Becoming Austronesian: mechanisms of language dispersal across southern Island Southeast Asia

Mark Donohue and Tim Denham
The Australian National University
mark.donohue@anu.edu.au, tim.denham@anu.edu.au

We examine the spread of Austronesian languages across Island Southeast Asia as a temporally disparate process that proceeded in different ways at different times, even in the same locale. The background for our examination includes the many ways a language can show Austronesian traits, the known earlier presence of non-Austronesian languages across all of Island Southeast Asia, and the similarity of social processes that can be inferred between mainland and Island Southeast Asia. We suggest that while many languages of Eastern Indonesia are exemplary Austronesian languages, many others should be considered to be the outcome of creolisation processes, other show the traces of scenarios involving language shift from earlier non-Austronesian languages, and yet others should be most parsimoniously considered to be non-Austronesian languages (‘Papuan’) with (in some cases minimal) Austronesian (lexical) veneers.

1. Introduction: the spread of Austronesian languages across Island Southeast Asia

In this paper we examine the spread of Austronesian languages across southern Island Southeast Asia. In contrast to most previous portrayals (eg., Bellwood, Fox and Tryon 1995, Bellwood 2007), we focus on the spread of Austronesian languages through this region as a process distributed in time even more than in space. The process is, in many senses, ongoing and needs to be considered in terms of the broader waxing and waning of interaction between mainland Southeast Asia and New Guinea. We illustrate the diversity of ways a language can ‘be Austronesian’, with particular reference to those languages with a ‘creole-like’ typology. In doing so we suggest that the distribution of different ‘types’ of languages in eastern Indonesia is best explained by assuming a linguistic mosaic such as that described for the languages of mainland Southeast Asia (eg., Enfield 2005), and that a range of social contact scenarios similar to those attested and evidenced in mainland Southeast Asia is also relevant across Island Southeast Asia.

* We acknowledge the Australian Research Council for funding this work under grant DP1093191, we thank the editors of this volume for their valuable comments, and valuable discussion over the years from our colleagues.

1 We use the term ‘Island Southeast Asia’ to refer to the islands found in a triangle with points at Taiwan, Sumatra, and New Guinea. This approximates Solheim’s ‘Nusantao’ region (1984-1985, 2006). While all these terms are ambiguous, we will concentrate on the southern part of the range, approximately the area in which the modern states of Malaysia, Singapore, Indonesia and Timor Leste are found.
The dispersal of Austronesian languages is an historical process, for which the precise antiquity, human migratory implications and cultural associations are not always clear (compare Bellwood 2005; Bulbeck 2008; Donohue and Denham 2010; Denham and Donohue 2012a). The dispersal of Austronesian linguistic forms from Taiwan is generally assumed to have begun around 4000 years ago, based on archaeological evidence across the Batanes Strait dating to this period (eg, Piper et al. 2009). However, the claimed association between material cultural linkages and linguistic dispersal is still fundamentally an assumption, despite recurrent attempts to provide a robust foundation (eg, Pawley and Green 1973; Shutler and Marek 1975; Pawley 2007). Consequently, although the dispersal of Austronesian languages southward across Island Southeast Asia from Taiwan is generally considered to date to c. 4000-3500 years ago, the precise antiquity is uncertain, and dates in regional locales are even less firm.

