconstructions: Tagalog, Ilokano, Malay, and Fijian mata 'eye' yield a parent form, \*mata (the asterisk signifies that the form is hypothetical—we have no surviving proof that it was so). Note that every element is identical: the sounds m-a-t-a, usage as a noun, and semantics. [Only a few of Blust's discoveries are of this type; some examples are his #22, 29, 33, 48, 59, etc.—but the absolute identities are more limited to form rather than to meaning.]

(2b) By far the most common type of correspondence set is regular, wherein a sound in one language regularly corresponds to a different sound in another, which probably descends from a distinct phoneme of the parent (proto) language. For example, Tag ka: 'in, Ceb ka: 'un 'eat' < \*ka: 'en [note Tag i: : Ceb u < \*e], Tag la: yag, Malay layar 'sail' < \*la: yaR [Tag g: : Malay r < \*R].

(2c) **Irregular correspondences** often, but by no means always, involve loans. Thus, while \*e goes to i in Tagalog (except when a neighboring syllable has \*u, e.g., \*penuq 'full' > Tag puno?), all instances of \*e > a should be viewed with suspicion, e.g., Tag andal 'jostle' < Malay endal 'stuff, push', #120) — numerous loans from Kapampangan and Malay establish this. However, a similar development

in Pangasinan is irregular and unexplained, e.g., talo 'three' (< \*telu), pano 'full' (< \*penuq), batik 'run' (< \*betik), sali 'foot' (< \*seli), etc. It is difficult to maintain a hypothesis that these are loans (many of the etyma in question are specific to the South Cordilleran subgroup to which Png belongs), since no donor language with the same words and shifts is identifiable. A similar situation exists with Ilokano reflexes of \*R—most scholars would agree that r is the expected reflex, yet g occurs on more numerous forms, some even as doublets (ba:go, baro 'new' < \*baqRuH, \*bibig, bibir 'lip' < \*bibiR), so that the treatment of forms with g as irregular correspondences is more attractive than a loan hypothesis (viz: a split of \*R > r and g might have occurred in the history of the language).

(2d) False correspondences involve loans or accidental similarities. In a family, such as Austronesian, where so many influences have been felt (Spanish, Portuguese, Indic/Sanskrit, Arabic, Persian, Chinese, Dutch, English, etc.), a scholar cannot be expected to be a master of all potential sources. Hence, a beginner might be excused for making spurious reconstructions, e.g., <sup>x</sup>gadapun 'water-jar' (< Sp garrafon 'large decanter') or <sup>x</sup>mani? 'peanuts' (< Mex-Sp mani 'peanut'). [The raised x indicates a false correspondence.] But even a veteran scholar might include foreign material in what may be a legitimate etymon (see 2.8 below). Blust in his earlier work, as did Dempwolff before him, included appropriately coded loan material in order to exemplify regular correspondences in the languages treated. However, some readers mis-interpreted this as 'a literal claim that Sanskrit or even Arabic loanwords were found in Proto-Austronesian' (29), which clearly was not the case.

(3) RECONSTRUCTION OF PARENT LANGUAGE PHONOLOGY. Blust, as well as most researchers nowadays, uses the PAN phonology reconstructed by Dempwolff as revised by Dyen (1947, 1951, 1953a, 1965). The following table indicates the status of the PAN phonemic inventory to date [excluding the sub-numerals as discussed in Dyen 1953b (\*R<sub>1</sub>–R<sub>5</sub>) or Tsuchida 1976 (\*S<sub>1</sub>–S<sub>6</sub>), see also 2.5 and 3.1 below for further details]:

