Journal of the Linguistic Society of Papua New Guinea

ISSN: 0023-1959

Vol. 31 No. 1, 2013

Towards a Papuan history of languages

MARK DONOHUE
Department of Linguistics,
College of Asia and the Pacific,
Australian National University
mark@donohue.cc

1. Introduction and overview

In this paper, I raise one simple point that must be taken into account when considering the history of the 'Papuan' languages – namely, the scope of the term 'Papuan'. I shall argue that 'Papuan' is a term that logically should include many languages that have generally been discussed as being 'Austronesian'. While much detailed work has been carried out on a number of 'Papuan' language families, the fact that they are separate families, and are not believed to be related to each other (in the sense of the comparative method) any more than they are to the Austronesian languages which largely surround their region, means that they cannot be considered without reference to those Austronesian languages.

I will argue that many of the Austronesian languages which surround the Papuan region (see the appendix) can only be considered to be 'Austronesian' in a lexical sense. Since historical linguistics puts little value on simple lexical correspondences in the absence of regular sound correspondences, and regularity of sound correspondence is lacking in the Austronesian languages close to New Guinea, we cannot consider these languages to be 'fully' Austronesian. We must therefore consider a Papuan history that is much more widespread than usually conceived.

In section 2, I shall discuss some logical problems with the term 'Papuan' as it is commonly used. In section 3, I introduce some details about what we can infer of social history in the 'Papuan' region, and the implications they have had for our understanding of linguistic processes in the area. In section 4, we discuss the notion of regularity in sound correspondences, and examine the implications for our understanding of the 'Austronesian dispersal' that arise from examining this metric for the languages west of New Guinea. Section 5 concludes the discussion.

2. 'Papuan' as a term

The term 'Papuan' has been used essentially to express the conjunction of two concepts:

1. on or near New Guinea
2. not Austronesian (or 'Australian').

These senses come out clearly in some descriptions by leading scholars, both from the 'Papuan' and the Austronesian camps:

> 'the Papuan languages occupy those areas of New Guinea and adjacent islands not claimed by Austronesian languages.' (Foley 1986:1; *The Papuan languages of New Guinea*)

> 'the term "Papuan" is a convenient term for the non-Austronesian languages of Papua New Guinea and eastern Indonesia, not all of which are demonstrably related.' (Tryon 1995:3; *Comparative Austronesian Dictionary*)

The definition raises a question as to the exclusion of the Austronesian languages: why? Given that the 'Papuan' languages are not assumed to be a genealogical grouping, what excludes the Austronesian languages of 'New Guinea and adjacent islands' from being included? The answer to this question is not simple: why are 'Austronesian' languages not 'Papuan'? Informally, linguists appeal to the non-Melanesian origin of the Austronesian languages, invoking a non-Melanesian and, thus, non-Papuan pedigree.

This analogy is demonstrably invalid. If we were, for instance, to talk of the languages of Southeast Asia, we would have to include languages belonging to four families: Austro-Asiatic, Austronesian, Kradai and Tibeto-Burman. Since the linguistic origins of Thai lie north of the region, in southern China, should Thai be excluded from the areal study of Southeast Asia? Or, to invoke another analogy, should a study of the languages of Eastern Europe exclude Romanian, since it is known to have originated outside the region?

We would not wish to be guilty of 19th century attempts to classify languages on the basis of human phenotypes or (in the modern scientific environment) genetic types, and so would not exclude Austronesian languages on that basis. Similarly, there are no consistent cultural distinctions between the Austronesian and non-Austronesian populations of Melanesia that are not better attributed to geography. The typology of the Austronesian and Papuan languages is not neatly split into opposing clusters (Hunley et al. 2008, Donohue et al. 2011), and the lexicon is well-known to be highly diverse in Melanesia, including the Austronesian languages of Melanesia (Dyen 1963, Blust 2000 for the Austronesian languages of the area, and studies such as Donohue and Denham 2009 for the region as a whole). In short, since 'Papuan' is an inherently areal group without genealogical implications, the failure to include the Austronesian languages when attempting to define 'Papuan' is fatal.

I shall not address here the question of how to best study areality, on which much has been written (e.g., Thomason 2001, Muysken 2010, and many others), and shall return briefly to the question of genealogical argumentation, namely the application of the comparative method, in section 4. For now, I hope to have established the need to at least consider the Austronesian languages when discussing 'Papuan', and wish to move on to some flaws often repeated when discussing the arrival of the Austronesian languages in Melanesia and adjacent areas.

3. Towards a study of social histories in the 'Papuan' region

Much thinking, implicit or explicit, assumes that, prior to the Austronesian linguistic colonization of its range, pre-existing human societies were not highly 'developed'. To cite just one recent example,

> '…the Austronesian expansion, which represents a complex demographic process of interaction between migrating Neolithic farmers and indigenous Mesolithic hunter-gatherer communities (Xu et al. 2012)

Let us review some uncontroversial facts that are relevant to a discussion of human settlement in the 'Papuan region' and adjacent areas:

- Human occupation has an antiquity of many tens of thousands of years;

- since New Guinea and the Asian mainland were not, at any time relevant to human settlement, connected, we know that the settlement of New Guinea necessarily involved travel between (intervisible) islands, over the sea;
- There is good evidence that maritime capabilities were *not* lost after settlement:
    - increasingly remote island chains, such as the Admiralties and Bougainville, continued to be settled intermittently (e.g., Kirch 2000, Irwin 1992, Spriggs 1997);
    - plants propagated west from New Guinea over long time periods prior to 4000 years BP (before present) (e.g., Denham and Donohue 2009, Donohue and Denham 2009, 2010, Perrier et al. 2011);
    - domesticated animals propagated east from mainland Southeast Asia prior to 4000 years BP (e.g., Dobney et al. 2008, Gongora et al. 2008);

All of this suggests that, rather than 'indigenous Mesolithic hunter-gatherer communities', the world of Island Southeast Asia and Melanesia was occupied by a mosaic of societies that engaged in plant cropping, animal husbandry, and maritime trade.