Additionally, discussions of Austronesian dispersal are repeatedly confused by the conflation, obfuscation and misapplication of categories that are not intrinsically interchangeable (see discussion in Oppenheimer 2004, Terrell 2004, Donohue and Denham 2010; Denham and Donohue 2012a). Although the term ‘Austronesian’ is often considered polyvalent, its application should be specific. ‘Austronesian’ refers unambiguously only to languages and neither to people nor to material culture. Archaeological remains from Taiwan should be considered Taiwanese material cultural traits, as opposed to ‘Austronesian’ in the sense that is relevant to the history of islands to the south of Taiwan, with other regions contributing to the long-term history of Island Southeast Asia. Similarly, the exact provenance of ‘Asian’ genetic characteristics within Island Southeast Asia and beyond are not clear; they likely include a small Taiwanese component, as well as components from other places (Jinam et al. 2012). Additional work (eg., Hill et al. 2007, Soares et al. 2011, Tumonggor et al. 2013) suggests that, regardless of the provenance of the ‘Asian’ genetic characteristics, any Taiwanese component represented is only minor. Indeed, Tumonggor et al. (2013: 170, 172) note that the genetic data is ‘consistent with a rapid expansion from Taiwan to the Philippines and Indonesia, but population dispersals in the opposite direction are equally likely’ and that ‘many aspects of culture—notably the widespread dispersal of Austronesian languages — are not obviously associated with genetics’. Importantly, the dispersal of linguistic, genetic and material cultural traits across Island Southeast Asia should not be assumed to correspond. Languages, genes and cultures may have dispersed separately and together at different times and in different places within the late Holocene (last 4000 years) history of Island Southeast Asia; indeed, various components of each are also likely to have distinct temporal and geographical patterns of dispersal (discussed in Denham and Donohue 2012a).

In this paper we argue that the dispersal of Austronesian languages across Island Southeast Asia should be considered against the broader historical backdrop evidencing widespread and long-term linguistic commonalities and interactions across the southern Southeast Asian region, including the Southeast Asian mainland. Just as on the mainland, extensive language contact and the widespread, stable co-existence of languages of different families must be posited across Island Southeast Asia. The emphasis of our argumentation is upon the temporal processes through which people became Austronesian, at least in terms of the languages they spoke.
We shall highlight a typological characteristic, ‘isolation’, that is found in many languages of the Southeast Asian region, both mainland and Island, but very rare from a global perspective. After a quick summary of some of the socio-historical similarities and differences between mainland and Island Southeast Asia, we shall remark on the difference in distributions of language families in the two regions. We then present compelling reasons to regard the proposed linguistic genealogical homogeneity across much of Island Southeast Asia as an error of analysis. A more nuanced account will reveal the same kind of diversity and the same kind of geographic intermeshing of language family ranges within Island as much as on mainland Southeast Asia. The intention is not to criticise the whole field of Austronesian historical linguistics, but rather to critique the often monodirectional nature of historical linguistic research and the search for language affiliations along single dimensions. Since Austronesian languages demonstrably represent a wide variety of ‘types’, in typological and historical linguistic terms, we believe that it is best to think of the process of ‘becoming Austronesian’ as not being a single, uniform process, but rather different processes involving different starting points, different trajectories, and different time spans in different societies (even those in the same local area). These differential ways of ‘becoming Austronesian’ are not simply asserted: they are measured in multiple dimensions.

2. The unnaturalness of isolation

Classification of languages according to scales of morphological character has been employed for a long time (see discussion in Sapir 1921). Here we shall focus on the nature of ‘isolating’ languages. This section will show that the cross-linguistically rare trait of ‘isolation’ is concentrated in the complex contact area that is mainland Southeast Asia, and is also found in numerous languages of southern Island Southeast Asia.

Isolation is defined as the absence of large amounts of bound morphology, such that the morpheme:word ratio approaches 1:1. Examples (1) – (3) show, with the English translations, four languages expressing one sentence. In the Iha example the two words contain six morphemes; the Tukang Besi case shows three words with seven morphemes. The English translation uses six words with eight morphemes, and the Papuan Malay in (3) has six morphemes in its six words.² The measures of average morphemes/word for the different languages are: Iha, 3; Tukang Besi, 2.3; English, 1.3, and Papuan Malay, 1.0.

Iha

(1) Komoh ni-ndo-nwe-mb-ih.
taro 1P-CAUS-eat-YESTERDAY-3
‘They made us eat taro yesterday.’

Tukang Besi

(2) No-pa-manga=kami te=opa dinggawi.
3R-CAUS-eat=1PA CORE=tuber yesterday
‘They made us eat taro yesterday.’