PAN	AUTHOR	NOTES	ACCEPTABILITY
*a	Demp		yes
*b	Demp		yes
*B	Prentice & Nothofer		no
*c	Dyen	Demp & Dahl *k'	yes-Blust; Wolff rejects
*C	Dyen	Dahl *t <sub>2</sub>	yes, on Formosan evidence
*d	Demp	Dahl d <sub>1</sub>	yes, but Wolff rejects
*D	Dyen	Demp *ḍ, Dahl *d <sub>2</sub>	problematic (Tagalic, Javanese, and Paiwan/Formosan evidence need comprehensive re-investigation)
*d <sub>3</sub>	Dahl		maybe; Blust rejects
*e	Demp	Demp *ə	yes
*é	Dyen	(Mid front vowel)	no
*g	Demp		yes, but Wolff rejects
*h	Demp	Revised by Dyen	yes, but = *S in Blust
*H	Dyen	Extended by Zorc	yes, on Formosan evidence
*i	Demp		yes
*j	Dyen	Demp & Dahl *g'	yes
*k	Demp		yes
*l	Demp		yes
*L	Dyen&Ts	Formosan evidence	maybe; Dahl rejects;
		recent paper by Dyen and Tsuchida may re-instate	
*m	Demp		yes
*n	Demp	Dahl *n <sub>1</sub>	yes

*ñ	Demp	Possibly only PMP	problematic next to *i
*N	Dyen&Ts	Dahl *ɳ	yes, on Formosan evidence
*ij	Demp	(Orthographic "ng")	yes
*o	Dyen	(Mid back vowel)	no
*p	Demp		yes
*q	Dyen	Demp *h	yes
*r	Dyen	Demp *ɭ	yes, but Wolff rejects
*R	Dyen	Demp & Dahl *ʀ	yes
*s	Dyen	Demp & Dahl *tʰ	yes
*S	Dyen&Ts		yes, on Formosan evidence
*Q	Tsuchida		problematic; Formosan and Malagasy evidence need re-investigation
*t	Demp	Dahl *tɿ	yes
*T	Dyen	Demp & Dahl *t̚	yes-Blust; Wolff and Dahl reject based on Indic influence in Javanese
*u	Demp		yes
*w	Dyen	Demp *v	yes, but Dahl rejects
*W	Dyen	Probably *u marker	no
*X	Dyen	Probably = *S	no
*y	Dyen	Demp *j	yes, but Dahl rejects
*z	Dyen	Demp & Dahl *dʰ	yes-Blust; Wolff rejects
*Z	Dyen	Demp *dʰ	yes, some problems
*?	Dyen & Zorc		yes, some problems
*:	Zorc	Vowel length	Blust rejects
*˘	Zorc	Vowel shortness	Blust rejects

(4) **SUBGROUPING.** Unlike determining correspondence classes, the establishment of subgroups depends on the isolation of **shared differences** (rather than similarities) – these constitute **innovations** which form the backbone of a subgrouping argument. The subgrouping hypothesis of each researcher must be made explicit. Since 1977, Blust has evolved a subgrouping that recognizes the indigenous languages of Formosa as falling into at least one first-order subgroup of Proto Austronesian, and the numerous languages spoken outside of Formosa taken together as the Malayo-Polynesian family. I am not in a position to take issue with the finer details of this hypothesis (but see 2.3). However, Blust's statement is a marked improvement over his earlier study, where (akin to a procedure used by Dempwolff) certain reconstructions were labelled PAN (Dempwolff's UAN) despite a limited distribution in Indonesian and Philippine languages [which would clearly indicate a label of 'Proto Western-Malayo-Polynesian' (in Blust's terminology) or 'Proto Hesperonesian' (in Dyen's)]. Hence, Blust's indication by a code [1 = Proto Austronesian (spoken around 5000 B.C.), 2 = Proto Malayo-Polynesian (spoken around 4000 B.C.), and 3 = Proto Western Malayo-Polynesian (spoken around 2000 B.C.)] is an honest appraisal of the status and also the approximate time-depth of each etymology. Any scholar who may disagree with these assignments can easily recompute the status based on the scope of the languages represented and his own subgrouping hypothesis.

