Typology and the Linguistic Macrohistory of Island Melanesia

Mark Donohue and Simon Musgrave

MONASH UNIVERSITY

Recent years have seen much discussion on the use and meaning of typological argumentation when reconstructing language history and language relations. We address the conclusions and methodology of a paper "Structural phylogenetics and the reconstruction of ancient language history" (*Science,* Sept. 23, 2005), which claims that, on the basis of a typological comparison, the non-Austronesian languages and (Austronesian) Oceanic spoken to the immediate east of New Guinea can be shown to belong to two unrelated genetic entities. We argue that the data and discussion in this paper do not allow us to conclude that the non-Austronesian languages in the study form a valid linguistic group in any historical sense, or that the methods they apply can be used to make claims about linguistic relatedness.

1. TYPOLOGY, STATISTICS, AND HISTORICAL LINGUISTICS. In this paper we intend to provide a qualitative critique of a recent well-publicized paper dealing with the historical interaction of Oceanic languages and non-Austronesian languages spoken in the east of mainland New Guinea and the islands extending east to the middle Solomons, an area its authors call "Island Melanesia."[1] The paper, Dunn et al. (2005), makes the claim that structural ("typological") features of a group of languages can be used to determine the genetic affiliations of those languages. This work has been popularized both within and without the linguistic community, and has been heavily cited as an example of the introduction of new (with the unspoken assumption: better) methods into historical linguistics. We show that the work in this paper does not establish this claim, and, moreover, does not establish any firm factual claims concerning their study area either.

1. The term is problematic because the area known as "Melanesia" is usually thought of as extending from Maluku in eastern Indonesia through to Vanuatu. Because Dunn et al. concentrate on the (allegedly related) non-Austronesian languages, we can contract this region away from Vanuatu, where no non-Austronesian languages are found, and finish in the Solomons (though the absence of any mention of the languages of Reefs and Santa Cruz in the eastern Solomons, the affiliation of which has been the subject of much dispute, is striking). The fact that non-Austronesian languages are spoken on the islands considered to be part of Melanesia *west* of the New Guinea mainland, as well as east of it, makes the reference ambiguous. It is, however, established by popular usage as referring to the islands east of the New Guinea mainland. Similarly, the term "Papuan" is unsatisfactory, and is synonymous with "non-Austronesian." Both terms are generally used to refer to the non-Austronesian language families of New Guinea and its vicinity, other than those families that "belong" to Australia.

Oceanic Linguistics, Volume 46, no. 2 (December 2007)

Dunn et al. (2005) apply a computational tool used in (biological) genetic classification to language data, using a dataset based on structural features of language rather than one containing comparative lexical data.[2] The authors apply this method to the question of the relationships between non-Austronesian languages in Island Melanesia. The Oceanic population arrived in this region approximately 3,500 years ago (e.g., Lynch, Ross, and Crowley 2002; Pawley forthcoming), and the relationships between these languages are therefore accessible to the methods of comparative linguistics. These relationships have indeed been studied, and the existence of an Oceanic subgroup within Austronesian is beyond question; further, there is general agreement on the subgrouping within Oceanic (Lynch, Ross, and Crowley 2002). However, it has proved impossible to establish any similar relationships between the non-Austronesian languages in the region, which share almost no lexical cognates (see, e.g., Ross, 2001). As Dunn et al. note, there is archaeological evidence for human presence in Island Melanesia going back 35,000 years; if linguistic lineages are even a third as long as that, any relationships that might have existed in the lexical data would be expected to have been lost. As an alternative, Dunn et al. use structural data to try to recover relationships. Their dataset consists of 125 binary characters (relating to phonological inventories, morphological patterns, word order patterns, and other syntactic features) coded for 31 languages, 16 Oceanic and 15 non-Austronesian. As a first step, they carried out a computational analysis of the Oceanic data, and the result they present is a tree that agrees well with the consensus picture obtained from the application of "standard" comparative linguistics methods, as mentioned previously. Dunn et al. take this as evidence that historical relationships between languages can be detected using such typological data, and that they are therefore justified in applying the same technique to the non-Austronesian data. The tree that results from this second analysis defines three groups of languages that match the geography of the region. That is, there is one group associated with each of the three archipelagos of Island Melanesia (New Britain, Bougainville, and the Solomons, with Bougainville grouping more closely with the Solomons). The authors interpret this result as showing that the non-Austronesian languages shared a common ancestor approximately 10,000 years ago, although they note that the observed pattern is also compatible with an areal patterning, and that quantitative measures of the robustness of the tree are rather low.

> "Assuming that the rate of vocabulary loss in the Papuan languages is similar to rates observed elsewhere, these languages are either unrelated or have been separated at least since the early Holocene or late Pleistocene" (2073).
>
> "The results show a remarkably geographically consistent pattern" (2074).
>
> "Bootstrap values, especially in the deeper branches, are low" (2074).

We are in favor of examining structural features of languages as a means of investigating the linguistic history of an area, in addition to the more accepted morphological and lexical occurrences of regular correspondence sets being used to determine genealogical entities. This direction of research, perhaps initiated in the modern sense by Greenberg (1963), and quantified by Nichols (1992) and more recently Haspelmath et al. (2005), has provided powerful insights into the distribution of languages and linguistic types

2. We are grateful to Martin Burd and Steve McKechnie for an illuminating discussion of biological phylogenetics, and to Robert Blust, Andrew Pawley, and Malcolm Ross for insightful comments that substantially improved this paper.

around the world. Unlike Dunn et al., we are not yet sure exactly what information is revealed by the occurrence of similarities (and differences) in these features and how such data should be interpreted, and find premature the assumption that commonalities in the appearance of such features can be equated to genetic relatedness.

Examining the value and interpretation of structural features of this sort and their distribution across the languages of the region continues the debate seen in the pages of this journal in recent years (e.g., Donohue 2004a, Klamer 2002, Ross 2003). Ross (2003: 507), summarizing earlier work, states that "typological features, especially phonotactic and syntactic features, are readily copied from one language to another, including across language family boundaries." If this is accepted, it would mean that conclusions such as those drawn by Dunn et al. are particularly prone to controversy: without establishing that a particular set of correspondences in grammatical features are *not* due to areal influence, there can be no argument about the correspondence indicating a genetic relationship. The existence of linguistic areas has been discussed for a long time (e.g., Sandfeld 1930, Jakobson 1962, Weinreich 1963, Masica 1976, Emeneau 1980, Thomason and Kaufman 1988, and recently Campbell 2006), and it is surprising that Dunn et al. failed to consider the possibility of areal factors explaining their results, which by their own admission showed an areal distribution.

We know on independent grounds that the languages in Dunn et al.'s sample are from at least two populations, an immigrant Austronesian population, all of whose representatives in the study are members of the (Western) Oceanic grouping, and so shall henceforth be called "Oceanic" languages, and the indigenous non-Austronesians. The crucial claims that are made in Dunn et al.'s paper relate to the extent to which a grouping of the non-Austronesian languages can be derived using their methodology, and a crucial step in making the argument is that the methodology derives a grouping for the Oceanic languages that agrees with the picture established using the comparative method. However, we suggest that the results reported in the study do not lend strong support to this position. Gray (2005), discussing the tree grouping the non-Austronesian languages presented by Dunn et al. (2074, figure 4—figures 3 and 4 from Dunn et al. are reproduced here as our figure 1) states that "the signal that is present is consistent with a scenario involving a time depth greater than 10,000 years." However, the Oceanic tree that is printed beside the Papuan tree (their figure 3, reproduced in our figure 1) has bootstrap values that are not significantly higher than those in their tree modeling the Papuan data, and the other numerical measures reported in the text are broadly similar for the two trees.[3] For example, the consistency index for the Oceanic tree is given as 0.42, and for the non-Austronesian tree as 0.35.[4] If Gray's opinion is correct, this means that the Oceanic tree also shows a signal that is consistent with a time depth of around 10,000 years (or perhaps a little less), although we know that the actual time depth involved is not greater than 3,500 years. We must conclude that the results presented by Dunn et al. are at least as consistent

3. Bootstrapping is a technique whereby a dataset is resampled many times, with a randomly selected portion of the data included on each iteration. Bootstrap values in a tree indicate the percentage of iterations that support a particular branch; they can vary from 0 (none of the samples support the branch posited) to 100 (every sample of the data supports the branch).

4. The consistency index runs from 0 to 1, with 1 indicating complete consistency of the tree with the data; in biology a value above 0.6 or 0.7 is considered to be potentially reliable.

FIGURE 1. DUNN ET AL.'S (2005:2074) FIGURES 3 AND 4

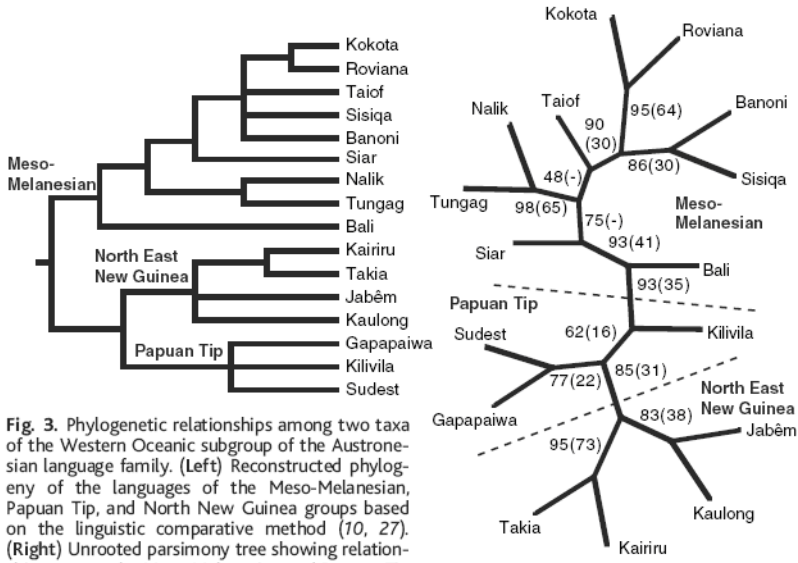


Fig. 3. Phylogenetic relationships among two taxa of the Western Oceanic subgroup of the Austronesian language family. (**Left**) Reconstructed phylogeny of the languages of the Meso-Melanesian, Papuan Tip, and North New Guinea groups based on the linguistic comparative method (*10*, *27*). (**Right**) Unrooted parsimony tree showing relationships among the Meso-Melanesian and Papuan Tip groups based on grammatical traits only (that is, discarding abundant lexical evidence) (the figure shows reweighted and raw bootstrap values). The two trees show a high degree of concordance, with monophyly in both major taxa and the similar geographical structuring of within-taxon diversity.

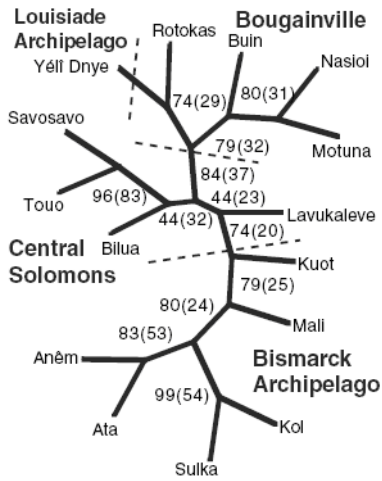


Fig. 4. Maximum parsimony tree of Island Melanesian Papuan languages with reweighted and raw bootstrap values. The tree shows a high level of geographic patterning by island group. Solomon Island languages are intermediate between Bougainville and Bismarck Archipelago languages, which is in violation of geographic progression.

with an interpretation that views the structural commonalities seen in the Island Melanesia region as being the result of (a) areal diffusion, and/or (b) the non-Austronesian languages having a shallow time depth.

We believe that the sort of typological investigation that is at the core of the methodology used by Dunn et al. is useful, and that it can produce valid results (though not necessarily in the area of determining genetic relatedness), but that:

- the methodology requires very careful checking and safeguards against redundancy and inconsistency;
- any such investigation needs to consider exactly what the central claim is carefully, and to evaluate the standards of proof that are required to assess the data; and
- appropriate methodology, especially appropriate tools and techniques, should be used to substantiate or quantify the claims.

We show, in the following sections, that in none of these regards do Dunn et al. meet minimum standards.

We also wish to emphasize that we are not arguing against the use of computational tools in analyzing data of this type. The algorithms developed for phylogenetic analysis in biology can be a powerful (and, importantly, impartial) tool in language research as demonstrated by, for example, Ringe, Warnow, and Taylor (2002) and McMahon and McMahon (2003), who combine the traditional methodology of historical linguistics with the rigor of computational verification. These studies apply computational methods to lexical data, but we see no problem in principle with the application of similar methods to structural data. A similar position is taken by Wichmann and Saunders (to appear), who address the methodological challenges involved in applying typological methods to historical classification. What remains constant in each case is the necessity for the application of stringent methodological standards, and care in the interpretation of results.

In the remainder of this paper, we discuss in detail the issues raised in this introduction. First, we turn to the question of exactly what claim Dunn et al. are making (section 2). Then in section 3, we discuss shortcomings in the dataset that they used, while section 4 considers the issue of relative time depth and appropriate comparison sets. The following section presents empirical data in support of our position that the data and analysis used by Dunn et al. do not make the kind of discriminations between groups of languages that might be expected of a reliable methodology, and that areal factors can be shown to be at least as significant as genetic relationships. In the concluding section, we raise briefly some other issues and summarize our arguments.

2. THE CENTRAL CLAIM IN DUNN ET AL. There are two logically possible factual claims that we might infer to be central to Dunn et al.'s study:

1. the non-Austronesian languages of Island Melanesia are related to each other (but not to other non-Austronesian languages in the New Guinea area).
2. the non-Austronesian languages of Island Melanesia are related to each other (and possibly to other non-Austronesian languages [of the region] as well).

The passages that bear on the interpretation of their claim are as follows:

> "The most plausible hypothesis to explain this result is the divergence of the Papuan languages from a common ancestral stock" (2072).

> "A plausible interpretation of the Papuan language tree is thus that the two language groups now located on the Solomons and Bougainville separated from a common ancestor" (2075).