² Donohue and Sawaki (2007) discuss the possibilities of cliticisation of forms such as dong= in Papuan Malay. The language also allows non-bound forms, as discussed here.
Papuan Malay

(3) **Dong kasi makan kladi kitong kemarin.**

3PL give eat taro 1PL yesterday

‘They made us eat taro yesterday.’

We can further note the nature of the bound morphology. In Iha and Tukang Besi we find agreement for two arguments on the verb; Iha has additional inflectional morphology in the form of the tense marking suffix, and both of these languages have a bound derivational prefix marking causative. Further, the Tukang Besi example shows a core case marker on the nominal object of the clause, *opa*. The English translation shows more than one morpheme per word only with the suppletive forms *make* and *us*, which include tense and case information. In the Papuan Malay sentence there is exactly one morpheme per word. Just these three examples show us that measures of isolation are sensitive to both inflectional and derivational possibilities; to agglutinative and suppletive forms; and, to possible differences between verbal and nominal morphology. We could easily add other widely-attested features to a discussion of the dimensions that make up ‘isolating’: just citing features coded in the WALS database (Haspelmath et al. 2005), the features listed in Table 1 all bear on the measure of degree of morphological isolation.

| Table 1. Morphosyntactic features bearing on the notion of ‘isolating’ (from WALS) |
|-----------------------------------------------|---------------------------------|
| Affixed gender                               | Affixed demonstratives          |
| Affixed tense/aspect/mood                    | Affixed subordinating           |
| Affixed (in)definiteness                     | Core case affixation            |
| Affixed possession                           | Subordinating structures        |
| Adpositional agreement                       | Verbal agreement                |
| Verbal affixation for: negative, optative, epistemic and situational possibility, evidentiality, reciprocals, passive, antipassive, applicative, causative, interrogative, desiderative | Verbal suppletion for: tense, aspect, number, imperative, hortative |

Map 1 shows the distribution of isolating languages, as coded by Bickel and Nichols in *WALS* (see Bickel and Nichols 2011 for a discussion on what this feature ‘means’ linguistically). On this map solid black circles represent languages with only isolating or tonal morphological profiles, and diamonds are those languages coded as having ‘isolating/concatenative’ profiles. From this map, it is clear that the distribution of languages with a predominantly isolating character is neither random, nor widespread. Bickel and Nichols (2011) write that ‘Languages with isolating formatives, or traces of isolating structure in mixed types, are mostly confined to the Sahel Belt of West Africa and to Southeast Asia and the Pacific.’
While useful, Map 1 is based on only 165 languages coded for the feature ‘± isolating’. Due to this lack of resolution, it is not straightforward to determine the location of the ‘isolating’ belts. Of concern to us here, the region of ‘Southeast Asia and the Pacific’ covers a lot of ground. In Map 2 we see the same feature coded for a database of approximately 1500 languages (black: ‘+ isolating’, small grey: ‘– isolating’). From Map 2 it is clear that while Sahel does feature as an area with isolating languages, it is at nothing like the density found in mainland Southeast Asia. And while ‘the Pacific’ hosts a number of additional isolating languages, most of them are found west of New Guinea, or on the (western) New Guinea mainland. It is clear that, while an isolating profile can arise in a language from almost any part of the world, it is rare for a region to be dominated by isolating languages.