Some uncontroversial aspects of the Austronesian component of the history of Island Southeast Asia and Melanesia are summarised below:

- the Austronesian languages developed in and later dispersed from Taiwan (e.g., Blust 2009a, and many others);
- Austronesian languages were first attested outside Taiwan approximately 3800 years BP (this is if we assume that the appearance of a particular design on pots should be equated with the arrival of a particular linguistic tradition) (Spriggs 2011, and many others);
- 'Austronesianisation' proceeded rapidly across Island Southeast Asia, arriving in Melanesia not more than a few centuries later than the first appearance in the northern Philippines (e.g., Bellwood 1985);
- the Austronesian languages of Melanesia (and much of what is now eastern Indonesia and East Timor) are typologically and lexically quite distinct from the Austronesian languages of Taiwan and the Philippines (e.g., Brandes 1884, Donohue 2005, 2007, Donohue and Schapper 2008, and many more);
- the lifestyles of the Austronesian-speaking peoples in Melanesia and surrounds are very different from speakers of Austronesian languages in Taiwan and The Philippines (e.g., Donohue and Denham 2010);
- the initial dispersal of pots (from which we assume we can date the dispersal of Austronesian languages) does *not* correspond to a change in the food crops utilised (Barker et al. 2011, Barton 2012, Donohue and Denham 2010).[1]

All of this paints a slightly confused picture: while 'Austronesianisation' has, to judge from the distribution and numbers of Austronesian languages, been incredibly significant in the Island Southeast Asia/Melanesia region, it did not involve a correspondingly massive social transformation. If there wasn't a large-scale food revolution, a massive introduction of new

[1] We know what the (sudden) appearance of agriculture looks like: The New Guinea highlands are one of the areas of early independent agricultural experimentation in the world. Stone axe/adzes were used to clear forest (from c. 7000 cal BP), resulting in massive and well-documented changes in the landscape (e.g., Denham et al. 2003).

material culture items, or a wholesale population replacement, what happened to the pre-Austronesian cultures of the area?

4. Pre-Austronesian continuity

What was the fate of the societies that flourished in Island Southeast Asia and Melanesia prior to the Austronesian linguistic colonization? To address this question we have to make a digression on the nature of argumentation in historical linguistics.

Evaluation of data in a historical linguistics sense must invoke employing the comparative method, but it can also involve other methods for detecting (aspects of) languages' histories, since a language history is the linguistic reflection of social events over time, and different social events (contact) are differently reflected in different parts of a language: the lexicon, the phonology, the morphology, the syntax, discourse structures such as topic uses and intonation patterns, etc.[2]

The comparative method involves identifying traits in languages that are unlikely to have arisen independently (by chance or by drift), and are unlikely to have been borrowed. In terms of the phonology, this essentially involves searching and testing for regular sound correspondences across languages (in morphological terms, we examine the languages for cognate morphological paradigms or cognate irregularities). In the case of the Austronesian languages of Island Southeast Asia and Melanesia, we cannot compare much in the way of cognate morphological paradigms, since the languages are so typologically different: the Philippine-style voice morphology is absent in most languages, and even the otherwise relatively well-preserved prefixes *ma- 'stative', *pa- 'causative and *təR- 'passive' are absent from the languages immediately west of New Guinea. We are left with regular sound correspondences.

4.1 Regular sound correspondences: quantifying the principle

The essential principle behind examining regular sound correspondences is that, while individual words are easily borrowable, and so can easily show resemblances, a true cognate will show the same regular correspondences of sounds that are present throughout the lexicon, and so the language will be structurally similar in a way that goes beyond the possibility of chance. For instance, examining just the word 'motor' in English and Mandarin we could be mistaken for thinking we had evidence of relatedness: English [mowtə] and Mandarin [mɤtwɔ] show good correspondences (by identity) in both consonants, and plausible correspondences in the vowels. (We should note, in passing, that the relationship between English and Mandarin is not arbitrary, even if it is not one of shared ancestry, since the Mandarin word is a loan from English.) We do not, however, have *any* evidence of regular sound correspondence**s**, since we have only examined one correspondence for each of the phonemes under consideration. A wider range of data is shown in Table 1, and, after examining these data, it is quite clear that the matches found in the first two words should be attributed to borrowing or chance, and not any historical relationship. In the light of this, we should consider most of the forms not to be cognates, leaving

[2] There are additional techniques invoked in the literature, which we might split into two: various lexical methods, starting with lexicostatistics and diverging in many directions; and typological methods, involving the comparison of different typological features, without the constraints of mapping form *and* function that the comparative method requires. These are not addressed here, as they are not immediately relevant to the main argument being made.

us, one could argue, with 100% regular sound correspondences between those forms which are likely cognates – but we are back to the problem of loans and chance resemblances.

Table 1. Putative correspondences between English and Mandarin

		English	Mandarin	m match?	t match?	Same match as:
1	'motor'	mowtə	mɤtwɔ	m	t	2
2	'mother'	mɐm	Mama	m		1
3	'moon'	muːn	jɥɛ	jɥ?		
4	'mountain'	mawntən	ʂan	ʂ?		11?
5	'mouth'	Mawθ	kow	k?		
6	'meat'	miːt	ʐow	ɹ?		
7	'tooth'	tuːθ	jatsz		j?	
8	'tongue'	tɐŋ	ʂɤtow		?	
9	'tail'	Tejl	wej		?	
10	'two'	tuː	əɹ		?	
11	'tree'	tɹiː	ʂu		ʂ?	4?