The first quote appears to take a stance on the relatedness of the "Island Melanesian" languages, but, in fact, does not preclude the possibility that they are simply a nonsubgrouping part of an unspecified greater entity. The second quote takes no position on the unity of the non-Austronesian languages of Island Melanesia, but offers the possibility that there is a group consisting of the languages of Bougainville and the Solomons.

In short, we cannot assess the factual claim in Dunn et al. with any certainty, though it appears that the argument is for an "Island Melanesian" group, roughly corresponding to what was previously referred to as the East Papuan Phylum.[5] Addressing the supposed unity of this group of languages, Ross (2001) concluded that there is no such thing as a genetic grouping that links the languages discussed in Dunn et al. Ross's conclusion is that there are eight unrelated groups of languages scattered about the archipelago.

In parts of Dunn et al.'s presentation we can find tacit support for Ross's notion that there is no geneological relationship between these languages. They acknowledge that "the lexical evidence for relationships between Papuan languages is minimal" (2073), and that "one possibility is that these trees reflect contact with local Austronesian neighbors, providing an areal rather than phylogenetic signal" (2074), and even that "a second possibility is the null hypothesis of no relatedness between the Papuan languages. In that case, we would not expect the orderly and geographically consistent phylogenetic signal that does emerge from the data" (2074).

TABLE 1. ROSS'S EIGHT FAMILIES OF ISLAND MELANESIA

1	Yele-West New Britain	West New Britain	Yele (Yélî-Dnye)
			Anêm
			Ata
2	Kol		Kol
3	Sulka		Sulka
4	East New Britain		Baining
			Taulil
			Butam
5	Kuot		Kuot
6	North Bougainville		Rotokas
			Konua
7	South Bougainville		Nasioi
			Nagovisi
			Buin
			Motuna
8	Central Solomons		Bilua
			Baniata (= Touo)
			Lavukaleve
			Savosavo

5. Dunn et al. do not include in their sample all of the languages that had been assigned to that grouping; none of the languages of the eastern Solomons, spoken on Santa Cruz and Reefs, was included, nor were Nagovisi and Konua, from Bougainville. The last two exclusions are surprising, given the fact that there is material available for them, as cited in Ross (2001)—namely, Decker (1981) and Müller (1954). Mali is included in Dunn et al.'s sample, but not overtly in Ross's, reflecting the fact that Mali is a variety of what Ross refers to as Baining (Baining is also known in the literature as Qaqet).

Of course, if there were no genetic relatedness, we would expect that languages close to each other would show more similarities than languages far from each other, due to the effects of contact and linguistic areas. Dunn et al. admit this: "regional diffusion also may account for the phylogenetic signal observed, a possibility that we cannot test without more detailed archaeological information" (2074). A final claim appears: "the lack of putative lexical cognates dates these relationships considerably before the Austronesian arrival" (2074).

We consider that the simplest explanation for the lack of "putative lexical cognates" is that the languages do *not* share any relationship in the historical sense, and that any *typological* convergence between the languages is the result of the languages being present in the same linguistic area or areas. It is, of course, impossible to prove that two (or more) languages are not related; we can, however, state that there is no convincing evidence that they are related, and this is what we believe to be true for the non-Austronesian languages in Dunn et al. This is important for the evaluation of the other, tacit, claim present in Dunn et al.:

3. analysis using typological features can determine historical relationships in the absence of any lexical signal, and in the absence of the application of the comparative method.

Because the bulk of the current article is devoted to assessing the kind of data used in their study, and how robust it is, we shall leave this claim for sections 3–5 that follow.

No justification for the choice of Oceanic languages in the sample is given. The Oceanic languages are described as being the "Austronesian neighbors" of the non-Austronesian languages in question, but some (e.g., Kairiru) are located far from any of these non-Austronesian populations, while others that are the immediate neighbors of non-Austronesians (e.g., Lusi, described in the same volumes that were used to collect data on non-Austronesian Anêm) were omitted.

As stated at the beginning of this section, it is hard to know exactly what the central claim in Dunn et al. is. However, as they do not include data from any non-Austronesian languages outside the region that is the focus of their study, their evidence can only bear on the question of whether the non-Austronesian languages of Island Melanesia are related to each other. Our discussion therefore concentrates on assessing Dunn et al.'s evidence for this claim, although in our conclusion we return to the issue of whether a wider-ranging sample would be desirable from the methodological perspective.

3. HOW IS THIS CLAIM INVESTIGATED/ESTABLISHED?

Leaving aside the question of what the claims of the paper are, we turn now to the way in which the study was conducted.

3.1 THE 125 FEATURES.

Dunn et al. base their subgrouping on the scores that their languages show when 125 structural features, ranging from phonology to discourse, are encoded for each language in a binary format (for instance, "Does the language have fricatives? Yes/No?"). These 125 features are problematic in many ways. First, and most obviously, not all language have values coded for all of the features in the survey; of the maximal 3,875 features coded (125 features x 31 languages), fully 373 data points,

almost 10 percent of the total, are not present. In 17 instances the coded value of a feature is absent for eight or more languages, one quarter of the languages in the sample; for one feature, 16 languages, over 50 percent of the sample, lacked a value (feature 114, "Say in desiderative constructions"). Only 40 features, less than one third of the entire sample, are represented with data from all 31 languages. Looked at from the perspective of successfully coding features for languages, we find that the languages vary between a 100 percent successful coding rate (Kilivila, Sulka, and Yélî-Dnye) down to a rate of less than 60 percent of the features coded (Buin, for which 52 features have missing values).

Why are data for so many features missing? Largely, this can be explained by looking at the data sources used. While data for the non-Austronesian languages were drawn from fieldwork carried out by a subset of the authors, or their colleagues, or else drawn from SIL grammatical manuscripts of unstated length, much of the Oceanic data came from grammatical sketches. Table 4 lists the length of the cited Oceanic sources used in the

TABLE 2. NUMBER OF LANGUAGES FOR WHICH A PARTICULAR FEATURE WAS NOT CODED IN EIGHT OR MORE LANGUAGES (i.e., ABSENT IN AT LEAST 25% OF THE LANGUAGES IN THE SAMPLE)

FEATURE	MISSING LANGUAGES	PERCENTAGE MISSING	FEATURE	MISSING LANGUAGES	PERCENTAGE MISSING
26	15	48%	92	8	26%
46	14	45%	93	13	42%
47	10	32%	94	10	32%
74	8	26%	114	16	52%
75	10	32%	116	11	35%
78	14	45%	117	10	32%
89	8	26%	121	10	32%
90	11	35%	123	15	48%
91	12	39%			

TABLE 3. FEATURES THAT WERE NOT CODED FOR LANGUAGES

LANGUAGE	OCEANIC ?	NUMBER OF MISSING VALUES	PERCENTAGE MISSING	LANGUAGE	OCEANIC ?	NUMBER OF MISSING VALUES	PERCENTAGE MISSING
Yelî Dnye	–	0	0.0%	Siar	+	12	9.6%
Sulka	–	0	0.0%	Sisiqa	+	12	9.6%
Kilivila	+	0	0.0%	Roviana	+	13	10.4%
Kuot	–	1	0.8%	Gapapaiwa	+	13	10.4%
Anêm	–	1	0.8%	Sudest	+	13	10.4%
Lavukaleve	–	1	0.8%	Nasioi	–	14	11.2%
Mali	–	3	2.4%	Bali-Vitu	+	16	12.8%
Kol	–	4	3.2%	Takia	+	17	13.6%
Nalik	+	4	3.2%	Taiof	+	18	14.4%
Motuna	–	5	4.0%	Rotokas	–	19	15.2%
Touo	–	5	4.0%	Ata	–	22	17.6%
Bilua	–	5	4.0%	Jabêm	+	27	21.6%
Kokota	+	6	4.8%	Kairiru	+	28	22.4%
Savosavo	–	8	6.4%	Kaulong	+	34	27.2%
Banoni	+	9	7.2%	Buin	–	52	41.6%
Tungag	+	11	8.8%				

study.[6] Apart from Kilivila, Kokota, and Nalik, the average length of their sources is 22 pages, not by any means an "authoritative" work, and certainly not one in which one would expect to find exhaustive exemplification of syntactic constructions or discussion about discourse structures, which are required given the questionnaire that Dunn et al. use.

While it is better to include some data rather than none, we must wonder why almost all of the Oceanic language data were taken from these short grammar sketches, when many full-sized grammars of Oceanic languages in the (ill-defined) area are easily available. Our own shelves contain monographs on Adzera (Holzknecht 1986), Mangap-Mbula (Bugenhagen 1995), Mekeo (Jones 1998), Motu (Lister-Turner and Clark 1931, Taylor 1970), Nakanai (Johnston 1980), Tawala (Ezard 1997), and Tigak (Beaumont 1979), and sketches of more than 22 (or even 46) pages' length about Oceanic languages in the area—Labu (Siegel 1984), Maisin (Ross 1984), Saliba (Mosel 1994), Sinaugoro/Balawaia (Kolia 1978, Tauberschmidt 1999). Of the Oceanic languages that Dunn et al. examine, more extensive documentation is available for Banoni, Jabêm, Kairiru, and Roviana (Corston 1996, Dempwolff 1939, Lincoln 1976, Ray 1926, Waterhouse 1928, Wivell 1981, Zahn 1940, 1982), but was not examined. Other languages in the geographic bounds of their study for which extensive materials are available include Manam (Lichtenberk 1983), Numbami (Bradshaw 1978, 1982, 1993, 1997, 1999, 2006), and Hoava (Davis 2003).

It might seem that missing data simply reduce the reliability of the study, but given the software that Dunn et al. use, the problem is greater than this. PAUP, the software employed to generate trees of "best fit," actively utilizes uncoded features to support trees as wild cards: a missing value for a feature can be used to support a link, not simply to not support it. The fact that the rate of missing data is so great then leads to even less confidence in the subgrouping they propose.

3.2 THE SELECTION OF THE FEATURES: PROBLEMS. Leaving aside the question of the completeness of the survey, the selection of the features themselves is also highly problematic. The description of the features used reads well. The authors state that the grammatical features chosen "were selected to provide broad typological coverage, reflecting the known linguistic variation of the region." The fact, however, that the

TABLE 4. LENGTH OF MATERIALS USED FOR OCEANIC LANGUAGES

Bali	25 pages	Kaulong	23 pages	Taiof	14 pages
Banoni	16 pages	Roviana	31 pages	Takia	33 pages
Gapapaiwa	25 pages	Siar	16 pages	Tungag	46 pages
Jabêm	27 pages	Sisiqa	11 pages	**Average sketch size:**	**22 pages**
Kairiru	12 pages	Sudest	25 pages		
Kilivila	2 books				
Kokota	1 book				
Nalik	1 book				

6. For some languages (Jabêm, Kairiru, Kaulong, and Takia) no sources are cited. Because the Oceanic data are overwhelmingly drawn from Lynch, Ross, and Crowley (2002), and because the "missing" languages are also found in that work, we have assumed that it also served as the source for these languages.

reference to sources for the "known linguistic variation" is an earlier article by a subset of the authors of the 2005 work here under review leads to concerns about the "independence" of the features selected. If the features are ones that maximize the distinction between the Oceanic and non-Austronesian languages in the region, then the statistics will not reflect an independent assessment of their relatedness, because the results would have been weighted before the statistical assessment began.

There are two ways to evaluate the extent to which this is a problem. The first is to use the features employed by Dunn et al. to attempt to evaluate relationships between different groups of languages. For instance, one could code values for English, Dutch, German, Spanish, and Basque, and then examine the results. English, Dutch, and German are known to be relatively closely related to each other, with Spanish a more distant cousin, while Basque is genetically unrelated to the four Indo-European languages. The second test is to use a different, independent, set of features that do successfully differentiate known language families to evaluate all or part of the languages in Dunn et al.'s sample, for example, the sample used in the *World Atlas of Language Structures* (Haspelmath et al. 2005).

Additionally, the features selected are not all independent of each other. The authors state that "traits invariant in the region (either entirely absent, … or present in all the languages …) were not coded. Characters [= grammatical features—D&M] that show strong implicational correlations were excluded …." This would represent "best practice" in responsible data-set design. Despite this statement, however, there are two instances in the dataset of features that are uninformative. Character 17 (number not marked on pronouns) has all 31 languages coded with the value 1, meaning that number is not marked on pronouns, and character 115 (relative clauses present/absent) has 28 languages coded with the value 1, with the remaining three unknown. Clearly these are features that are invariant in the region, and yet they were included. There are also typological dependencies present in the character set, where the value of one character allows us to predict the value of other characters with better than chance accuracy. For example, characters 28, 29, and 30 code (respectively) the presence of dual marking, the presence of plural marking, and the presence of number marking other than singular, plural, and dual. But it is well known that these possibilities are related implicationally: dual marking implies plural marking, and any other marking implies dual marking. We identify nine examples of this type of dependency. This discussion refers to typological dependencies, which are implicational relationships across human languages in general. We suspect that there are also implicational relationships contingent to this particular sample of languages, cases where the value of two or more characters covaries consistently in this sample but not in languages more generally. We have not analyzed such relationships in this research.

We also found numerous instances in which one value is wholly or partly dependent on the value given for another feature. Here, we refer to instances in which there is a logical relationship between the definitions of two or more characters, such that the value assigned a language for one character restricts the possible value of other characters. A simple example of such a dependency can be seen by considering characters 82, 83, and 84. Character 82 codes the presence of intransitivizing morphology, while characters 83 and 84 code the presence of (respectively) reflexive and reciprocal morphology. Reflexive and reciprocal morphology function to reduce the transitivity of a verb; therefore, we sug-

gest that if either character 83 or 84 is assigned a value 1, then character 82 must also have the value 1. Taking a more complex example, a value of 1 ("present") for either of feature 39 or 40 ("Prefix-marked possession" and "Suffix-marked possession") implies that there must be a value of 1 ("present") in either feature 41 or 42 ("Marked possessor" or "Marked possessee"). An even more complex example can be seen with character 33 and characters 58–63 and 68–69. Character 33 codes the presence of concord beyond the NP, the possibility that elements outside the noun phrase can agree with the class of a noun. We take it that this refers to agreement or cross-referencing on the verb, and it follows from that assumption that a value of 1 for character 33 implies a value of 1 for at least one of the characters 58–63, which code affixal marking of core participants, or a value of 1 for one of characters 68 and 69, which code suppletive marking on verbs. We have found 22 dependencies of this type. In several instances no meaningful value can be assigned to one character if a certain value of another character occurs. For example, character 105 codes whether the same negator is used for verbal and nonverbal predicates. But no meaningful value can be assigned to this character if character 87, which codes the presence of nonverbal predicates, has the value 0.[7] We identify five dependencies of this type, referred to in the table below as privative relationships. The different types of dependencies that we found between features in the data set are summarized in table 5.