Map 2 is an improvement over Map 1 in terms of coverage, but it suffers from the same collapsing of multiple dimensions that could be used to measure isolation. Does a language count as isolating if it displays an isolating setting for only one of the parameters mentioned.
in Table 1? Do the different parameters have independent existences, or is there a morphological conspiracy? Maps 3–8 present the distribution of the features discussed with respect to examples (1)–(3), and the addition of ‘± subordinating morphology’; the frequencies of these different features are shown in Table 2.³ Note that there is a significant degree of correlation between the different values; if the distribution of the five features (rows 2–6 in Table 2) was truly random we would expect only one tenth of the languages to show all the features together.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Global frequency</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 isolating</td>
<td>11.4%</td>
<td>1384</td>
</tr>
<tr>
<td>2 no verbal agreement</td>
<td>30.8%</td>
<td>1245</td>
</tr>
<tr>
<td>3 no tense marking</td>
<td>31.3%</td>
<td>804</td>
</tr>
<tr>
<td>4 no bound causative</td>
<td>26.6%</td>
<td>863</td>
</tr>
<tr>
<td>5 no core case marking</td>
<td>48.8%</td>
<td>1167</td>
</tr>
<tr>
<td>6 no subordinating morphology</td>
<td>30.5%</td>
<td>606</td>
</tr>
<tr>
<td>7 all of features 2–6</td>
<td>4%</td>
<td>476</td>
</tr>
</tbody>
</table>

An examination of these features shows that the Sahel belt is less prominently featured, and that the correlation between the features in Table 2 is geographical. The ‘Southeast Asia and the Pacific’ described in WALS contracts or expands according to the feature examined, and is not evenly distributed across all of that region. Of note is the fact that for some features we do not see the disjunct distribution found in Map 2, with mainland Southeast Asia separated from the (western) New Guinea region. Map 8 shows those languages for which three or more of the features shown in Maps 3–7 are in an isolating setting; from this map we can see that there is a clear connection between mainland Southeast Asia and western New Guinea. For some features this trail has been muddied, probably by the spread of languages typologies from the north, but it remains detectable and shows how this region stands out against the sparse distribution of isolating languages elsewhere in the world.

Maps 3–8. Decomposing ‘isolating’ according to features identified in (1)–(3)

Map 3. No verbal agreement

Map 4. No tense marking

³ The features are similarly drawn from, and defined in, WALS. The interested reader is referred to chapters 66, 67, 98, 99, 102, 111, 125, 126 and 127 (Dahl and Velupillai 2011a, 2011b, Comrie 2011a, 2011b, Siewierska 2011, Song 2011, Cristofaro 2011a, 2011b and 2011c). The feature ‘no verbal agreement’ follows chapter 102; ‘no tense marking’ is based on the negative union of WALS features 66 and 67; ‘no bound causative’ follows chapter 111, and ‘no core case marking’ is coded for languages showing negative values for chapters 98 and 99. The feature ‘no subordinating morphology’ is calculated for languages that show ‘balanced’ values for all of WALS chapters 125-127, and no ‘deranked’ values.
Judging from the information we have seen in Table 2 and Maps 2 – 8, we can state that ‘isolation’, measured as a single parameter or decomposed into separate variables, and is not a normal state of affairs for most languages or language areas around the world. Rather, some degree of inflectional (and derivational) morphology is the norm. In one area, however, stretching from the (eastern) Himalayas to (western) New Guinea (depending on the feature examined), isolating characteristics are modal with a much higher than expected frequency, as can be informally judged from the maps. The same area is remarkable for the widespread adoption of SVO order in Austronesian languages, independent of any recognised subgroup that includes these languages, a region that also excludes Austronesian languages to the north in Taiwan and the Philippines (Donohue 2005, 2007). We shall examine this distribution in terms of the processes that are known to have brought about such widespread language contact in mainland Southeast Asia, compared to our assumptions about language and population dispersals in Island Southeast Asia.

3. Southern Southeast Asia

In this section we will point out some of the similarities, and some of the differences, between mainland and Island Southeast Asia. We shall combine this exposition of differences, often superficial, with the many similarities that we have seen in section 2, and the social histories that are implied by those data.