This can be compared to Table 2, which shows the fate of the same words in a comparison between English and Dutch. The correspondences have been marked as if they were matched automatically, but it is quite clear that the words for 'mountain', 'meat', 'tail' and 'tree' are not cognate. Even ignoring this (important) filter, we could argue, on the basis of the data in Table 2, that the /m/ correspondences were 67% regular, since four of the six putative matches show the same form. The /t/ correspondences are, again not taking into account cognacy or otherwise, also 67% regular. If we considered only those words that were cognate we would find 100% regularity in the words: given the Dutch (or English) form, and the knowledge that the word is cognate in the other language, we can predict the form in the other language with complete confidence. Not only the forms, but, importantly, the *regular* correspondences between the forms show a resemblance that goes beyond chance, and so relatedness can be proven.

Table 2. Putative correspondences between English and Dutch

		English	Dutch	m match?	t match?	Same match as:
1	'motor'	mowtə	moːtər	m	t	2,3,4,5,6
2	'mother'	mɐm	Ma	m		1,3,4,5,6
3	'moon'	muːn	maːn	m		1,2,4,5,6
4	'mountain'	mawntən	bɛrəx	(b?)		1,2,3,5,6
5	'mouth'	Mawθ	mond	m		1,2,3,4,6
6	'meat'	miːt	fleːs	(f?)		1,2,3,4,5
7	'tooth'	tuːθ	tɑnd		t	1,4,6,8,9,10
8	'tongue'	tɐŋ	tɔŋ		t	1,4,6,7,9,10
9	'tail'	Tejl	staːrt		t?	1,4,6,7,8,10
10	'two'	tuː	tʊeː		t	1,4,6,7,8,9
11	'tree'	tɹiː	boːm		b?	

If we were to continue this method, examining different consonants in the basic vocabulary of different languages and the regularity of their correspondence, we arrive at an average figure

of 96% for the degree to which sound correspondences are regular in European languages.[3] Similar figures for similar datasets can be obtained for Finno-Ugric (94%), and for language groups that have relatively uncomplicated migration histories (e.g., Polynesian: 92%).

Examining this sort of data for the Austronesian languages relevant to New Guinea is revealing. Only a part of the data will be presented here, but the overall results are significant.

Table 3 shows the correspondences that hold for an initial Proto-Malayo-Polynesian *b in Indonesian, and the reader can verify that 15/15, 100%, of the reflexes are regular. Clearly, the relationship between Proto-Malayo-Polynesian and Indonesian is more like the relationship between English and Dutch than the relationship between English and Mandarin. If we examine similar data, comparing PMP to Tukang Besi, we find that 80% of the reflexes are regular; again, this is more like the English-Dutch relationship than the borrowing relationship that we see between English and Mandarin. In Ngadha, the figure is 75%, and in Arguni, spoken on the western coast of New Guinea, the majority reflex is found in only four out of the seven identifiable reflexes of PMP *b, 57%. The locations of the languages portrayed in Tables 3 – 6 are shown in Map 1. It is clear that the closer the languages are to New Guinea, the lower the regularity of reflexes of PMP *b (this generalisation proves to be valid when a larger set of languages is examined, as summarised in section 4.2). The goal of this exercise is not to overturn existing comparative method techniques for assessing the affiliation of a language, but to propose a new methodology for examining the kinds of correspondences that are found, and so to assess whether the degree of regularity in correspondences is enough to eliminate borrowing(s), or independent developments, as the source of apparently cognate vocabulary.

Table 3. Putative correspondences between PMP and Indonesian

		PMP	Indonesian
1	‘kill’	***b**unuq	**b**unuh
2	‘split’	***b**əlaq	**b**əlah
3	‘body hair’	***b**ulu	**b**ulu
4	‘flower’	***b**uŋa	**b**uŋa
5	‘stone’	***b**atu	**b**atu
6	‘wet’	*ma-**b**asəq	**b**asah
7	‘buy’	***b**əli	**b**əli
8	‘new’	*ma-**b**aqəRu	**b**aru
9	‘heavy’	*ma-**b**əRəqat	**b**ərat
10	‘open’	***b**uka	**b**uka
11	‘rotten’	*ma-**b**usuk	**b**usuk
12	‘true’	*ma-**b**ənər	**b**ənar
13	‘hide’	***b**uni	səm/**b**uɲi
14	‘below’	***b**a**b**aq	**b**awah
15	‘star’	***b**ituqən	**b**intaŋ

[3] This measures the degree of regularity of *initial plosives* found in an approximation of a Swadesh wordlist. Different classes of sounds are (anecdotally) known to show different degrees of regularity, and different positions in the word are more or less conducive to regularity. Plosives were chosen for this study because they can be expected to be less regular than sonorants (and are thus able to sufficiently differentiate the data), and yet more regular than vowels. The initial position was chosen because that is the part of the word where conditioning environments are minimized.

Table 4. Putative correspondences between PMP and Tukang Besi

		PMP	Tukang Besi
1	'split'	*bəlaq	βola
2	'body hair'	*bulu	βulu
3	'flower'	*buŋa	βuŋa
4	'stone'	*batu	βatu
5	'buy'	*bəli	βoli
6	'new'	*ma-baqəRu	βoʔohu
7	'heavy'	*ma-bəRəqat	moɓoha
8	'hide'	*buni	heɓuni
9	'below'	*babaq	βaβo
10	'star'	*bituqən	βetuʔo