Clearly there are problems in the selection of features in the study, such that less than 125 *independent* aspects of grammar were used to determine the typological assessment. We have not entered into a debate on coding decisions, except where it involves apparently contradictory values for related features, because that would move from the objective into the subjective. But we believe that we have established that there are not 125 separate grammatical features in use, and that those features are not objective, in that they were selected with discrimination in mind.

The following section presents a brief interlude, pursuing the implications of one of the conclusions reached by Dunn et al., before we examine the value of the features that Dunn et al. used in terms of discriminating the Oceanic and non-Austronesian populations.

TABLE 5. DEPENDENCIES OBSERVED BETWEEN FEATURES

TYPE OF DEPENDENCY	PAIRS OR GROUPS OF FEATURES
Implicational relationship: value of one character strongly predicts value of another character or group of characters based on typological knowledge	10
Logical relationship: value of a character entails value of another character or group of characters	22
Privative relationships: value of a character makes value of another character or group of characters meaningless	5
Total dependencies:	37

7. Confusingly, only one language in the sample in Dunn et al., Rotokas, is coded as *not* having nonverbal predicates, and yet Rotokas is given a value of 1 for feature 105 ("verbal and nonverbal predicates are marked by the same negator"). It is hard to see how this coding decision was reached.

4. TIME DEPTH. Dunn et al. appear to conclude that the non-Austronesian languages of Island Melanesia are related at a time depth of approximately 10,000 years. This is based on the lack of any lexical resemblances beyond those that could be ascribed to chance (and, we presume, the lack of any regular sound correspondences), and the presence of similar typological profiles: "A plausible interpretation of the Papuan language tree is thus that the two language groups now located on the Solomons and Bougainville separated from a common ancestor. This could have happened while they could still freely migrate on a common landmass, a time depth (~ 10,000 years) in accord with that required to erode traces of common vocabulary" (2075).

The 10,000-year time depth is based on the assumption that vocabulary "decays" at a relatively constant rate: "Assuming that the rate of vocabulary loss in the Papuan languages is similar to rates observed elsewhere, these languages are either unrelated or have been separated at least since the early Holocene or late Pleistocene." In fact, such rates vary considerably from one language to another and from one period of a single language's development to another (Embleton 1986:99–101; Chowning 1985; Pawley 2006, 2007) and such variation has been demonstrated in detail for Austronesian languages by Blust (1999). This point is not clarified in the bibliography or in the supplemental materials.

Regardless of the methodology involved in declaring a 10,000-year time depth, if we assume that the non-Austronesian languages do represent a very ancient lineage in Island Melanesia, then there is a major methodological flaw in the design of their study involving the comparison of uneven datasets.

The comparison sample, the languages that are not "Papuan languages of Island Melanesia," represent a much smaller time depth, at the most generous estimate 3,500 years, and would therefore be expected to display less diversity than the other sample.[8] Dunn et al. note that the Austronesian languages with which they make their comparison are all part of the Oceanic subgroup—further, they are all Western Oceanic languages within that subgroup. Given that the Austronesian family is perhaps the best understood large language family outside Indo-European, the reasons for not including languages that would allow a greater genetic sample of this family are obscure. An analogy of this methodological error would be to compare features of Finnish and its relatives with the North Germanic (i.e., Scandinavian) branch of Indo-European with which they are in contact. A feature such as "Language shows agreement for S and A" would be ranked positively for Finnish, but negatively for the Germanic languages. An analysis of this would conclude that "Language shows agreement for S and A" is a feature on which the Finno-Ugric languages and the Indo-European languages differ. Even the most cursory examinations of other Indo-European languages, even within the Germanic family, would reveal that agreement is a common feature of these languages as well, and that the comparison with North Germanic is not representative of the family that contains it (that is, Germanic, or Indo-European) when the family is considered as a whole.

Therefore we propose that comparison with Austronesian languages should include representative Austronesian languages from a wider geographic region in order to obtain an idea of the degree of diversity of these features that can be expected in a family over a 10,000-year (in the Austronesian case, 5,500-year) time frame.

8. The fact that the Oceanic languages in the sample are all members of the Western Oceanic group means that the time depth is in fact less than 3,500 years.

5. WHAT DO THE SELECTED FEATURES ACHIEVE? We noted above that Dunn et al.'s reported results do not clearly differentiate between a well-established group of languages which originate from a common ancestor about 3,500 years ago and a group whose status is in dispute, and which is claimed to originate at least 10,000 years ago. Before considering the status of the putative East Papuan group, it therefore is useful to consider how successful this dataset is in making the basic discrimination between Oceanic and non-Austronesian languages, and between Austronesian languages more widely and non-Austronesian languages.

To this end, we have examined the distribution of each of the characters across the two groups of languages. Where the proportion of languages with (or without) a particular feature varies to a significant extent between the two populations, we can consider this grammatical feature to be one that usefully marks the difference between the Oceanic and non-Austronesian languages. The features in Dunn et al.'s dataset that show a significant difference (measured by both t-test and chi-square test) are shown in table 6. Nine features show highly significant distributions ($p < 0.01$), and another 13 are significant, but not to the same degree ($p < 0.05$). True to the bias in the features selected, 15 of these 22 features are ones that are present more frequently in the non-Austronesian languages. Of these 22 features, 12 are what we call "categorial": the feature is either present in all of the sample(s)

TABLE 6. FEATURES THAT DISTINGUISH THE OCEANIC AND NON-AUSTRONESIAN POPULATIONS IN DUNN ET AL.'S SAMPLE

	FEATURE NO.		OCEANIC (% OF LANGUAGES)	NON-AUSTRONESIAN (% OF LANGUAGES)
$p < 0.01$	**28**	dual marked on noun	0%	**64%**
	29	plural marked on noun	33%	93%
	33	concord beyond the NP	0%	**73%**
	38	multiple possessive constructions	**100%**	60%
	39	prefix-marked possession	0%	**60%**
	40	suffix-marked possession	**100%**	40%
	51	prepositions	**100%**	46%
	65	verb agreement varies according to verb class	6%	80%
	79	conjugation classes	13%	67%
$p < 0.05$	**16**	inclusive/exclusive distinction	**100%**	67%
	19	2nd and 3rd persons conflate	0%	**27%**
	24	noun declensions determined by number	0%	**33%**
	46	collective nouns	0%	**55%**
	47	adjectives function as verbs	67%	17%
	58	S marked by suffix	19%	64%
	63	O marked by prefix	0%	**27%**
	70	person and number marked separately in the verb	6%	47%
	94	'give' is irregular	0%	**46%**
	106	ergative/absolutive morphology in basic constructions	6%	46%
	110	syntactic conflation of S and O	0%	**33%**
	118	language has complement clauses	93%	50%
	120	bound affix/clitic marks causation	86%	43%

from one population, or else is entirely absent from the other population; these features have been shown with bolding on the relevant population. In no case is there a grammatical feature that is universally present in one population, and universally absent in the other: the closest approximations to this ideal are found in feature 33, concord beyond the NP, and feature 40, suffix-marked possession. Many of the categorial features cannot be used as diagnostics, in any sense, because the rate at which they are attested is too low: while feature 19, the conflation of second and third persons, is attested only in the non-Austronesian part of the sample and nowhere in the Oceanic languages sampled, it is attested in only just over one quarter of those languages (four of the 15 languages, Mali, Rotokas, Savosavo, and Yélî-Dnye).

Despite the issues in independence of features, and low representation in one population or another, it is true that these features do distinguish the Oceanic languages from the non-Austronesian languages in the sample. Is this, then, a significant result?

We argue that it is not. First, the differences among the allegedly related non-Austronesian languages are *much* greater than the differences between the Oceanic languages. Table 7 shows the 31 languages of the sample ranked according to how they behave with respect to the features listed in table 6. The scores in the first column represent the number of positive values for features that are significantly more common in non-Austronesian languages, while the second column shows positive values for the features that are more common in Oceanic languages. The third column is calculated by subtracting the value for "Oceanic features" from the "non-Austronesian features" value, and shows the difference between the two scores, giving an overall ranking of how close any given language is to an Oceanic or Papuan ideal 'type'. Note that the greatest break in this column is not between the Oceanic and the non-Austronesian languages, but between Yélî-Dnye and the other languages.

It is clear that the variation within the non-Austronesian group is considerably greater than that within the Oceanic group. The range of scores in the rightmost column is 21 in the one case (from Yélî-Dnye at 14 to Bilua at –7) and 9 in the other case (from Kilivila at –11 to Takia, Banoni, Gapapaiwa, and Tungag at –20), with standard deviations of 5.6 and 2.6. In addition, we can see that there is a clustering toward the Oceanic end of the scale, such that six non-Austronesian languages (Kol, Anêm, Ata, Touo, Bilua, and Buin) have negative scores, and four of these (Ata, Touo, Bilua, and Buin) fall on the Oceanic side of the midpoint of the range (which is at –3). It is worth noting that this group includes at least one language from each of the three major geographic groupings in the sample: Ata is from the Bismarck Archipelago, Buin is from Bougainville, and Bilua and Touo are from the Solomons. This distribution suggests that the Oceanic languages form a coherent group typologically, but that the non-Austronesian languages do not. Indeed, several of the non-Austronesian languages have an Oceanic look to them on this measure.[9]

Second, and most significantly, all that we have shown in table 6 is that, given a set of features that are known to be significant in distinguishing the languages in the region, we can distinguish the languages in the region. This is, of course, circular argument.[10]

In the following sections, we examine the diagnostic power of these features on the micro- and the macrolevel. First, we present a small case study, examining the structural features used by Dunn et al., showing that in this case geography plays as much a role as

TABLE 7. RANKING OF LANGUAGES ON THE BASIS OF THE FEATURES SHOWN IN TABLE 6

LANGUAGE		NON-AUSTRONESIAN FEATURES	OCEANIC FEATURES	DIFFERENCE
Yélî-Dnye	non-AN	18	4	14
Nasioi	non-AN	13	6	7
Kuot	non-AN	14	7	7
Mali	non-AN	13	8	5
Savosavo	non-AN	12	8	4
Lavukaleve	non-AN	13	9	4
Motuna	non-AN	12	9	3
Rotokas	non-AN	10	9	1
Sulka	non-AN	11	10	1
Kol	non-AN	10	11	−1
Anêm	non-AN	10	11	−1
Ata	non-AN	8	12	−4
Touo	non-AN	9	13	−4
Buin	non-AN	6	11	−5
Bilua	non-AN	7	14	−7
Kilivila	Oceanic	4	15	−11
Jabêm	Oceanic	2	15	−13
Taiof	Oceanic	2	17	−15
Bali	Oceanic	1	18	−17
Kairiru	Oceanic	1	18	−17
Roviana	Oceanic	1	18	−17
Sisiqa	Oceanic	1	18	−17
Nalik	Oceanic	2	19	−17
Kaulong	Oceanic	0	18	−18
Siar	Oceanic	1	19	−18
Kokota	Oceanic	1	20	−19
Sudest	Oceanic	1	20	−19
Takia	Oceanic	0	20	−20
Banoni	Oceanic	0	20	−20
Gapapaiwa	Oceanic	1	21	−20
Tungag	Oceanic	1	21	−20

9. It is true that these figures (and our interpretation of them) must be affected by the position of Yélî-Dnye as an outlier. We would suggest that the fact that the most geographically separated of the non-Austronesian languages should behave in this way is already a cause for suspicion, especially when we compare Yeli-Dnye with its close geographical neighbor, Sudest, which is a central member of the Oceanic grouping in table 7. If we remove Yélî-Dnye from the table, the conclusions still hold: the standard deviation of the final values for the remaining non-Austronesian languages is still dramatically higher than the Oceanic languages, 4.5, over a range of 14 points; the recalculated midpoint for the whole sample is –6, which still allows one non-Austronesian language, Bilua, on the Oceanic "side" of the table.

10. Dunn et al. make further claims about their selection of features, which turn out not to be true. They claim that features were also selected based on their being "features that would typically be described in a published sketch grammar," and that "traits invariant in the region (either entirely absent, such as polysynthesis or proximate/obviative case distinctions; or present in all the languages, such as the existence of a word class of verbs) were not coded." As discussed in section 3, a number of the features in the sample do *not* vary across the languages in the region, and so these features are not helpful in differentiating the languages (though they do indicate to some extent the degree to which an unqualified analysis of the distribution of typological features is *not* useful in distinguishing language descent). To claim that features such as "tail-head linkage" is one commonly found in published sketch grammars is easily disproved, and the feature "S=O in complex constructions" definitionally requires the description of complex constructions, something typically absent in 22-page sketch grammars.

does putative affiliation in the distribution of these features. Then we examine more closely the "significant" features from the perspective of Austronesian languages beyond the Oceanic subgroup, a methodology that was argued for in section 4.

5.1 THE FEATURES IN MICROAREAL PERSPECTIVE: NEW BRITAIN AND NEW IRELAND. In this section, we present a detailed examination of Dunn et al.'s data for a small, geographically motivated sample of languages. The languages we examine are both Oceanic and non-Austronesian languages from the Bismarck Archipelago and from the northern part of Bougainville: New Ireland (Tungag, Nalik, and Siar [all Oceanic] and Kuot [non-Austronesian]), two languages from the east end of New Britain (Mali and Sulka [both non-Austronesian]), and two languages from the north of Bougainville (Taiof [Oceanic] and Rotokas [non-Austronesian]).