(5) **DOUBLET vs DISJUNCT.** Blust has made a very useful distinction between reconstructions that are formally and semantically similar (**doublets** = Dempwolff's 'Nebenformen', e.g., \*adaduq/\*anaduq 'long', \*kambing/\*kanding 'goat') and reconstructions which have an overlap of cognate sets (**disjuncts**, e.g., Fijian kumi which could come from either \*kumis or \*gumi 'beard', Tagalog gata? from \*Rataq or (irregularly from) \*getaq 'coconut milk'). However, Blust recognizes that the term 'doublet' is still used to describe several quite distinct phenomena (27): phonologically similar reconstructions, e.g., \*bingag and \*bingaR [volute shell],

\*baNaw and \*baNaR *Smilax* (= true doublets) and etyma containing a monosyllabic root [or 'phonestheme', which he has since treated in more detail in a recent article (Blust, in press b)], e.g., \*ket 'sticky, adhesive' or \*ngaC, \*ngeC, \*ngiC, \*nguC 'gnash the teeth as in anger'. To this list can be added suspect synonyms, e.g., \*beRngi and \*Rabi:ʔiH 'night'. I propose that the terms and abbreviations used can be adapted to include: doublets (Dbl), disjuncts (Dsj), monosyllabic roots (Mon), and synonyms (Syn). Where there is still some potential ambiguity as to the mixture of types, a convention can be adopted to mark suspect morpheme boundaries, e.g., \*ti+kuʔ 'bend, curve' (Dbl+Mon). At least some of the remaining difficulties would appear to have to do with the quality of specific language evidence rather than problems in labelling reconstructions.

(6) TEST, CRITERION, AND WITNESS LANGUAGES. In an earlier paper, I indicated that a reconstruction is founded upon three different kinds of language evidence (Zorc 1982:114). (1) A test language is one wherein a phoneme directly reflects one—and only one—proto phoneme, e.g., Paiwan ts < \*C, Ilokano e < \*e. (2) A criterion language is one where a phoneme can relate to two or more proto phonemes, but with the evidence of additional languages, the comparativist can 'triangulate' on the most probable correspondence set, e.g., Akl ʔ < \*q or \*ʔ and Iban ʔ < \*ʔ or \*h so Akl ʔ + Ib ʔ < \*ʔ. (3) A witness language is useful primarily in determining the antiquity of an etymon—not its phonemic shape. Thus, with the loss of final consonants and/or numerous mergers in the Oceanic languages, a word might be descended from any of several etyma within a range of formal and semantic similarities. Amongst the languages of the North Cordilleran subgroup (Ibanag, Atta, Agta, Yogad, etc.), only Malaweg differentiates \*s from \*t, and is thus a test language for both phonemes; whereas Ibanag, which merges the two, as well as all final voiceless stops into glottal stop, is a witness language.

## 2. BLUST'S BOOK

The organization of Blust's book is as follows: 30 pages are dedicated to textual introduction, 7 to language abbreviations (including the subgroup affiliation of each language) and data sources, 6 to notes, 19 to references, 7 to a very useful index of languages, and a full 121 to the reconstructions themselves. This proportion is very much as it should be in a study of this kind. The brevity of the nine introductory sections belies the depth of time, research and study so obviously involved, and the importance of Blust's insights for future etymological work.

(1) The 'Brief History of Research' is an excellent survey of the field since Reland proposed 23 Malagasy-Malay comparisons in 1708 (representing the earliest Austronesian etymologies known), through Humboldt, Klaproth, van der Tuuk, Brandstetter, to Dempwolff (1938). Although there seems to have been a suspension of innovative etymological work in the two or more decades after Dempwolff, this need not reflect a tacit assumption that 'the work had been done'. (3) I suspect that this slow-down was in part due to the reorganization of academic priorities during and after World War II (when teaching and learning 'foreign' languages was foremost in the visions and finances of politicians and educationists). Also, scholars needed to understand more clearly the correspondences of reflexes in the numerous languages for which data was finally being published and the subgrouping relationships within the entire family. That might best be characterised as a period of consolidation, resulting in: the refinement of several correspondences sets (\*q, \*h, \*D, \*Z), including the introduction of new phonemes based on Formosan evidence (\*S, \*C, \*N/L, \*ʔ, \*H), and a productive debate over higher and lower order Austronesians subgroups. These endeavors consumed (certainly not wasted) a good deal of time, and were the necessary prerequisites for further careful etymological work.

Blust's statistical analysis of the contributions to Austronesian lexical reconstruction are fair and accurate, and not at all self-serving. He has, after all, almost single-handedly doubled the 2,215 etymologies made available by Dempwolff