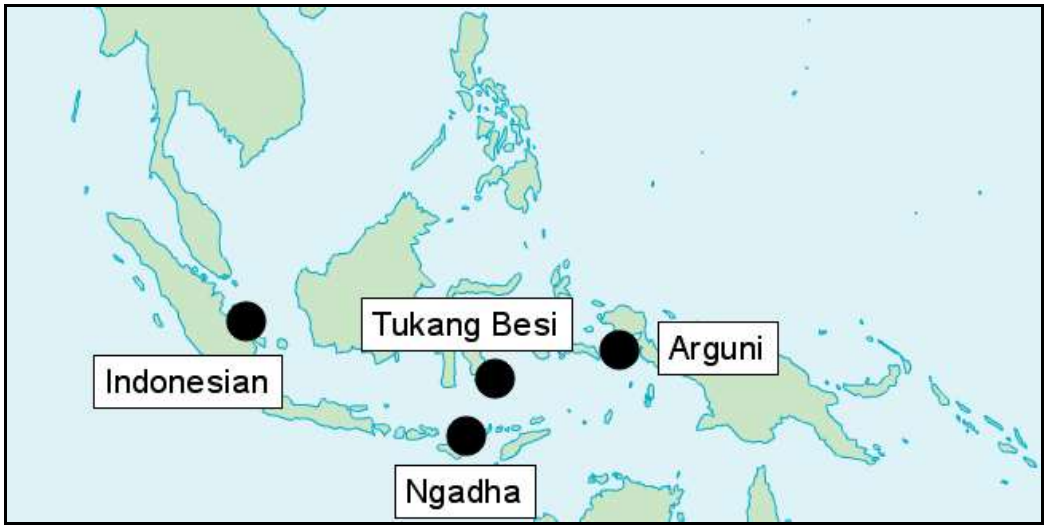
Table 5. Putative correspondences between PMP and Ngadha

		PMP	Ngadha
1	'woman'	*bahi	faʔi
2	'hair'	*buhək	fu
3	'flower'	*buŋa	vuŋa
4	'stone'	*batu	vatu
5	'moon'	*bulan	vula
6	'buy'	*bəli	vəli
7	'fruit'	*buaq	vuʔa
8	'split'	*bəlaq	vəla

Table 6. Putative correspondences between PMP and Arguni

		PMP	Arguni
1	'woman'	*ba-b<in>ahi	popin
2	'moon'	*bulan	purin
3	'body hair'	*bulu	pu\pure
4	'stone'	*batu	puat
5	'pig'	*bəRək	mbo
6	'fruit'	*buaq	mbu
7	'paddle (n.)'	*bəRsay	pʷores

Map 1. Locations of Indonesian, Tukang Besi, Ngadha and Arguni



The question, of course, concerns the 'cut-off' point at which we can legitimately declare a language to show an inheritance relationship, rather than a borrowing relationship. We have seen that when there is no discernible regularity in correspondences, there is no justification in declaring the languages to share a common ancestry. Is 75% a high enough rate for regularity? If we consider what this means, it is not so high a score: 75% *regularity* is a declaration that, of the lexemes that are apparently cognate (thus excluding pairs such as English [mi:t], Dutch [fle:s], fully one-quarter do not show the formal resemblances expected – in other words, they should be considered to have been acquired *not* through inheritance. When we examine a wide range of data of this type, considering correspondences amongst initial plosives in words drawn from Swadesh-like lists of basic vocabulary, we find the figures shown in Table 7 for Indo-European languages of Europe (data drawn from Buck 1949).[4]

[4] Tadmor et al. (2010) present a discussion of the (non-)borrowability of basic vocabulary. Their discussion does not consider complex (imperfect) language shift scenarios such as those envisioned here.

Table 7. Regularity between PIE and European languages

Language	regularity	
Serbo-Croatian	84%	Balto-Slavic
French	85%	Romance
Spanish	89%	Romance
Lithuanian	92%	Balto-Slavic
Polish	92%	Balto-Slavic
Church Slavonic	94%	Balto-Slavic
Latvian	94%	Balto-Slavic
Italian	94%	Romance
Latin	96%	Romance
Rumanian	96%	Romance
Czech	100%	Balto-Slavic
Russian	100%	Balto-Slavic
Danish	100%	Germanic
Dutch	100%	Germanic
English	100%	Germanic
German	100%	Germanic
Gothic	100%	Germanic
Old Norse	100%	Germanic
Swedish	100%	Germanic
Average	**96%**	

We can view this data in terms of how many languages show a regularity metric at what level. So, for instance, in the data in Table 7, we can see that nine languages, almost 50% of the sample, show a 100% rate of regularity of correspondence In Tables 7 and 8, we can, therefore, see the profile of regular sound correspondence: an average level of 96%, and very few languages with less than 90% regular correspondences.

Table 8. Regularity for European languages

Regularity (%age)	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99	100
Languages (no.)	-	-	-	-	-	-	-	-	-	3	7	9
Languages (%age)	0	0	0	0	0	0	0	0	0	16%	26%	47%

Similar results are found with other families from Eurasia; Table 9 shows the values obtained by measuring 18 Uralic languages (again, considering only initial plosives in basic vocabulary items). To demonstrate that this is not an artifact of Eurasian historical linguistics, Table 10 shows the values we find if we examine initial plosives in basic vocabulary items for 34 Polynesian languages. In both cases, the average is high (Uralic: 94%; Polynesian: 92%), and the majority of the languages have regularity scores above 90%. This confirms our ability to identify 'regular' across families.

Table 9. Regularity for Uralic languages

Regularity (%age)	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99	100
Languages (no.)	-	-	-	-	-	-	-	-	1	5	7	5
Languages (%age)	0	0	0	0	0	0	0	0	6%	28%	39%	28%

Table 10. Regularity for Polynesian languages

Regularity (%age)	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99	100
Languages (no.)	-	-	-	-	-	-	-	1	4	5	19	5
Languages (%age)	0	0	0	0	0	0	0	3%	12%	15%	55%	15%

These data should be contrasted with the results found for Turkic, a family that shares the edges of its range with a number of other families, in Eastern Europe and in Western China. Table 11 shows a significant drop in regularity from the values seen in Tables 8 – 10, with outliers scoring very low values. The location of the outliers is not random: the low-regularity languages (Salar: 54%, Yugur: 61%, Tuva: 69%) are all spoken in areas where language shift, from Tibeto-Burman or Mongolic languages, is either a likely scenario or else is the documented reality. The languages, in short, show the morphosyntactic and lexical profile of Turkic languages, but have acquired that lexicon in an irregular fashion and, in some cases, show aberrant phonologies as well. Following Donohue (in press), these languages appear to be examples of language shift.[5]

Table 11. Regularity for Turkic languages

Regularity (%age)	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99	100
Languages (no.)	-	-	-	-	-	-	1	2	4	9	8	-
Languages (%age)	0	0	0	0	0	0	4%	8%	17%	38%	33%	0

We have been able to quantify the notion of 'regular sound correspondence', and to calibrate the results against languages with known histories – languages, among which the notion of 'regular sound correspondences' was born. We have further been able to identify different kinds of language transmission and social histories on the basis of our calibration of the metric of regularity. We shall now apply those conclusions to an examination of Austronesian languages in Indo-Malaysia.