For this subsample of languages, 40 of the features from Dunn et al.'s set of 125 lack data points, and these features will be ignored in the discussion. A further 21 features are uninformative, with all languages showing the same value (for 10 features, all languages have the value 1, for 11 they all have the value zero). Of the remaining 64 features, four split the languages in this subsample into groups of four, as shown in table 8.[11]

Two of these characters define groups that contain equal numbers of Oceanic and non-Austronesian languages (characters 20 and 60), and are therefore not discriminating at all between the two groups. One character (65) groups three non-Austronesian languages with an Oceanic language. As seen in table 6, this character is one that shows a significant difference in occurrence between the two groups. We note that in the group defined by this character the Oceanic language (Nalik), which would not be expected to have a positive value, is the close geographical neighbor of one of the non-Austronesian languages (Kuot), and borrowing seems a plausible scenario. A single character separates the Oceanic and non-Oceanic languages in this subsample, and this is character 29—Plural Marked Noun ("plural number can/cannot be marked on the noun itself"). The four Oceanic languages are coded as not allowing plural marking on nouns, while the non-Oceanic languages do allow such marking. This is a character that shows a significant difference in its distribution, but that nevertheless has a positive value in one-third of the Oceanic sample. This again suggests a trait that has diffused across an area.

When we look at the characters that define groups of three languages, the evidence for grouping the non-Austronesian languages appears stronger. Kuot, Mali, and Sulka are isolated as a group by four characters. They lack suffixes marking possession (character 40), they have verbal prefixes marking both S and A (characters 59 and 61), and they lack causatives formed by bound affixes or clitics (character 120). Of these characters, only 40

TABLE 8. FEATURES THAT GROUP FOUR LANGUAGES

LANGUAGES	NUMBER OF FEATURES	FEATURE CODE
Nalik + Kuot + Mali + Rotokas	1	65
Nalik + Kuot + Taiof + Rotokas	1	60
Kuot + Siar + Taiof + Rotokas	1	20
Kuot + Mali + Sulka + Rotokas	1	29

11. The table shows the languages which have the value 1 for the character in question. The complementary set is, of course, an equally valid grouping.

has a significant difference in distribution. This character has a complete set of data for the whole sample, with all Oceanic languages having the positive value, and 40 percent of the non-Austronesian languages also having a positive value. Suffixal marking for possession is well attested in Austronesian languages elsewhere, and it cannot be ruled out that this feature spread from Austronesian languages to their non-Austronesian neighbors. In such a scenario, the fact that the four non-Austronesian languages in the current subsample all lack the feature can be seen as an areal phenomenon.[12] The presence of verbal prefixes marking S and A is also not uncommon in Austronesian languages elsewhere; indeed it is one of the criterial characteristics of a typological grouping identified by Himmelmann (2005:175) in the non-Oceanic Austronesian languages. Three characters define a group containing Kuot, Mali, and Rotokas (characters 28, 33, and 79), and for all three of these there are significant differences in distribution. Further, characters 28 ("dual marked on noun") and 33 ("concord beyond the NP") are categorial: neither occurs in an Oceanic language in the whole sample. Two characters (characters 43 and 44) define a group including Mali, Sulka, and Rotokas, but the distribution of these two characters does not show a significant difference. One group of three Oceanic languages is defined by a single character (character 41—"marked possessor"), again a character whose distribution across the sample is not significant.

Several groups of two languages are isolated from the subsample by one or more characters. Such groupings have little value in examining the power of these characters to discriminate between Oceanic and non-Austronesian languages, but it further reflects on the choice of characters that more exclusively non-Austronesian groups are isolated (five groups of two as against one purely Oceanic group and three mixed groups), and that the non-Austronesian groups tend to be supported by more characters. These tendencies were also evident in the groups of three languages discussed above, and it is clear that the characters used tend to isolate non-Austronesian languages, while Oceanic languages tend to share more character values.

Our detailed examination of the data for this subsample of eight languages brings out the following points:

- the Oceanic languages show a high degree of commonality;
- there is some evidence for the influence of geographic factors when we look at groups of languages that are isolated by sharing some features;
- many patterns that we are observing in the data may be artifacts of the way in which the features were chosen.

In the following section we examine all the languages in Dunn et al.'s survey, focussing on those features that show statistically significant different distributions across the Oceanic languages compared to the non-Austronesian languages.

5.2 THE "SIGNIFICANT" FEATURES IN MACROPERSPECTIVE. In this section we examine the "significant" features from a macroareal perspective. Departing from the small region included in Dunn et al.'s study, we examine the features seen in table 6 in the wider context of the region, looking at their occurrence in the Aus-

12. In addition, the well-known cross-linguistic preference for suffixing weakens the evidentiary value of this feature (Dryer 2005).

tronesian family more broadly and in other non-Austronesian languages from New Guinea and Australia.

28 Dual marked on noun. This feature appears to be a strong one for distinguishing the languages, and the lack of attestation is true of the Austronesian languages west of New Guinea as well as those to its east. We note, however, that this feature is not independent of feature 29, because cross-linguistic studies have shown that a language with dual marking necessarily also displays plural marking. We also note that this feature is well attested in other non-Austronesian languages of New Guinea, as shown in the Abinomn and Yimas examples.

(1) ABINOMN NUMBER AFFIXATION

	SINGULAR	DUAL	PLURAL
'bandicoot'	aine	aine**rom**	aine**kon**
'tree kangaroo'	we	we**rom**	we**kon**
'praying mantis'	tig^{w}rere	tig^{w}rere**rom**	tig^{w}rere**kon**
'house'	pr	pr**dom**	pr**kon**
'swamp'	ok^{w}i	ok^{w}i**rom**	ok^{w}i**gon**
'father's father'	moi	moi**rom**	woi**gon**
'pot'	jek	jek**rom**	jek**igon**
'jungle'	g^{w}ek	g^{w}ek**rom**	g^{w}ek**igon**
'younger brother'	ai	ai**rom**	a**kon**
'night'	siwi	siwi**rom**	siw**kon**
'star'	skin	ski**rom**	ski**di**
'cassowary'	komosin	komosi**rom**	komosi**di**
'owl'	weimn	weim**rom**	weim**ti**
'river tortoise'	fan	fa**rom**	fa**ti**
'louse'	jen	jend**rom**	je**ti**
'prawn'	beresmin	beresmind**rom**	beresmi**di**
'knife handle'	tam	ta**brom**	ta**tom**
'sago pudding'	midam	mida**brom**	mida**tom**
'bamboo knife'	abisiam	abisia**brom**	abisia**som**
'toe'	g^{w}esiam	g^{w}esia**brom**	g^{w}esa**som**
'centipede'	sm	s**brom**	s**kr**
'headband'	k^{w}etam	k^{w}eta**brom**	k^{w}eta**kr**

(2) YIMAS NUMBER AFFIXATION

	SINGULAR	DUAL	PLURAL
'sago palm'	tɨnum	tɨnum**ul**	tɨnn**i**
'coconut palm'	irɨpum	irɨpum**ul**	irɨpŋ**i**
'shell'	kaŋk	kaŋkɨ**l**	kaŋk**i**
'frog'	krayŋk	krayŋkɨ**l**	krayŋk**i**
'basket'	impramp	imprampɨ**l**	imprampa**t**
'bone'	tanɨmp	tanɨmpɨ**l**	tanɨmpa**t**
'axe'	awi	awɨ**l**	aw**mpɨt**
'breast'	nɨŋay	nɨŋay**l**	nɨŋa**mpɨt**
'knee'	trukwaw	trukw**ul**	trukw**ut**

'road'	yaw	y**ul**	y**ut**
'mouth'	antuk	antuk**ul**	antuk**at**
'bandicoot'	awruk	awruk**ul**	awruk**at**
'egg'	awŋk	awŋk**ul**	awŋk**wi**
'elbow'	mpunawŋk	mpunawŋk**ul**	mpunawŋk**wi**

Note, parenthetically, that the marking of [dual] in both Abinomn and Yimas is more consistent, formally, than is the morphology used to mark [plural].

29 Plural marked on noun. Marking plurality is more commonly associated with the non-Austronesian languages in the sample than with the Oceanic ones, though it is attested in fully one-third of the Oceanic sample. Related Austronesian languages west of New Guinea also show plural marking, as seen in (3–6). The Selaru and Alune forms are possibly, even probably, cognate, while the Tugun example clearly shows plurality marked by reduplication, raising a question of the interpretation of the feature that Dunn et al. code: they state that their feature 29 involves the opposition "plural number can be marked on the noun itself / plural number cannot be marked on the noun." While the Tugun example clearly does not involve any paradigm of number affixation, it does involve plural marking on the noun itself.[13] Example (6) shows that in the Austronesian languages of northern Borneo, plural marking is still a feature, as shown with the Tatana' data (the N in *po*N- represents a placeless nasal that assimilates to the following consonant [see various references from Newman 1984 to Blust 2004, including especially Pater 1999, 2001]).

SELARU (Coward 1990:23)

(3) kader-are
chair-PL
'(the) chairs'

TUGUN (Hinton 1991:51)

(4) Ni-ala ni-enu-enu.
3S-take 3S-RED-necklace
'She took her necklaces.'

ALUNE (Florey 2005:29))

(5) Ami alena-'e atue ami 'i-ebe-ru.
1P.EXCL narrate-APPL BEN 1P.EXCL 1P-friend-PL
'We told (the story) to our friends.'

TATANA' (Chan and Pekkanen 1989:73)

(6) … om poN-sumad mu do ngo-anak mu?
and IV-feed 2S.GEN DAT PL-child 2S.GEN
'(What will you eat) and feed to your children?'

This feature is not convincing as a separator of Austronesian and non-Austronesian languages. Elsewhere in New Guinea we also find plenty of examples of plurality being

13. Abbreviations used in glossing follow sources in all cases (adapted to the Leipzig conventions). The following list includes only abbreviations that are *not* standard abbreviations in the Leipzig Glossing Rules: H, NH: human, non-human; 6, 9: noun classes numbers (in Yimas); ACT, active voice; ANIM, animate class; DIM, diminutive; DRINK, drinking possession class; EDIBLE, edible possession class; GENERAL, general possession class; IV, instrumental voice; LM, lexical marker; PRO, pronoun; R, realis; RED, reduplication; SI, "subject" infix.

marked on nouns; in addition to the Abinomn and Yimas examples seen earlier in (1) and (2) showing singular-dual-plural paradigms, note the following singular-plural paradigms from Lani (Dani family, western New Guinea highlands) and One (Torricelli family, eastern Bewani ranges). The Torricelli languages, as well as the Lower Sepik languages to which Yimas belongs, have long been noted as languages that show singular:plural alternations (Laycock 1975, Laycock and Z'graggen 1975).

LANI

(7)				
	kwe	'woman'	kumi	'women'
	apuluk	'child'	apologwe	'children'
	wulaga	'boy'	wulogwe	'boys'
	kweliga	'girl'	kumulogwe	'girls'
	kuwagaluk	'short'	kuwagalogwe	'short.PL'

ONE

(8)				
	pino	'woman'	pini	'women'
	mala	'child'	meli	'children'
	toma	'stone'	tomu	'stones'
	plona	'daughter'	plonapi	'daughters'
	sou	'frog'	souwo	'frogs'
	suwol	'snake'	suwolwa	'snakes'

33 Concord beyond the NP. Unless Dunn et al. are consciously making a distinction between NPs and DPs, this feature effectively identifies a language with an agreement system in which gender (or noun class) is part of that agreement system. If so, then this is something that is widely attested in the non–Trans New Guinea family languages of New Guinea. The following examples illustrate concord in Skou, Yimas, and Bukiyip, unrelated languages from northern New Guinea (Skou belongs to the Greater Skou family, Yimas to the Lower Sepik family, and Bukiyip to the Torricelli family). In both Skou and Bukiyip, we can see that the difference between masculine and feminine (or feminine and nonfeminine) is indicated on the verb. In Yimas the different noun class of the objects (class 9, class 6) in the two clauses results in the use of different P-agreement prefixes.

SKOU

(9) Ke=angku=ing a ke=ti.
3S.NF=child=the 3S.NF=3S.NF:go
'The boy went.'

(10) Pe=angku=ing a pe=te.
3S.F=child=the 3S.F=3S.F:go
'The girl went.'

YIMAS

(11) Wanwa wa-ka-tar-wapi.
knife:9S 9S-1S-CAUS-sharp
'I sharpened the knife.'

(12) Murakl kla-ka-yamal.
paddle:6DU 6DU-1S-carve
'I carved the two paddles.'

BUKIYIP

(13) Énan n-a-leh.
3S.M 3S.M-R-cry
'He cried.'

(14) Okok kw-a-gak.
3S.F 3S.F=R-die
'She died.'

Selaru and many Austronesian languages of eastern Indonesia show a distinction between animate and inanimate, indicated by verbal prefix, instantiating concord beyond the NP as well.

(15) SELARU SINGULAR SUBJECT PREFIXES (Coward 1990:15)

	V-INITIAL	C-INITIAL	CC-INITIAL
1S	k-	k-w-	ku-
2S	m-	m-w-	mu-
3S.ANIMATE	y-	-y-	i-
3(S/P).INANIMATE	ky-	k-y-	ki-

And it is common in languages of Maluku for a distinction to be made between human and nonhuman. Again, this is seen in cross-referencing clitics attached to verbs and must count as concord beyond the NP, illustrated with examples from the Alune language of Seram Island (Florey 2001) in (16–19).

(16) Ela' inai-je i=ombe …
elder CLF-DET 3S.H-say
'My parent, s/he said …'

(17) Apa-le e=betu bei au.
pig-LM 3S.NH-get.up ABL 1S
'The pig got up from me.'

(18) Au susu=i pela le'i au be 'ite leu le 'petu-re pene'a.
1S milk=3S.H finish then 1S COMP 1P return.home because dark-DET PFV
'I finished breastfeeding him, then I (said): "We're going home because it's dark."'

(19) Au sabe tasi-anai 'ena leu-'we=le.
1S buy salt-DIM OBL take.home-APPL=3S.NH
'I bought a little salt to take home.'