4.2 Regular sound correspondences: Malayo-Polynesian

If the modern Austronesian languages of the Philippines, Malaysia, Indonesia and East Timor have developed regularly, we will expect to see a regularity profile similar to that found for the European sample, or the Uralic or Polynesian samples. If, on the other hand, the transmission and dispersal of Austronesian languages involved language shift, then we should expect to see more languages with low scores for regularity, as was the case for some parts of the Turkic tree.

Since the subgrouping of the Malayo-Polynesian languages is controversial (e.g., Blust 1993, 2009a; Donohue and Grimes 2008, Blust 2009b), with many nodes established through lexical sets rather than by the application of the comparative method, we shall examine a set of languages that does not include Oceanic (since that is a well-established subgroup), and do not include the non-Malayo-Polynesian languages of Taiwan (since they did not participate in the

[5] The distribution of low-regularity languages within Polynesian is similarly non-random: The low-scoring languages (Pileni: 70%; West Uvea: 71%; Kapingamarangi: 75%, Bellona: 78%) are all found outside the 'Polynesian triangle' in Melanesia or (for Kapingamarangi) Micronesia, and are substantially contact-affected by the non-Polynesian languages in their region. Again, the phonologies of most of these languages are aberrant, from a Polynesian perspective.

sound changes that characterise Malayo-Polynesian). 75 languages were examined, selected for geographic dispersal (and availability of data), with the results summarised in Table 12 and Maps 2 – 7 (Language-by-language data is shown in the appendix).

Table 12. Regularity for Austronesian languages

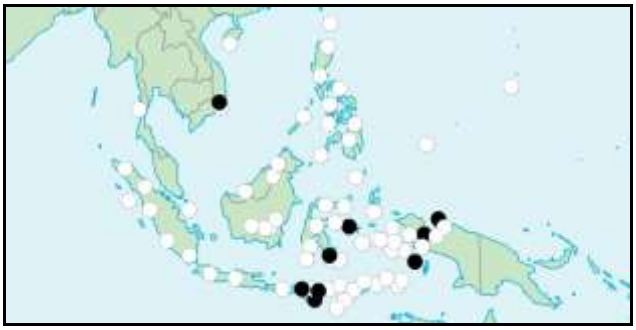
Regularity (%age)	0	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99	100
Languages (no.)	-	-	-	-	-	1	-	8	13	23	18	12
Languages (%age)	0	0	0	0	0	1%	0	11%	17%	30%	24%	16%

The profile in Table 12 does not clearly match any of those we have seen previously. The overall impression is similar to Turkic, but with a large number of more regular languages. The best way to understand the different regularity scores is to examine the geographic distribution of the different languages with different levels of regularity; this is shown in Maps 2 – 7. In Maps 2 – 3 we can see that the distribution of languages with low levels of regularity (shown with black dots) is split between those on the mainland of Southeast Asia, and those closer to New Guinea. The languages concerned are from a range of Malayo-Polynesian subgroups: Malayo-Sumbawan, Celebic, South Halmahera-West New Guinea (within Eastern Malayo-Polynesian) and 'Central Malayo-Polynesian', indicating that the question of high or low levels of regularity cannot be ascribed to membership of particular subgroups that show inherited irregularities (since the larger groups have members with high regularity measures). The distribution in Map 3 is almost perfectly split by the appearance of languages in Map 5, those with perfect regularity scores (like the Germanic languages, seen in Table 7). Maps 5 and 6 show that there is a band of languages with high degrees of regularity, running from the Philippines down to the western Sunda islands of Sumatera and Java, with languages outside this belt being rare and, when present, often showing typological behaviour that is, from the perspective of the conservative languages from the north, highly aberrant (Donohue 2007). Maps 4 and 7 are instructive: we cannot absolutely associate regularity with particular geographic positions, since there are members of the irregular (below 80%) group on all major islands south of the Philippines, and many languages with regularity scores between 80 and 90% in the eastern region in Maluku.

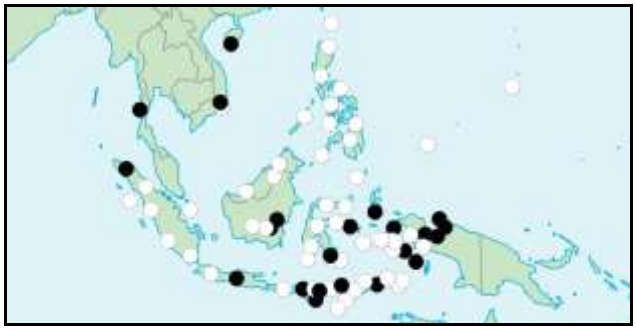
The distribution of regularity in Austronesian

Map 2. Below 50% regular

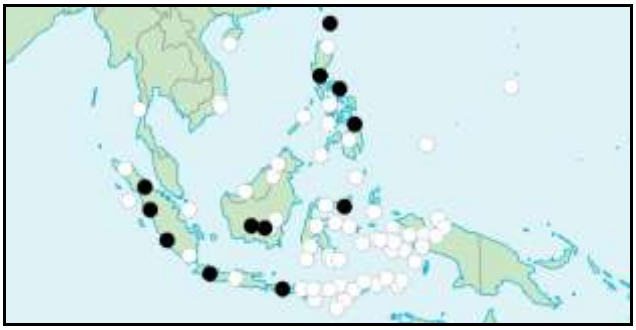
Map 3. Below 70% regular



Map 4. Below 80% regular



Map 5. 100% regular



Map 6. Above 90% regular

Map 7. Above 80% regular

The Austronesian data is best described as bimodal: there are two 'norms', and ignoring one of them does not allow for a full description of the data. On the one hand, there is a set of languages that are best thought of as displaying the hallmarks of regular transmission: they have a large number of reflexes of Proto-Malayo-Polynesian roots, they display those roots with highly regular sound correspondences, and, as 'circumstantial evidence', they have the morphosyntactic profile that would be expected of a language related to Proto-Malayo-Polynesian.