38 Multiple possessive constructions. It is well known that languages of the Oceanic branch of Austronesian show multiple possessive constructions (Lynch, Ross, and Crowley 2002:40–43). These usually divide into what has been termed "direct" possession, in which a suffix on an inalienably possessed noun shows agreement with the possessor, and "indirect" possession, in which the suffix appears on a classifier indicating some salient property of the possessed noun. These two classes of possession are illustrated with Fijian data in (20–23). Direct possession can be seen in (20), while (21–23) show the three different classes of indirect possession.

FIJIAN

(20) ulu-mu
head-2S
'your head'

(21) na me-mu wai
ART DRINK-2S water
'your water' (to drink)

(22) na ke-mu dalo
ART EDIBLE-2S taro
'your taro (to eat)'

(23) na no-mu vale
ART GENERAL-2S house
'your house'

Among Austronesian languages other than those in the Oceanic group, we find that multiple possessive constructions are at best sketchily attested; some (but not all) languages immediately west of New Guinea show multiple possessive construction markers, such as can be seen in Selaru, but the Austronesian languages further west have no such multiple constructions, as evidenced by the examples from Tukang Besi and Indonesian.[14]

SELARU

(24) a. ara wasi-my asw-Vre
1P.EXCL GENERAL-1P.EXCL dog-PL
'our dogs' (Coward 1990:33)

b. hina-mw hahy-desy-ke
EDIBLE-2S pig-that-ART
'your pork there' (Coward 1990:34)

c. ity ama-t-ke
1P.INCL father-1P.INCL-ART
'our father' (Coward 1990:39)

TUKANG BESI

(25) a. 'obu=nto
dog=1P.GEN
'our dog'

b. wawu=nto
pig=1P.GEN
'our pork'

c. ama=nto
father=1P.GEN
'our father'

INDONESIAN

(26) a. anjing kita
dog 1P.INCL
'our dog'

b. babi kita
pig 1P.INCL
'our pork'

c. bapak kita
father 1P.INCL
'our father'

The existence of multiple possession classes is thus a feature of the Austronesian languages near, or beyond, New Guinea, but not of Austronesian languages as a whole. This phenomenon might even be an innovation of the Austronesian subgroup that contains Oceanic and those languages of Maluku that also show multiple possession; but it is certainly not something that can be reconstructed to any great time depth for Austronesian languages in general.[15]

39 Prefix marked possession. While prefixal possession is categorially not found in the Oceanic languages in the study, but is present for almost two thirds of the non-Austronesian languages, it is also in Austronesian languages from Maluku, such as Tugun, in (27) (see also Alune in [5]).

TUGUN (AUSTRONESIAN, EASTERN INDONESIA)

(27) a. u-mumu
1S.POSS-machete
'my machete'

b. lasoni-pei
mouse3S.POSS-feces
'mouse's feces'

This distribution suggests that prefixal (and prenominal) possession is a feature of the New Guinea region, a fact that has been noted with respect to the Austronesian languages of eastern Indonesia since Brandes (1884) (see also Donohue 2004a, van Hoëvell 1877).

14. The Selaru plural suffix is -V*re*, the V indicating that the suffix supports the syllabification of a final glide in the root. The Selaru word for 'dogs' is thus pronounced [asure].

15. Dunn et al. have separate characters that code an alienability distinction in possessive constructions (37) and the existence of multiple possessive constructions (38). Our reading of the definition given for character 38 ("two or more different possessive constructions") is that if a language has an alienability distinction (Dunn et al.: "distinction between semantic alienable and inalienable possession is present in the grammar [inalienable possession interpreted loosely as special grammatical treatment of a class defined by some of the prototypical targets of inalienable possession …]"), it must qualify as having multiple possessive constructions (a similar conclusion is evident in Nichols and Bickel 2005). If this interpretation is followed, then most of the Austronesian languages of Maluku would be considered to have multiple possessive constructions—see Florey (2005) for examples. The existence of this alienability distinction is well known, being one of Himmelmann's (2005) criteria for the "preposed possessor" type of Austronesian language found in eastern Indonesia.

40 Suffix-marked possession. Suffix-marked possession occurs in all the Oceanic languages in the study sample, but as seen in the immediately preceding discussion, it is not found in all Austronesian languages, even though there is a strong tendency for it to occur, almost certainly reflecting the morphology of Proto-Austronesian, which has been inherited in many of the daughter languages.[16]

However, many non-Austronesian languages also show suffixally marked possession, as seen in the following examples from north-central New Guinea showing both suffixes on the possessor and suffixes on the possessum.

	ONE	
(28)	auna	i-enu
	son	1S-GEN
	'my son'	

	SKOU	
(29)	kc=angku-nì=nc	
	3S.NF=child-1S.DAT=1S.GEN	
	'my son'	

51 Prepositions. Prepositions are found in all the Oceanic languages in the study. There are, nonetheless, many Austronesian languages that either lack adpositions altogether (such as Kéo, from southern Indonesia), or use (suffixal) case marking instead, such as Tobati (northern New Guinea, and Oceanic).

In Kéo (Baird 2002) all 'preposition'-like elements are simply serial verb constructions (some of which have grammaticalized to varying extents), and the same is true of related languages in southern Indonesia. In any event, the presence of prepositions (as opposed to postpositions) shows such a strong correlation with VO order at the clause level that it cannot be considered a wholly independent variable.

65 Verb agreement varies according to verb class; 79 Conjugation classes. We interpret feature 65 to mean that there are conjugation classes in the verbs based on agreement marking. If this is so, these two features are obviously heavily interdependent: if a language has verbs that show agreement that varies according to verb class, then that language must be coded as having conjugation classes. Despite this, four languages (Motuna, Touo, Bilua, and Nalik) are coded as having a positive value encoded for feature 65, but a negative value for feature 79.

Including "conjugation classes" as a feature begs the question of whether phonologically defined classes "count" as verb classes / conjugation classes, and whether irregular verbs count. If either of these conditions do count, then a great number of languages of the world must be reckoned to have positive values for 65/79; it is not clear if there are any (inflecting) languages that do not have some degree of irregularity or phonological conditioning. For instance, Tugun shows phonologically conditioned inflectional classes, as shown in (30) (Hinton 1991:76).

(30)		-#C…	-#V…	-#/…	'MAKE'	'TAKE'	'FEED'
	1S	u-	v-	v-	ugisan	vala	varo
	2S	o-	m-	o-	ogisan	mala	oaro
	3S		n-		gisan	nala	aro
	1P.EXCL	am-	am-	am-	amgisan	amala	amaro
	1P.INCL	it-	itt-	it-	itgisan	ittala	itaro
	2P	mi-	mir-	mi-	migisan	mirala	miaro
	3P	ra-	r-	ra-	ragisan	rala	raaro

16. Given the general preference for suffixation as opposed to prefixation cross-linguistically, this feature has little predictive value; see also fn. 12.

Here the differences in subject agreement prefixes force us to establish three classes of verbs, based on the initial segment in the verb root. Examples of a verb from each of these classes are shown in the right of (30), following the same arrangement used with the prefixes.

We have already seen inflectional information from Selaru in (15), which also shows (phonologically conditioned) variation in verbal agreement, and in (31) we present data from Roma (Steven 1991), which is analogous to both Tugun and Selaru (given the fact that Roma is geographically midway between these other two languages, this is not surprising).

(31)

	V-INITIAL	C-INITIAL	CC-INITIAL
1S	kaw-	a-w-	a-
2S	mw-	m-w-	mu-
3S	n-	n-	na-

At a more micro level, Tukang Besi includes two verbs that can count as irregular. In (32a–d) we see the (partial) inflections of two verbs, *hada* 'want' and *nde'u* 'not want'. *Hada* shows the regular behavior when prefixing (for S/A) and infixing (the 'S/A infix', which appears in irrealis clauses, among other uses). In (32c–d) we see that *nde'u* shows irregular behavior with infixation: rather than appearing aligned to the left edge of the root, the infix appears aligned to the left edge of the inflected verb. This difference cannot be ascribed to different phonological shapes, as can be seen by examining (32e–f); these examples have phonological shapes analogous to that of *nde'u*, yet they show regular infixation.

(32) a. ku-hada
1S-want
'I want'

b. ku-h<um>ada
1S-want<SI>
'I will want'

c. ku-nde'u
1S-not.want
'I don't want'

d. k<um>u-nde'u
1S<SI>-want
'I won't want'

e. ku-nduha
1S-fall
'I fall'

f. ku-nd<um>uha
1S-fall<SI>
'I will fall'

Another verb showing irregular behavior of the <*um*> infix is 'give'. In (33a–d) we see regular infixation of <*um*>; (33e–f) shows that with some *h*-initial roots the morpheme appears as substitution of the initial segment of the root with *m*. This is the allomorph found with *hu'u* 'give', but with 'give' we also see a vowel alternation in the root itself; such a vowel alternation is not attested in any other verbs.

(33) a. 'awa
get

b. '<um>awa
get<SI>

c. hada
want

d. h<um>ada
want<SI>

e. hesowui
wash

f. <m>esowui
wash<SI>

g. hu'u
give

h. <m>o'u
give<SI>

16 Inclusive/exclusive distinction. The presence of an inclusive/exclusive distinction is probably a good genetic marker for the Austronesian languages. It appears in 100 percent of the Oceanic sample in Dunn et al. The distinction is known to be lost in some

Austronesian languages, such as Tukang Besi (Donohue 1999a:113–14, Donohue and Smith 1998), but it shows a high degree of stability cross-linguistically (Nichols 1992).

Balancing this support for the use of the inclusive/exclusive distinction as an Austronesian feature is the fact that it appears in more than 50 percent of the non-Austronesian languages in the sample as well, and is widely, though erratically, attested in the non-Austronesian languages of mainland New Guinea (other than those in the Trans New Guinea family), as pointed out in Donohue (2005).[17]

19 2nd and 3rd persons conflate in the plural. Pronoun systems with this characteristic are widely reported in the eastern New Guinea highlands; see, for example, the description of Hua given by Haiman (1980). The feature does, however, discriminate between Austronesian and non-Austronesian languages.

24 Noun declensions determined by number. Noun classes in which number plays a role in determining membership are a cross-linguistically rare occurrence, and the complete absence of this feature in the Oceanic languages, compared to its appearance in five out of 15 of the non-Austronesian languages in the sample in Dunn et al., suggests that it is a strong separator of the Oceanic languages from the non-Austronesian languages of the area. This feature is not unique to the languages in Island Melanesia, however; Donohue (2000) documents the classification system of Orya in north-central New Guinea, where dual functions as a noun class as well as masculine and feminine, for instance.

46 Collective nouns. It is not clear how this category is different from feature 35, numeral classifiers. The description given by Dunn et al. ("words for particular amounts of a thing, e.g., ten possums, ten canoes") does not allow us to evaluate this adequately.

47 Adjectives function as verbs. This feature is probably robust in differentiating Austronesian languages from non-Austronesian ones, if we leave aside the issue of precategoriality. Adjectives do not function as verbs across a very large number of southwestern Austronesian languages, where they are a separate lexical category (e.g., Donohue 1999b, forthcoming), and indeed across most of New Guinea and Australia we also find adjectives not functioning as verbs.

Examples of nonverbal adjectives are shown in (34–35), contrasting verbal and adjectival predicates. In Skou we can see that the complex agreement complex—consisting of a proclitic as well as a prefix on the verb—that marks the verbal predicate in Skou (Donohue, 2003) is absent from adjectival predicates. In One we can see that agreement prefixes, which are one of the characteristics of verbs in this language, are absent from adjectives.

SKOU

(34) a. Í pɛ=w-íng toe ɛ tuɛ kong.
snake 3S.F=3S.F-crawl 3S.F:come 3S.F.be 3S.F.do down
'A snake is crawling this way down there.'

b. Í pɛ=bà fɛ̀ng.
snake 3S.F=ANIM bad
'Snakes are bad things.'

17. Peripherally, we may note the near-universal presence of an inclusive/exclusive distinction in the languages of Australia.

ONE

(35) a. Suwol suwe y-eri po'u.
snake come.out 2/3s-come.up closed.hole
'The snake is crawling out of its hole.'

b. Suwol sa fampono (mana).
snake TOP bad person
'Snakes are bad (thing).'

In the languages of Australia a distinction between verbal and "adjectival" predicates (usually referred to as nominal predicates in the Australianist tradition) is almost, if not completely, universal.

58 S marked by suffix. More non-Austronesian languages than Austronesian languages in the Dunn et al. sample show S marked by suffix (though still less than half). Suffixal marking for subjects is widely reported in non-Austronesian languages of New Guinea (e.g., Foley 1998; Wurm, Laycock, and Voorhoeve 1975), and simply reflects a global preference for suffixation (Dryer 2005), a preference that is not dominant in the Austronesian family. If we examine the historical origins for subject prefixing in Austronesian languages, however, we see a very different story. What are now prefixes in many Austronesian languages of the Oceanic group were for most of their history enclitic forms—more suffixal than prefixal. As first outlined in Starosta, Pawley, and Reid (1982), and elaborated on by Blust (1993) and Himmelmann (1996), a series of second-position clitics that more and more frequently appeared on preverbal auxiliaries were transferred from being preverbal enclitics on auxiliaries to being preverbal proclitics, and later prefixes, on the verbs themselves.[18]

(36)

	NO AUXILIARY		AUXILIARY		
Stage 1	V	Pros, NPs	Aux	V	Pros, NPs
Stage 2	V=**clitic**	NPs	Aux	V=**clitic**	NPs
Stage 3	V=**clitic**	NPs	Aux=**clitic**	V	NPs
Stage 4	—		Aux=**clitic**	V	NPs
Stage 5	—		Ø=**clitic**	V	NPs
Stage 6	prefix-V	NPs	**prefix**-V	NPs	

The point is that any linguistic sample represents a snapshot in time: if we had chosen an earlier date, at, say, stage 2 in (36), we would conclude that the Austronesian languages do not show S marked by prefix, but by suffix (assuming that the clitic-to-affix development is as well attested on the other side of the verb).

63 O marked by prefix. The marking of objects by prefix is a characteristic of the vast majority of languages across New Guinea and northern Australia, as noted by Wurm, Laycock, and Voorhoeve (1975); Foley (1998, 2000); Haiman (1980); Evans (2003); and is not characteristic of Austronesian languages. Judging by the fact that only just over one quarter of the non-Austronesian languages in the sample show objects marked by prefix, we can only conclude that, in this respect, three quarters of the Island Melanesian languages are quite unlike most of the languages of (mainland) New Guinea and northern Australia, and more closely resemble the Austronesian languages.