On the other hand, we also find languages that, even when they do show lexical items that appear to reflect Proto-Malayo-Polynesian roots, do not reflect those items with the degree of regularity that has been used to define 'regular sound correspondence'. Further, they tend to show less roots that can be traced to Proto-Malayo-Polynesian, and their typological profile is significantly different – in many cases, not easily distinguished from that of the coastal non-Austronesian languages of New Guinea.

To account for these languages, the 'irregular Austronesian' languages, we could take one of two paths. The conservative view is to consider them to be Austronesian languages that have been contact-affected. Under this view, they are Austronesian 'at the core', but show the effects of considerable influence from non-Austronesian languages near New Guinea. They show the effects of contact in their lexicon, in their morphology, and in their syntactic structures. Further, the lexical items that are reflected in these languages are not reflected in such a way as to lead to the identification of regular sound correspondences with the degree of precision that is expected of regular sound correspondences in European languages and language groups, such as Germanic.

The alternative scenario is to suppose that these languages are essentially non-Austronesian – Papuan, in effect – but that they have been, over the millennia, affected by the increasingly Austronesian-dominated social networks in which they are found. This has not been dramatic:

the languages have acquired lexical items, in many cases, to the extent that they could be considered to have been relexified from a number of Austronesian sources. We must posit more than one lexifier language in order to account for the erratic sound correspondences; this would be expected in any case, if we are talking of an Austronesian *network* of contact, rather than one isolated instance. The contact has not, however, changed the local nature of the phonological systems, or the morphological profile, or the syntactic organisation of the languages, or indeed a large part of even the basic vocabulary of the language. By a consensus of different metrics, the languages remain members of the non-Austronesian linguistic lineage from which they originate, even though in some cases those lineages may no longer be extant, and it might be unlikely that we can determine what that lineage would have been.

Clearly, the most parsimonious account of the provenance of these 'irregular Austronesian' languages in the vicinity of New Guinea is that they are relexified 'Papuan' languages. The existence of most of the technologies, including food technologies, that are present in the region prior to the Austronesian diffusion, the absence of a dramatic change in food crops associated with the spread of Austronesian languages (again, assuming that pots can be equated with languages), the evidence for long-standing trade, or at least dispersal, networks, and the absence of any large-scale replacement of genes (HUGO 2009, Soares et al. 2011, Donohue and Denham 2011), all attest to the cultural momentum of the societies that were present prior to the arrival of the Austronesian lexicon.

Parenthetically, we might note that some of the languages in question might be considered 'creoles' (e.g., McWhorter 2005) – the languages of Flores, for instance, have many of the features 'typical' of creole languages (McWhorter 2005), and languages with this profile appear in an arc that runs from mainland Southeast Asia across Sumatra and along the south of Indonesia and East Timor, terminating in the Bird's Head and Onin peninsulas (in such languages as Onin ('Austronesian') and Abun ('non-Austronesian')). This is not a problem for the account; in fact, creole languages tend to show highly regular correspondences with their principal lexifier language, as shown in Table 13 (selecting the same words that were used in Tables 1 and 2, for maximum comparability). All cognates show completely regular correspondences; the only term in Table 13 that shows any irregularity, 'tree', is not cognate. (A comparison of Papuan Malay and Standard Indonesian would show even closer correspondences.)

Table 13. Putative correspondences between English and Tok Pisin

		English	Tok Pisin	m match?	t match?	Same match as:
1	'motor'	mowtə	mota	m	T	2,3,4,5,6
2	'mother'	mɐm	mama	m		1,3,4,5,6
3	'moon'	mu:n	mun	m		1,2,4,5,6
4	'mountain'	mawntən	mawntɛn	m	T	1,2,3,5,6
5	'mouth'	mawθ	maws	m		1,2,3,4,6
6	'meat'	mi:t	mit	m	T	1,2,3,4,5
7	'tooth'	tu:θ	tit		T	1,4,6,8,9,10
8	'tongue'	tɐŋ	tɐŋ		T	1,4,6,7,9,10
9	'tail'	tejl	tɛl		T	1,4,6,7,8,10
10	'two'	tu:	tu		T	1,4,6,7,8,9
11	'tree'	ʈi:	diwaj		d?	

Given that many of the regular (in terms of sound correspondences) 'Austronesian' languages of the islands west of (and on) New Guinea are the ones with the most aberrant typologies (from the perspective of a conservative Austronesian language of the Philippines or Taiwan), we can best consider these languages not to be conservative exemplars of the Austronesian 'type', but rather good exemplars of a creole 'type'.[6]

5. Conclusions

'Papuan' cannot be studied without its context, and much of the 'context', in many ways, is extremely relevant for our understanding of the histories and classification of Papuan languages. In short, we need to reconsider what it means to be 'non-Papuan'. It is not in dispute that Non-Austronesian languages are dominant on mainland New Guinea, but it is not so clear that the converse, that non-Papuan languages are dominant on the off-shore islands, can be argued for. If we decompose 'languages' into separate parts, we could argue that we have a mixed, but overwhelmingly negative, response: the phonologies (of the 'Austronesian' languages of the Melanesian region) are not in line with Austronesian languages of Taiwan and the Philippines; the morphology and syntax of these languages is not in line with Austronesian languages of Taiwan and the Philippines; and, while the lexicons of many languages are clearly related in many places to the Austronesian linguistic tradition that comes from Taiwan and the Philippines, the putative reflexes in many cases fail to show the degree of regularity that would provide convincing evidence of relatedness in families such as Indo-European.