18. For another study of the unreliability of prefixing and suffixing typology as a stable historical marker, see Harvey, Green, and Nordlinger (2006).

70 Person and number marked separately in the verb. This feature is reported in widely separated (non-Austronesian) languages of eastern New Guinea (e.g., Koiari [Dutton 1996]). To cite just a single example, Kanum (southern New Guinea) shows a split in the places in which agreement features appear in the verb. The final suffix in (37d–e) codes the plural number of the subject, as well as future tense; note that person is not coded here (in [37f] there is a more explicit 3P.FUT suffix); singular subjects have no agreement suffix. For second person subjects there is no further agreement on the verb, but for first and third persons we can see the prefix *r-*; this does not encode number features, but does encode person features (or, more specifically, the absence of a positive second person feature).

(37) a. Ngkay tr s-r-wr-nt.
1S.ERG tooth FUT-**1/3.FUT.SUBJ**-bite-FUT
'I will bite him/them.'

b. Mpay tr s-wr-nt.
2S.ERG tooth FUT-bite-FUT
'You will bite him/them.'

c. Pyengkw tr s-r-wr-nt.
3S.ERG tooth FUT-**1/3.FUT.SUBJ**-bite-FUT
'He/she will bite him/them.'

d. Ny-nta tr s-r-wr-nt-ey.
1P-PL.ERG tooth FUT-**1/3.FUT.SUBJ**-bite-FUT-**PL.SUBJ**
'We will bite him/them.'

e. Mpw-nta tr s-wr-nt-ey.
2-PL.ERG tooth FUT-bite-FUT-**PL.SUBJ**
'You (PL) will bite him/them.'

f. Py-nta tr s-r-wr-nt-eme.
3-PL.ERG tooth FUT-**1/3.FUT.SUBJ**-bite-FUT-**3P.SUBJ**
'They will bite him/them.'

It is not clear how suppletive verb stems would be excluded from being counted for this feature. In the Skou examples in (38) we can see that person is marked by a proclitic, bundled in a single morpheme with number and gender. Number is additionally marked by the choice of verb stem. Should this count as separate marking of person (in the proclitic) and number (in the verb stem, as well as in the proclitic), or not?

SKOU

(38) a. Pále pe=wung.
pig 3S.F=SG.die
'A pig died.'

b. Pále te=bìng.
pig 3P=PL.die
'Pigs died.'

Numerous other examples can easily be garnered from grammars of mainland New Guinea languages; the phenomenon is also found in Oceanic languages further east in the Pacific in southern Vanuatu (e.g., Lenakel, as described in Lynch 1978), adding to the argument in Blust 2005 that there is a Papuan substrate in these languages.

94 'Give' is irregular. As seen earlier in (33), the verb for 'give' in Tukang Besi shows irregular morphophonemics: *hu'u* + <*um*> = *mo'u*. Looking at morphosyntax, *beri* 'give' is the only Indonesian verb for which the recipient is the object.

INDONESIAN

(39) a. Dia mem-beri saya hadiah.
3S ACT-give 1S present
'She gave me a present.'

b. Dia meng-[k]irim hadiah kepada saya.
3S ACT-send present to 1S
'She sent a present to me.'

Other examples outside the Austronesian languages are easy to find (for instance, in Skou *ké leng* 'give' is the only underived predicate that takes a postverbal object as well as a preverbal one). As Newman (1996) shows, it is normal for the verb for 'give' to show irregularities.

106 Absolutive morphology in basic constructions. This feature is a good marker for the non-Austronesian languages, although less than 50 percent of non-Austronesian languages in the sample are coded as having the feature. It is also well attested in New Guinea generally (Foley 1986), and is common in Australia (Dixon 1980). Despite ongoing debate concerning the languages of the Philippines (see Foley, to appear, for a summary), ergative structuring is not common in the western Austronesian languages. However, several eastern Oceanic languages (in the Polynesian group) are generally analyzed as having absolutive morphology (Lynch, Ross, and Crowley 2002).

110 Syntactic conflation of S and O. It is not a simple matter to argue about whether or not there are absolutive pivots in a language. English, almost a "poster boy" for the nominative camp of syntactic alignment, shows constructions that conflate an absolutive category, rather than a nominative one, whereby an A or a P appears with the preposition "of" (*the sleeping **of** the children*, *the eating **of** the spinach*), but an A uses "by" (*the eating of the spinach **by** the children*).[19] Donohue (2004b) discusses the way in which floated quantifiers can be *expected* to show an absolutive pivot (restriction) for interpretation depending on their constituency, even in otherwise strictly nominative-accusative languages such as Japanese. It is unlikely, however, that much of this sort of material will make its way into a 22-page grammar sketch, such as represents most of the Oceanic languages in the sample, and so absolutive pivots would remain overlooked.

118 Language has complement clauses. We find it hard to imagine what a language without complement clauses might look like (pace Everett 2005). If Dunn et al. refer to the presence of overt complementizers, or clauses that are overtly distinct (in terms of morphosyntax) from main clauses, then that should be stated.

This is not a trivial question. We must ask whether, for instance, there is a complement clause in (40–42). There are not complementizers, or special subordinate verb or clause forms, but we can establish the nonmain status of these embedded clauses without too much difficulty. The fact that some complements involve complete nominalization, as in (42), and that nominalization is indeed the normal strategy used in complementation in some languages, does not stop the construction from being a complement clause (albeit nominalized) (Givón 2001).

19. These are not minor categories, either; the well-known distribution of *by* and *of* in nominalizations has been discussed in detail since at least Chomsky (1970).

(40) I know [they're going to attend my wedding].

(41) It is also worth recognising [conservative women have abortions, too,] … (*Sydney Morning Herald,* January 13, 2007)

(42) I suspect [their involvement in the failed coup].

Despite the absence of an overt complementizer, it is clear that syntactically there is an embedded clause; if this is not called a complement clause, then we are dealing with terminology and not with linguistic facts.[20]

120 Bound affix/clitic marks causation. As Nichols (1992) shows, the presence and type of valency-changing morphological processes in a language is genetically stable, and this feature does serve to separate the Oceanic (and other Austronesian) languages from the non-Austronesian languages of New Guinea (and, to a lesser extent, Australia—a causative affix reflecting *-ma is common in many Pama-Nyungan languages), including the non-Austronesian languages in the Dunn et al. sample. It does not, however, identify the non-Austronesian languages of Island Melanesia (or beyond) as a genetically coherent group: the absence of a character cannot be taken as showing as much significance as the presence of one. The languages that do not have a bound affix or clitic marking causation are simply a negative group, not a positive one.

5.3 SUMMARY. We showed in table 6 that Dunn et al.'s dataset includes 22 features that show statistically significant differences in their distribution between Oceanic and non-Austronesian languages in their sample area. We have now evaluated these features in the light of the distribution of these structural traits in non-Austronesian languages beyond those in Island (eastern) Melanesia, and in Austronesian languages beyond the Oceanic subgroup. Table 9 summarizes the discussion in the preceding sections. It is worth pointing out that the distribution of the non-Austronesian languages that display the "distinguishing" features is not random: we can identify clear "hotbeds" in which various traits appear. Some traits do appear to show a random dispersal among the non-Austronesian languages in the sample, while others are easily geographically defined. Other than plural marking, dual marking is the most widely attested such feature, but it is absent from central and western New Britain; the use of prefixes to show agreement for an O on the verb is found only in the Solomons languages, and in Kuot (where a great variety of prefixes, infixes, and suffixes are used, depending on verb class); the separation of pronominal features for a single argument in verbal morphology is prevalent in eastern New Britain, plus a number of languages of Bougainville, as well as Lavukaleve; languages with an irregular 'give' verb are concentrated in the Solomons; languages with absolutive morphology are concentrated on New Britain; and the use of a bound morpheme to mark causative is limited to the Solomons and Bougainville.

In sum, we are left with between one and six (depending on how we wish to count things) structural traits that might be used to distinguish the languages genetically. Note that four of these, numbers 40, 47, 51, and 120, are characteristic of Austronesian languages, and so merely provide the necessary criteria for not considering the Island

20. Mark Libermann discusses similar examples of subordination without overt embedding in English (such as 'You ever see it they take it out of the shell?') in a Language Log posting archived at http://itre.cis.upenn.edu/~myl/languagelog/archives/003162.html (accessed 02/16/07).

Melanesian languages in the sample to be Austronesian (from this typological perspective). The remaining two features, the presence of ergative/absolutive morphology in basic constructions and the presence of a dual marked on the noun itself, were found in 46 percent and 64 percent (respectively) of the non-Austronesian languages in the sample.

The appearance of ergative/absolutive morphology is anything but uncommon in the (non-Austronesian) languages of the New Guinea region; Li and Lang (1979) discuss the question in detail for Enga, and similar examples are easy to find across the highlands and adjacent areas of New Guinea.

The presence of a dual marker is perhaps more convincing, especially because this feature is categorial: none of the Oceanic languages in the Dunn et al. sample shows this feature, it is not strongly attested elsewhere in the Austronesian world, and it is not common among the non-Austronesian languages of the New Guinea region beyond those in the Dunn et al. sample. This feature, however, is found positively in only two-thirds of the non-Austronesian languages in the Dunn et al. sample, nine languages. Those nine languages are not evenly distributed: they are Motuna, Rotokas, and Nasioi, all the languages in the sample from Bougainville (no data was available for Buin); Touo,

TABLE 9. FEATURES THAT DISTINGUISH THE AUSTRONESIAN AND NON-AUSTRONESIAN POPULATIONS IN DUNN ET AL.'S SAMPLE

FEATURE		REMAINS SIGNIFICANT?	NOTES
16	inclusive/exclusive distinction	no	distinction absent in some AN languages, present in many non-AN languages
19	2nd and 3rd persons conflate	no	widespread in New Guinea
24	noun declensions determined by number	no	found in widely dispersed non-AN languages
28	dual marked on noun	(yes)	
29	plural marked on noun	no	found in AN and non-AN languages
33	concord beyond the NP	no	found in AN and non-AN languages
38	multiple possessive constructions	no	not found in most AN languages
39	prefix marked possession	no	found in AN and non-AN languages
40	suffix marked possession	(yes)	not found in all AN languages
46	collective nouns	?	uninterpretable
47	adjectives function as verbs	(yes)	
51	prepositions	(yes)	
58	S marked by suffix	no	found in most non-AN languages; part of the grammaticalization into prefixes attested in AN languages
63	O marked by prefix	no	widely found in non-AN languages
65	verb agreement varies according to verb class	no	all languages?
70	person and number marked separately in the verb	no	found in widely dispersed non-AN languages
79	conjugation classes	no	all languages?
94	'give' is irregular	no	found in AN and non-AN languages
106	ergative/absolutive morphology in basic constructions	yes	
110	syntactic conflation of S and O	no	data not sufficient to decide
118	language has complement clauses	no	question not meaningfully defined
120	bound affix/clitic marks causation	(yes)	

Lavukaleve, and Savosavo from the Solomons; Yélî-Dnye from Rossel Island; and Kuot and Mali from the eastern side of the New Britain/New Ireland chain. In other words, the distribution of dual markers shows a strong areal trend, appearing in the east, and being absent from the languages in the west (Kol, Ata, Anêm, and Sulka, from New Britain). If, as Dunn et al. suggest, the logical direction of migration for an ancient ancestor to a putative Proto–Island Melanesian linguistic homeland is from the west, we can only conclude that either the New Britain languages closest to New Guinea all (coincidentally?) lost dual morphology, or else the dual morphology found in the east is attributable to an "eastern Melanesian" areal trend. In any case, the geographic distribution of this feature does not suggest a pan–Island Melanesian unity.[21]

6. ADDITIONAL ISSUES. On the basis of the preceding discussion of the features that form the basis of Dunn et al.'s study, we propose that there is an alternative, and simpler, default hypothesis that can, and should, be advanced to account for the data: the "Island Melanesian" languages that Dunn et al. contrast with the Oceanic languages in the area represent a number of unrelated linguistic lineages, either the descendants of ancient linguistic entities in a contact area prone to the diffusion of linguistic function (while forms remain distinct, as pointed out by Terrill 2002), or else recent linguistic innovation, either creolizations at least partly based on languages that are no longer extant, or else imperfect learning of now extinct lineages. In addition to the problems relating to the empirical base of the analysis, we also wish to draw attention to two further methodological issues: the nature of unrooted trees as a representation of relationships, and the lack of negative controls in this study.

An unrooted tree, such as that shown by Dunn et al., does not allow for two (or more) languages to show no relationship. All it is capable of is demonstrating which languages are, according to the criteria given, more closely related, or less closely related, to each other. In other words, given a hypothetical case with three unrelated families, they will necessarily appear in a single tree, apparently "related." An unrooted tree representing (for instance) English, Dutch, German, Basque, Mandarin, Cantonese, and Swahili can be constructed, and will show them all in one tree—if that is the way the researchers wish to represent the "relationship" between the languages.

In Dunn et al.'s supporting material (available online at http://www.sciencemag.org/cgi/content/full/309/5743/2072/DC1) a single unrooted tree contains all of the languages in their sample; they observe that the non-Austronesian languages of New Britain appear to be related as closely to a branch of Oceanic languages as to the non-Austronesian languages of other areas in the sample. In addition to this anomaly, we note that the topology of the trees varies radically from instantiation to instantiation. In the published article there are two trees, one for the Oceanic languages and one for the non-Austronesian languages. When we examine the tree from the supporting materials that includes all 31 languages, we find that many of the subgroups defined in the published trees vanish.

21. Bilua, the fourth non-Austronesian language of the Solomons, does not attest this feature. This reflects the widespread "oceanicization" that is found throughout the grammar of Bilua (Obata 2003).