Many of the Austronesian languages of and near New Guinea have been described as Austronesian, and can be described as Austronesian, but in the same sense that Tok Pisin, the creole that is one of the national languages of Papua New Guinea, is Germanic; or that Michif is a Romance language, and Romanian is Slavic. Romanian is *not* classified as Slavic by the comparative method, but it has clearly been influenced by Slavic languages, notably in the area of phonology (e.g., Donohue et al. 2008). Given the known history of Romanian and of its origins, it is clear that the Slavic-patterning phonology reflects the influence of the original phonologies of the region prior to the shift to the ancestor of what is now Romanian. In a similar vein, when assessing the histories of the Austronesian languages of New Guinea, we can best examine them from a 'Papuan' perspective, and accept them as relexified, but still inherently 'Papuan', languages. Similarly, when assessing the history and classification of the 'Papuan' languages, we cannot neglect the putatively 'Austronesian' languages of the region, which hold as many clues to local past linguistic ecologies as do the modern languages which have not been relexified.

References

Barker, Graeme, Lindsay Lloyd-Smith, Huw Barton, Franca Cole, Chris Hunt, Philip J. Piper, Ryan Rabett, Victor Paz, Katherine Szabo. 2011. Foraging-farming transitions at the Niah Caves, Sarawak, Borneo. *Antiquity* 85: 492-509.

6 Other work on 'aberrant' Austronesian languages (e.g., Chowning 1985, Grace 1992, Pawley 2006) reaches different conclusions, essentially on the basis of the presence of (many) recognisably Austronesian lexical items and, in some cases, the absence of nearby non-Austronesian languages in historical times. This paper, arguing that the lexical items carry less weight than has been thought (due to their lack of regular correspondences and, thus, their lack of immunity from accusations of borrowing), takes a different stance.

Barton, Huw. 2012. The reversed fortunes of sago and rice, *Oryza sativa*, in the rainforests of Sarawak, Borneo. *Quaternary International* 249: 96-104.

Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian archipelago*. London: Academic Press.

Blust, Robert. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32 (2): 241-293.

Blust, Robert. 2000. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. In Colin Renfrew, April McMahon & Robert Lawrence Trask, eds., *Time depth in historical linguistics (Papers in the prehistory of languages 2), 311-331. Cambridge: McDonald Institute for* Archaeological Research.

Blust, Robert. 2009a. *The Austronesian languages*. Canberra: Pacific Linguistics.

Blust, Robert. 2009b. The position of the languages of eastern Indonesia: a reply to Donohue and Grimes. *Oceanic Linguistics* 48 (1): 36-77.

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

Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European languages: a contribution to the history of ideas*. Chicago: University of Chicago Press.

Chowning, Ann. (1985) Rapid lexical change and aberrant Melanesian languages: Sengseng and its neighbours. In Andrew K. Pawley and Lois Carrington, eds., *Austronesian linguistics at the 15th pacific science congress* (Pacific Linguistics: Series C-88), 169-198. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Denham, Tim, and Mark Donohue. 2009. Pre-Austronesian dispersal of banana cultivars west from New Guinea: linguistic relics from eastern Indonesia. *Archaeology in Oceania* 44: 18-28.

Denham, Tim, Simon Haberle, Carol Lentfer, Ruth Fullagar, J. Field, M. Therin, N. Porch and B. Winsborough. 2003. Origins of Agriculture at Kuk Swamp in the highland sof New Guinea. *Science* 301: 189-193.

Dobney, Keith, Thomas Cucchi, and Greger Larson. The pigs of Island Southeast Asia and the Pacific: new evidence for taxonomic status and human-mediated dispersal. *Asian Perspectives* 47 (1): 59-74.

Donohue, Mark. 2005. Word order in New Guinea: dispelling a myth. *Oceanic Linguistics* 44 (2): 527-536.

Donohue, Mark. 2007. Word order in Austronesian: from north to south and west to east. *Linguistic Typology* 11 (2): 351-393.

Donohue, Mark. in press. Who inherits what, when? contact, substrates and superimposition zones. In Balthasar Bickel, David Peterson, Lenore Grenoble and Alan Timberlake, eds., [*Festschrift for Johanna Nichols*]. Amsterdam: John Benjamins.

Donohue, Mark, and Tim Denham. 2009. Banana (Musa spp.) Domestication in the Asia-Pacific Region: Linguistic and archaeobotanical perspectives. *Ethnobotany Research and Applications* 7: 293-332.

Donohue, Mark, and Tim Denham. 2010. Farming and Language in Island Southeast Asia: reframing Austronesian history. *Current Anthropology* 51 (2): 223-256.

Donohue, Mark, and Tim Denham. 2011. Language and genes attest different histories in Island Southeast Asia. *Oceanic Linguistics* 50 (2): 536-542.

Donohue, Mark, and Charles E. Grimes. 2008c. Yet more on the position of the languages of eastern Indonesia and East Timor. *Oceanic Linguistics* 47 (1): 115-159.

Donohue, Mark, Simon Musgrave, Bronwen Whiting and Søren Wichmann. 2011. Typological features analysis models linguistic geography. *Language* 87 (2): 369-383.

Donohue, Mark, and Antoinette Schapper. 2008. Whence the Oceanic indirect possessive construction? *Oceanic Linguistics* 47 (2): 316-327.

Dyen, Isidore. 1963. *The lexicostatistical classification of Malayo-Polynesian languages*. New Haven.

Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Gongora, Jaime, Nicolas J. Rawlence, Victor A. Mobegi, Han Jianlin, Jose A. Alcalde, Jose T. Matus, Olivier Hanotte, Chris Moran, Jeremy J. Austin, Sean Ulm, Atholl J. Anderson, Greger Larson, and Alan Cooper. 2008. Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of the National Academy of Sciences of the USA* 105 (30): 10308-10313.

Grace, George W. 1992. How Do Languages Change? (More on "Aberrant" Languages). *Oceanic Linguistics* 31(1): 115-130.

HUGO Pan-Asia SNP consortium, et al. 2009. Mapping Human Genetic Diversity in Asia. *Science* 326: 1541-1545.

Hunley, Keith, Michael Dunn, Eva Lindström, Ger Reesink, Angela Terrill, Meghan E. Healy, George Koki, Françoise R. Friedlaender, and Jonathan S. Friedlaender. 2008. Genetic and linguistic coevolution in Northern Island Melanesia. *PLoS Genetics* 4.10.e1000239. Online: http://www.plosgenetics.org/article/info:doi/10.1371/ journal.pgen.1000239.

Irwin, Geoff. 1992. *The Prehistoric Exploration and Colonisation of the Pacific*. Cambridge: Cambridge University Press.

Kirch, Patrick V. 2000. *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact*. Berkeley: University of California Press.

McWhorter, John. 2005. *Defining Creole*. Oxford: Oxford University Press.

Muysken, Pieter. 2010. Scenarios for Language Contact. In Raymond Hickey (ed.), The *Handbook of Language Contact*, 265-281. Oxford: Wiley-Blackwell.

Pawley, Andrew K. (2006) Explaining the Aberrant Austronesian Languages of Southeast Melanesia: 150 Years of Debate. *Journal of the Polynesian Society* 116 (3): 215-258.

Perrier, Xavier, Edmond De Langhe, Mark Donohue, Carol Lentfer, Luc Vrydaghs, Frédéric Bakry, Françoise Carreel, Isabelle Hippolyte, Jean-Pierre Horry, Christophe Jenny, Vincent Lebot, Ange-Marie Risterucci, Kodjo Tomekpe, Hugues Doutrelepont, Terry Ball, Jason Manwaring, Pierre de Maret, and Tim Denham. 2011. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences of the United States of America* 108 (28): 11311-11318.

Soares, Pedro, Teresa Rito, Jean Trejaut, Maru Mormina, Catherine Hill, Emma Tinkler-Hundai, Michelle Braid, Douglas J. Clarke, Jun-Hun Loo, Noel Thomson, Tim Denham, Mark Donohue, Vincent Macaulay, Marie Lin, Stephen Oppenheimer and Martin Richards. 2011. Ancient Voyaging and Polynesian Origins. *American Journal of Human Genetics* 88: 239-247.

Spriggs, Matthew. 1997. *The Island Melanesians*. Oxford: Blackwell Publishers.

Spriggs, Matthew. 2011. Archaeology and the Austronesian expansion: where are we now? *Antiquity* 85: 510-528.

Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226-246.

Thomason, Sarah. 2001. *Language contact: an introduction*. Edinburgh: Edinburgh University Press.

Tryon, Darrell. 1995. In Darrell Tryon, ed., *Comparative Austronesian Dictionary*, xxx-xxx. Berlin: Mouton de Gruyter.

Xu, Shuhua, Irina Pugach, Mark Stoneking, Mandred Kayser, Li Jin and The HUGO Pan-Asian SNP Consortium. 2012. Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proceedings of the National Academy of Sciences of the USA* 109 (12): 4574-4579.

Appendix: overall regularity measured for 75 Austronesian languages

Languages, which should be suspected of having been formed via (incomplete) language shift on the basis of their displayed regularity of sound correspondence, and by calibration with other language families, are shown with highlighting. All are found in the eastern half of Indonesia, in the area known to have been home to non-Austronesian languages into historical times, or else are spoken on mainland Southeast Asia. The cut-off point at 70% is essentially arbitrary, but defensible; the next four languages are also in the eastern half of Indonesia, or part of mainland Southeast Asia. See Map 3 for the locations of these languages. Note that the low-regularity languages share their regions with languages that have much higher regularity scores.

Language		Language	
Bima	48%	Buru	90%
Soboyo	62%	Gorontalo	90%
Ngadha	63%	Agta	90%
Chru	67%	Eastern Cham	91%
Kei	68%	Makassar	91%
Wandamen	68%	Tugun	92%
Biak	68%	Batak	93%
Muna	69%	Nias	94%
Kambera	69%	Kowiai	94%
Kedang	72%	Tausug	95%
Leti	73%	Indonesian	96%
Tsat	73%	West Bukidnon Manobo	96%
Matbat	74%	Kelabit	96%
Javanese	74%	Kagayanen	96%
Maanyan	74%	Palauan	96%
Ambai	75%	Chamorro	96%
Watubela	75%	Lampung	97%
Gayo	75%	Sangir	97%
Giman	76%	Timugon Murut	97%
Moken	78%	Aklanon	97%
Mor	80%	Iban	98%
Tunjung	80%	Ivatan	100%
Manggarai	81%	Ngaju Dayak	100%
Banda	81%	Sundanese	100%
Selaru	82%	Toba Batak	100%
Wolio	83%	Bikol	100%
Yamdena	83%	Katingan	100%
Tukang Besi	83%	Mamanwa	100%
Kemak	83%	Bolaang Mongondow	100%
Pamona	83%	Tagalog	100%
Alune	84%	Minangkabau	100%
Mambai	84%	Rejang	100%
Bugis	84%	Sasak	100%
Bobot	85%		
Masiwang	85%		
Sekar	86%		
Paulohi	86%		
Banggai	87%		
Dai	88%		
Sika	88%		
Uab Meto	89%		
Rote	89%		