In the first two groupings shown in figure 2 we compare the trees presented for some of the non-Austronesian languages of the Bismarck archipelago. In both of them, Kuot is shown as maximally distanced from the other languages (as would be predicted based on geography: Kuot is found in central New Ireland, while all the other languages are found on New Britain). In the published tree, the next split involves Mali, which is sister to a group consisting of Kol, Sulka, Ata, and Anêm. In the tree that appears in Dunn et al.'s supporting data, however, Mali is grouped at a low level with Kol.

When we examine the languages of Bougainville we find an even less convincing picture. In the published subtree from figure 4, we can see that Nasioi's closest relative is Motuna, while in the tree from the supporting materials, Nasioi is the sister of Buin. Yélî-Dnye is sister to a Nasioi-Buin group in the tree from the supporting materials, while in the published tree Yélî-Dnye and Rotokas form a group together. There is no point at which the two trees representing the Bougainville non-Austronesian languages agree. Because these trees were generated, and published, using the same datasets and the same methodology, and yet produce such different topologies, our confidence in this technique cannot be high. Indeed, the low figures for the consistency index (see section 1) support this lack of confidence.[22]

The software package used by Dunn et al. is one of those accepted by genetics researchers in biology (PAUP is used as a teaching tool in the relevant section of the biology department in our own university). Therefore we attribute this instability to the weaknesses of the underlying data, a point that further supports the arguments made in previous sections.

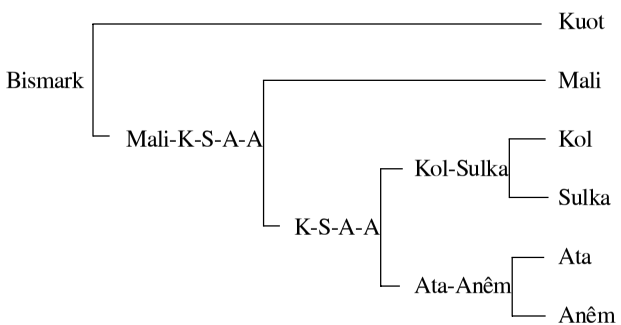
We have suggested in section 4 that data from Austronesian languages outside of Melanesia should have been included in the dataset for this study. Similarly, the authors need to abstract away from the areal trends that pervade Melanesia, either as a whole or in smaller, local areas.[23] To do this, the sample needs to include not just languages of their putative Island Melanesian group, but also non-Austronesian languages of the Melanesian area that are not thought to be related to each other or to the "Island Melanesian" languages. It has not been proved that the "Papuan" languages of Island Melanesia show any particular time depth. The simplest explanation for the lack of cognate lexicon must surely be that the languages are unrelated—a point that Dunn et al. concede and that sets the stage for their study. There is, however, no evidence in this study for an ancient Island Melanesian linguistic entity—simply evidence for the uncontroversial position that there are typological differences between the various non-Austronesian languages and some of the Oceanic languages. There are at least three explanations for the similarities shared by the modern non-Austronesian linguistic lineages: first, they might represent recent inno-

22. It is interesting that in the tree presented in their supporting materials, Motuna is grouped at the maximum distance from any of the other Bougainville languages. Our attempts to replicate the results, using the same data and coding that is reported in Dunn et al., running on the same software, produced trees in which Motuna was consistently grouped at a maximum distance to *all* other languages in the sample, Oceanic and non-Austronesian alike. This results in a de facto grouping of all the Oceanic languages and all the non-Austronesian languages other than Motuna. The issue of the replicability of Dunn et al.'s results is not one we wish to pursue in this qualitative assessment of the work, but see Wichman, Musgrave, and Donohue, in preparation.

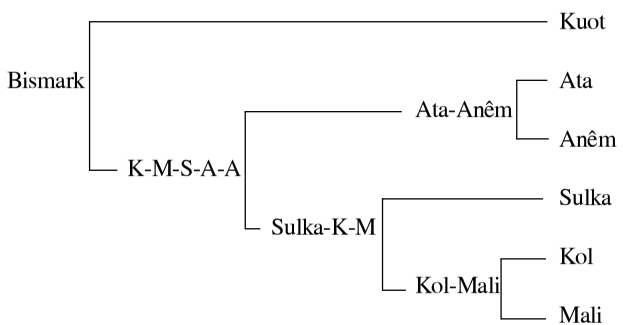
23. The existence of a Melanesian linguistic area is well known, as attested by the many Austronesian languages in the area that show grammatical features that are not typical of Austronesian languages generally, but that are well attested in the non-Austronesian languages of Melanesia.

FIGURE 2. FOUR SUBGROUPING SELECTIONS FROM DUNN ET AL. (2005), COMPARING THE PUBLISHED AND SUPPORTING DATA

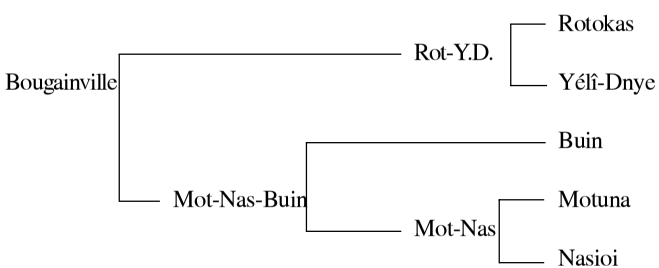
(a) Bismarck languages, from *Science*, figure 4



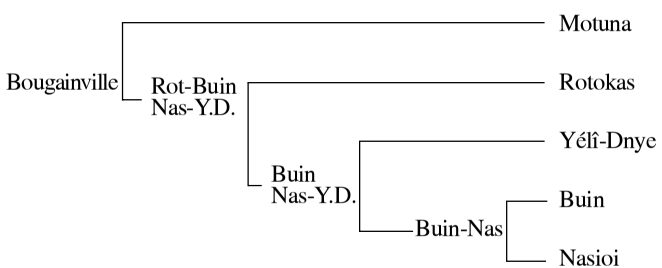
(b) Bismarck languages, from Supporting Material, figure S1



(c) Bougainville languages, from *Science*, figure 4



(d) Bougainville languages, from Supporting Material, figure S1



vations (effectively, creolizations) that arose during the period of initial contact between the Oceanic population and their new non-Austronesian neighbors. We know that there was a large trading cycle in Island Melanesia (see, e.g., Pawley 2006, among others), and the appearance of a number of effectively new lineages, showing the continuation of pre-existing typologies in the area, as well as the admixture of some Oceanic elements. We know (e.g., Thurston 1982) that modern contact between coastal Oceanic and inland non-Austronesian languages results in mutual influence, and there is no reason to suppose that a different sociolinguistic setting held in the previously vastly more non-Austronesian dominated islands (see Blust 2005 for discussion of a likely scenario that would have prevailed to the east, and probably west, of New Guinea at the time of initial Austronesian contact). A somewhat simpler solution is to suppose, as does Ross (2001; see table 1, in section 2), that there are several unrelated lineages present in Island Melanesia. A third scenario would see some (or none) of the non-Austronesian languages in Island Melanesia as being truly ancient in the area, and some as being more recent arrivals from further west in mainland New Guinea. The lack of any established links between the non-Austronesian languages of Island Melanesia with the non-Austronesian languages of mainland New Guinea makes this a difficult hypothesis to defend, but it cannot be ruled out.[24] A combination of any of the above three scenarios could also hold in the region: some recent koineizations, some ancient lineages, and some recent immigrations. Importantly, the similarities in typology observed in the area can just as easily be explained by areal as by genealogical scenarios.

To address this issue, we propose the inclusion of a number of diverse non-Austronesian languages from Melanesia beyond those in the study. Note that we do not make this suggestion with a view to testing the latter part of the claims outlined in section 2; the wider perspective is needed in our view to control for the effect of areal trends. In order to establish relationships between the non-Austronesian languages of Island Melanesia and non-Austronesian languages elsewhere in New Guinea, it is necessary to include data from further afield to assess areal factors in the larger region. The fact that the "control" group in the search for Island Melanesian unity is the Austronesian family is fortuitous, because the Austronesians have such a wide spread that we can certainly abstract away from any local areal factors by examining languages beyond that area. Further, we suggested in 3.2 that data from unrelated languages should have also been included as a negative control. At this point, we have no idea what an unrelated language would look like when assessed against this set of features, or how a truly and unambiguously unrelated language would appear in an unrooted tree. Given that their methodology involves the use of these unrooted trees that—as stated earlier—cannot assign noncognacy to any two pairs of languages, this is a particularly great failing.

Finally, we wish to address the question: why did we go to this trouble in pointing out deficiencies in Dunn et al.? The reason is that their work has been widely popularized, and this raises issues both for linguists and for scholars in other disciplines. First, within the linguistic community the use of new methodologies such as employing unrooted

24. We note that most of the typological features the distribution of which is found to be significantly different between the Oceanic and non-Austronesian languages in Dunn et al.'s sample are also present in north-central New Guinea, and that it is along the north coast of New Guinea that the immigrant Oceanic population traveled, both west to east and later east to west.

trees is for most people unassessable, and we believe that it is therefore an important contribution to the general debate to examine such work in detail. And, outside the mainstream linguistics community, we believe that assessment of this work may not be based on a good understanding of the relevant linguistic issues.[25] Thus we find references such as Gray (2005), who states that "the Dunn et al. approach is an important step forward in this interdisciplinary endeavour" (presumably the unraveling of the prehistory of the Pacific). Gray also gives the following description of Dunn et al.'s results: "The Papuan tree of Dunn et al. shows some geographic clustering at its tips. The signal toward the base of the tree is very weak, suggesting that few structural features support these historical links. However, the signal that is present is consistent with a scenario involving a time depth greater than 10,000 years. Dunn et al. are careful to emphasize that the signal is weak and discuss alternative hypotheses. … [Dunn et al.'s paper] does not conclusively demonstrate deep historical signals in structural features" (2008).

To state that the results are "consistent" with an ancient time depth is not to claim that they support such a time depth; they are also consistent with a time depth of 1,000 years in an area with strong diffusion of linguistic features, or a time depth considerably greater than 10,000 years and multiple groups of converging languages. Despite the reservations expressed in this passage, Gray elsewhere downplays possible alternative interpretations of the results. Our view is that, at this stage in the endeavor, caution is the appropriate response to work along these lines, and we hope that our discussion has clarified the reasons for taking this position.

7. CONCLUSION. Our discussion in this paper makes it clear that Dunn et al.'s study has not convincingly shown that the non-Austronesian languages of Island Melanesia are related to each other as a group. Both their methodology and their interpretation of their results are seriously flawed.

We acknowledge that, while we have criticized the assumptions about the distribution of features in the languages in the study, and provided spot exceptions from elsewhere, we have not examined these exceptions using quantitative tools. As stated in the introduction, the aim of this paper is to provide a qualititative critique of Dunn et al. (2005); we have argued that, without taking a wider perspective and controlling for the presence of linguistic areas in which features diffuse, the application of their quantitative methods is premature. Adapting the methodology to take account of the issues that we have raised will require a widening of the scope of any future inquiry along these lines; but, as we have shown, in order to claim that grammatical features can persist for 10,000 years, it is necessary to take into account the Austronesian languages that do not belong to the Oceanic branch, as we have done in section 5, and so deal with an Austronesian standard of comparison that shows a time depth of more than the approximately 3,500 years that is accepted as the maximum time depth of Proto-Oceanic, the ancestor of the Oceanic languages that Dunn et al. examined.

The necessary expansion of the inquiry also means including material from languages that are in a known contact area, eastern Indonesia, and therefore subject to diffusion (e.g., Donohue 2005, Donohue to appear, and references cited there), but that is the central point

25. For comments on a related foray into historical linguistics, see Trask (2003).

of this article: the entire New Guinea area shows evidence of wide-ranging linguistic sharing, as well as small, localized linguistic areas (this point is widely discussed in Foley 1986, 1998, 2000). Our suggestion that any typological similarities between the non-Austronesian languages of Island Melanesia (in the absence of any lexical similarities and—even more importantly—in the absence of any evidence of regular sound correspondences) are due to areal rather than genetic factors remains the simplest explanation for the patterns observed.

REFERENCES

Baird, Louise. 2002. A grammar of Kéo, an Austronesian language of East Nusantara. PhD diss., Australian National University.

Beaumont, C. H. 1979. *The Tigak language of New Ireland*. Series B-58. Canberra: Pacific Linguistics.

Blust, Robert. 1993. Central and central-eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–93.

———. 1999. Why lexicostatistics doesn't work: The "universal constant" hypothesis and the Austronesian languages. In *Time depth in historical linguistics*, ed. by Colin Renfrew, April McMahon, and Larry Trask, 311–32. Cambridge: MacDonald Institute for Archaeological Research.

———. 2004. Austronesian nasal substitution: A survey. *Oceanic Linguistics* 43:73–148.

———. 2005. Review of *The Oceanic languages*, by John Lynch, Malcolm Ross, and Terry Crowley. *Oceanic Linguistics* 44:544–58.

Bradshaw, Joel. 1978. The development of an extra series of obstruents in Numbami. *Oceanic Linguistics* 17:39–76.

———. 1982. Genitives and relatives in Numbami, a New Guinea Austronesian language. In *Gava': Studies in Austronesian languages and cultures dedicated to Hans Kähler* (*Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg,* vol. 17), ed. by Rainer Carle, Martina Heinschke, Peter W. Pink, Christel Rost, and Karen Stadtlander, 123–39. Berlin: Dietrich Reimer.

———. 1993. Subject relationships within serial verb constructions in Numbami and Jabêm. *Oceanic Linguistics* 32:133–61.

———. 1997. The population kaleidoscope: Another factor in the Melanesian diversity v. Polynesian homogeneity debate. *Journal of the Polynesian Society* 106:222–49.

———. 1999. Null subjects, switch-reference, and serialization in Jabêm and Numbami. *Oceanic Linguistics* 38:270–96.

———. 2006. Grammatically marked ideophones in Numbami and Jabêm. *Oceanic Linguistics* 45:53–63.

Brandes, Jan Laurens Andries. 1884. *Bijdragen tot de Vergelijkende Klankleer der Westersche afdeeling van de Maleis-Polynesische taalfamilie.* Utrecht: P. W. van de Weijer.

Bugenhagen, Robert D. 1995. *A grammar of Mangap-Mbula, an Austronesian language of Papua New Guinea*. Series C-101. Canberra: Pacific Linguistics.

Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. In *Linguistic areas: Convergence in historical and typological perspective*, ed. by Yaron Matras, April McMahon, and Nigel Vincent, 1–31. Houndsmill: Palgrave.

Chan, Phyllis, A. Dunn, and Inka Pekkane. Sabah Museum Series C3. Kota Kinabalu: Sabah Museum and state archives.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, ed. by R. Jacobs and P. Rosenbaum, 184–221. Waltham, MA: Blaisdell Publishing.

Chowning, Anne. 1985. Rapid lexical change and aberrant Melanesian languages: Sengseng and its neighbours. In *Austronesian linguistics at the 15th Pacific Science Congress*, ed. by Andrew Pawley and Lois Carrington, 169–98. Series C-88. Canberra: Pacific Linguistics.

Corston, Simon H. 1996. *Ergativity in Roviana*. Series B-113. Canberra: Pacific Linguistics.

Coward, David F. 1990. An introduction to the grammar of Selaru. MA thesis, University of Texas at Arlington.

Davis, Karen. 2003. *A grammar of the Hoava language, Western Solomons*. No. 535. Canberra: Pacific Linguistics.

Decker, Margaret. 1981. Notes on Nagovis grammar. Ukarumpa, Papua New Guinea: Summer Institute of Lingusitics.

Dempwolff, Otto. 1939. *Grammatik der Jabêm-Sprache auf Neuguinea*. Hamburg: Friederichsen, de Gruyter.

Dixon, R. M. W. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.

Donohue, Mark, and John Charles Smith. 1998. What's happened to us? Some developments in the Malay pronoun system. *Oceanic Linguistics* 37:65–84.

———. 1999a. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.

———. 1999b. Syntactic categories in Tukang Besi. *Revue Québécoise de Linguistique* 27:71–90.

———. 2000. Pronouns and gender: Exploring nominal classification systems in northern New Guinea. *Oceanic Linguistics* 39:1–11.

———. 2003. Agreement in the Skou language: A historical account. *Oceanic Linguistics* 42:479–98.

———. 2004a. Typology and linguistic areas. *Oceanic Linguistics* 43:221–39.

———. 2004b. Floating quantifiers and universal grammar. In *Proceedings of the 2003 Conference of the Australian Linguistics Society*, ed. by Christo Moskowsky. Newcastle: University of Newcastle.

———. 2005. Review of *The languages of the Eastern Bird's Head*, by Ger Reesink. *Oceanic Linguistics* 44:287–300.

———. Forthcoming. Covert word classes: Seeking your own syntax in Tukang Besi. *Studies in Language*.

———. To appear. Word order in Austronesian: From north to south and west to east. *Linguistic Typology*.

Dryer, Matthew S. 2005. Map 26: Prefixing vs. suffixing in inflectional morphology. In *The World Atlas of Language Structures*, ed. by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. Oxford: Oxford University Press.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, and Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309(5743): 2072–75.

Dutton, Tom E. 1996. *Koiari*. Languages of the World/Materials 10. Munich: Lincom Europa.

Embleton, S. M. 1986. *Statistics in historical linguistics*: Quantitative Linguistics vol. 30. Bochum: Studienverlag Brockmeyer.

Emeneau, Murray B. 1980. *Language and linguistic areas: Essays*. Stanford: Stanford University Press.

Evans, Nicholas, ed. 2003. *The non-Pama-Nyungan languages of Northern Australia*. No. 552. Canberra: Pacific Linguistics.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Cultural Anthropology* 46:621–46.

Ezard, Bryan. 1997. *A grammar of Tawala, an Austronesian language of the Milne Bay area, Papua New Guinea*. Series C-137. Canberra: Pacific Linguistics.
Florey, Margaret. 2001. Verb and valence in Alune. Explorations in valency in Austronesian languages. *La Trobe Papers in Linguistics* 11:73–120.
———. 2005. Overview: The languages of Central Maluku. *Monash University Linguistics Papers* 4:21–32.
Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
———. 1998. Toward understanding Papuan languages. In *Perspectives on the Bird's Head of Irian Jaya, Indonesia: Proceedings of the conference, Leiden, 13–17 October 1997*, ed. by Jelle Miedema, Cecilia Odé, and Rien A. C. Dam, 503–18. Amsterdam: Rodopi B.V.
———. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29:357–404.
———. To appear. The place of Philippine languages in a typology of voice systems. In *Voice and grammatical relations in Austronesian languages*, ed. by Simon Musgrave and Peter K. Austin. Stanford: CSLI Publications.
Givón, Talmy. 2001. *Syntax: An introduction*. Amsterdam: John Benjamins.
Gray, Russell. 2005. Pushing the time barrier in the quest for language roots. *Science* 309(5743): 2007–08.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of grammar*, ed. by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
Haiman, John. 1980. *Hua, a Papuan language of the eastern highlands of New Guinea*. Amsterdam: John Benjamins.
Harvey, Mark, Ian Green, and Rachel Nordlinger. 2006. From prefixes to suffixes: Typological change in Northern Australia. *Diachronica* 23:289–311.
Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, eds. 2005. *World atlas of language structures*. Oxford: Oxford University Press.
Himmelmann, Nicklaus. 1996. Person marking and grammatical relations in Sulawesi. In *Papers in Austronesian linguistics No.3*, ed. by Hein Steinhauer, 115–36. Series A-84. Canberra: Pacific Linguistics.
———. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In *The Austronesian languages of Asia and Madagascar*, ed. by K. Alexander Adelaar and Nikolaus P. Himmelmann, 110–81. London: Routledge.
Hinton, Bryan Douglas. 1991. Aspects of Tugun phonology and syntax. MA thesis, University of Texas at Arlington.
Holzknecht, Susanne. 1986. A morphology and grammar of Adzera (Amari dialect), Morobe province, Papua New Guinea. In *Papers in New Guinea linguistics No. 24*, 77–166. Series A-70. Canberra: Pacific Linguistics.
Jakobson, Roman. 1962. Über die phonologischen Sprachbünde. In *Selected writings*, ed. Roman Jakobson, 137–43. The Hague: Mouton, 1931.
Johnston, R. L. 1980. *Nakanai of New Britain: The grammar of an Oceanic language.* Series B-70. Canberra: Pacific Linguistics.
Jones, A. A. 1998. *Towards a lexicogrammar of Mekeo (an Austronesian language of West Central Papua)*. Series C-138. Canberra: Pacific Linguistics.
Klamer, Marian. 2002. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41:363–83.
Kolia, J. A. 1978. A Balawaia grammar sketch and vocabulary. In *Studies in language of Central and South-East Papua*, ed. by T. E. Dutton, 107–226. Series C-29. Canberra: Pacific Linguistics.
Laycock, D. C. 1975. The Torricelli phylum. In *New Guinea area languages and language study,* vol 1, *Papuan languages and the New Guinea linguistic scene*, ed. by S. A. Wurm, 767–80. Series C-38. Canberra: Pacific Linguistics.

Laycock, D. C., and Z'graggen, J. 1975. The Sepik-Ramu phylum. In *New Guinea area languages and language study Vol. 1: Papuan languages and the New Guinea linguistic scene*, ed. S. A. Wurm, 729–63. Series C-38. Canberra: Pacific Linguistics.

Li, Charles N., and Rainier Lang. 1979. The syntactic irrelevance of an ergative case in Enga and other Papuan languages. In *Ergativity*, ed. by Frans Plank, 307–24. London: Academic Press.

Lichtenberk, Frantisek. 1983. *A grammar of Manam*. Honolulu: University of Hawai'i Press.

Lincoln, Peter C. 1976. Describing Banoni, an Austronesian language of southeast Bougainville. PhD diss., University of Hawai'i.

Lister-Turner, R., and J. B. Clark. 1931. *Revised Motu grammar and vocabulary.* (rev. by P. Chatterton). Sydney: Government Printer.

Lynch, John. 1978. *A grammar of Lenakel*. Series B-55. Canberra: Pacific Linguistics.

Lynch, John, Malcolm Ross, and Terry Crowley. 2002. *The Oceanic languages*. Richmond, UK: Curzon.

Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: University of Chicago Press.

McMahon, April, and Robert McMahon. 2003. Finding families: Quantitative methods in language classsification. *Transactions of the Philological Society* 101:7–55.

Mosel, Ulrike. 1994. *Saliba*. Munich: Lincom Europa.

Müller, Adam. 1954. *Grammar and vocabulary of the Konua language.* Microbibliotheca Anthropos (MBA) 12. Fosieux, Switzerland: Anthropos Institut.

Newman, John. 1984. Nasal replacement in western Austronesian: An overview. *Philippine Journal of Linguistics* 15:1–17.

———. 1996. *Give: A cognitive linguistic study.* Berlin: Mouton de Gruyter.

Nichols, Johanna. 1992. *Linguistic diversity in space and time.* Chicago: University of Chicago Press.

Nichols, Johanna, and Balthasar Bickel. 2005. Map 59: Possessive classification. In *The world atlas of language structures*, ed. by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. Oxford: Oxford University Press.

Obata, Kazuko. 2003. *A grammar of Bilua, a Papuan language of the Solomon Islands*. No. 540. Canberra: Pacific Linguistics.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. In *The prosody-morphology interface*, ed. by René Kager, Harry van der Hulst, and Wim Zonneveld, 310–43. Cambridge: Cambridge University Press.

———. 2001. Austronesian nasal substitution revisited. In *Segmental phonology in optimality theory: Constraints and representations*, ed. by Linda Lombardi, 159–82. Cambridge: Cambridge University Press.

Pawley, Andrew. 2006. Explaining the aberrant Austronesian languages of Southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 116:213–256.

———. 2007. Recent research on the historical relationships of the Papuan languages, or, What does linguistics say about the prehistory of Melanesia? In *Population genetics, linguistics, and culture history in the Southwest Pacific: A synthesis*, ed. by Jonathan Friedlaender, 96–136. New York: Oxford University Press.

———. Forthcoming. Where and when was proto Oceanic spoken? Linguistic and archaeological evidence. In *Language and text in the Austronesian world: Studies in honour of Ülo Sirk*, ed. by Yury Lander. Munich: Lincom Europa.

Ray, Sidney H. 1926. *A comparative study of the Melanesian island languages*. Cambridge: Cambridge University Press.

Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100:59–129.

Ross, Malcolm. 1984. Maisin: A preliminary sketch. In *Papers in New Guinea linguistics No. 23*, 1–82. Series A-69. Canberra: Pacific Linguistics.

———. 2001. Is there an East Papuan phylum? Evidence from pronouns. In *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton*, ed. by Andrew Pawley, Malcolm Ross, and Darrel Tryon, 301–21. Canberra: Pacific Linguistics.

———. 2003. Typology and language families: A comment on Klamer's "Typical features of Austronesian languages of Central/Eastern Indonesia." *Oceanic Linguistics* 42: 506–10.

Sandfeld, Kristian. 1930. *Linguistique balkanique: Problèmes et resultats*. Paris: Edouard Champion.

Siegel, Jeff. 1984. Introduction to the Labu language. In *Papers in New Guinea linguistics No. 23*, 83–159. Series A-69. Canberra: Pacific Linguistics.

Starosta, Stanley, Andrew Pawley, and Laurence Reid. 1982. The evolution of focus in Austronesian. In *Papers from the Third International Conference on Austronesian linguistics,* vol. 2, *Tracking the travellers*, ed. by Amran Halim, Lois Carrington, and Stephen Wurm, 145–70. Series C-75. Canbera: Pacific Linguistics.

Steven, Lee Anthony. 1991. The phonology of Roma, an Austronesian language of eastern Indonesia. MA thesis, University of Texas at Arlington.

Tauberschmidt, Gerhard. 1999. *Sinaugoro grammar, Papua New Guinea*. Series C-143. Canberra: Pacific Linguistics.

Taylor, Andrew. 1970. Syntax and phonology of Motu: A transformational approach. PhD diss., Australian National University Department of Linguistics.

Terrill, Angela. 2002. Systems of nominal classification in East Papuan languages. *Oceanic Linguistics* 41:63–88.

Thomason, Sarah Grey, and Terrence S. Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Thurston, William R. 1982. *A comparative study in Anêm and Lusi*. Series B-83. Canberra: Pacific Linguistics.

Trask, Larry. 2003. Re: 14.1825, Media: NYT: Celtic found to have ancient roots. Linguist List 14:1876. View at http://linguistlist.org/issues/14/14-1876.html (accessed Feb. 16, 2007).

van Hoëvell, G. W. W. C. 1877. Iets over de vijf voornaamste dialecten der Ambonsche landtaal (bahasa tanah). *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch Indië* 24:1–136.

Waterhouse, J. H. L. 1928. *A Roviana and English dictionary, with English-Roviana index and a list of natural history objects*. Guadalcanal: Melanesian Mission Press.

Weinreich, Uriel. 1963. *Languages in contact*. New York: Linguistic Circle of New York, 1953. Reprint, The Hague: Mouton.

Wichman, Søren, Simon Musgrave, and Mark Donohue. In prep. The East Papuan languages: A case study in the reproducibility of results in typology.

Wichman, Søren, and Arpiar Saunders. To appear. How to use typological databases in historical linguistic research. In *Proceedings of the 17th International Conference on Historical Linguistics,* Madison, WI, July 31–August 5, 2005.

Wivell, Richard. 1981. Kairiru grammar. MA thesis, University of Auckland.

Wurm, S. A., D. C. Laycock, and C. L. Voorhoeve. 1975. General Papuan characteristics. In *New Guinea area languages and language study,* vol. 1, *Papuan languages and the New Guinea linguistic scene*, ed. by S. A. Wurm, 171–89. Series C-38. Canberra: Pacific Linguistics.

Zahn, Heinrich. 1940. *Lehrbuch der Jabêmsprache (Deutsch-Neuguinea)*: Beihefte zur Zeitschrift für Eingeboren-Sprachen 21.

———. 1982. *Jabêm-English dictionary.* Rev. and trans. from the German edition of 1917 by Johannes F. Streicher. Series C-68. Canberra: Pacific Lingusitics.

mark@donohue.cc simon.musgrave@arts.monash.edu.au