The area of ‘Southeast Asia’ has been described in terms of two subparts, in many ways unified and yet frequently treated disparately. In the west on the mainland, the area most proto-typically referred to without qualification as ‘Southeast Asia’ (e.g., Enfield 2005) is centred approximately on Thailand and includes surrounding regions. Enfield defines it as ‘the region encompassing Vietnam, Laos, Cambodia, and Thailand, with some extension west into Burma, south into Peninsular Malaysia, and north into southern China’ (2005: 182). This (very practical, and functional) definition explicitly excludes the other subpart, Island Southeast Asia (also known as Maritime Southeast Asia, or Insular Southeast Asia, Indo-
Malaysia or (earlier) ‘the Malay archipelago’). This second region can be characterised as the islands lying between the Malay peninsula and New Guinea; whether a northern extension to include the Philippines is counted or not varies from author to author, and for the purposes of this paper we will take the more restricted definition, that roughly corresponds to the insular part of the modern states of Malaysia, Indonesia and Timor Leste. For clarity, we shall refer to Island Southeast Asia and Mainland Southeast Asia when we want to differentiate the two regions.

There are many similarities between these two regions. We can summarise just a few (eg, Bellwood 2007, Bellwood and Glover 2004, Bellina et al. 2010, Wright et al. 2013):

• Both areas have a long history of modern human occupation, especially around coastal regions.
• Both regions host multiple ecological zones, with high mountainous hinterlands leading to coastal plains that in some cases extend far inland.
• Both regions have seen long-distance trade and interaction spheres, in which goods from one zone are transported to another zone, or out of the region entirely. These kinds of long-distance interactions are repeated throughout the Holocene period and into modern times.
• Both regions have complex proto-historic traditions, including the arrival of Indic states in the last 2000 years (eg., Coedes 1968), with the concomitant presence of Hindu architectural styles and cosmologies, and the traces of Sanskrit vocabulary in languages of the regions.

At the same time, substantial differences can be ascribed to the two regions.

• language family distributions
Mainland Southeast Asia is a mosaic of intertwined language families. Austroasiatic, Austronesian, Hmong-Mien, Tai-Kradai and Tibeto-Burman languages are all found, with massively disjunct distributions, throughout most of the region. It is not possible, just looking at the distribution of different families, to identify exclusive ‘heartlands’ for the different families in the region. In some cases communities are distinguished from their geographically close neighbours by altitude and lifestyle, in others only by local custom and tradition. By contrast, the map of Island Southeast Asia is dominated by Austronesian languages, which are the only languages found in the Philippines and on most of the islands of Indonesia west of New Guinea. There are pockets of non-Austronesian languages, but the overall level of diversity is nothing like that accepted for mainland Southeast Asia.

• written traditions
Both regions were the recipients of Indic written traditions during the period of Indic empires (starting circa 2,200 years ago). This led to the development of local scripts, which flourished in local states. In mainland Southeast Asia these scripts continued in use and in development into modern times, allowing us to examine written histories spanning centuries if not millenia. In Island Southeast Asia the use of the Indic-derived scripts was either abandoned earlier on or else never so prevalent inscriptionally, leading to fewer historical records with the kind of detail that is found on the Asian mainland.
• political traditions

Both regions saw the rise of local states of various degree of durability. In mainland Southeast Asia some of these states, particularly away from the coast, involved extreme control of the local societies that they dominated, and ended up driving nonconformist members of their populations into more marginal environments (see Scott 2010). In Island Southeast Asia, and in coastal regions of mainland Southeast Asia, the independence of food production and personal mobility offered by a coastal locale, meant that state control was never as pervasive as it was on the mainland.

The question that we must address is why these differences in language distributions are found in two otherwise similar areas. It is trivial, but possibly relevant, to note that Island Southeast Asia consists of islands, while mainland Southeast Asia is a single land mass; but in what ways did these different geographic facts affect the social milieux of the two regions? In maps 9 and 10 we can see the difference in scale between one small area of northern Laos, and the vast area in which the Austronesian languages are described. Map 9 shows an area less than 200km from west to east, and containing 23 languages from four different families: Tibeto-Burman (red with shading), Kradai (light brown), Hmong-Mien (orange with dots) and Austroasiatic (green). The area occupied by Map 9 is shown on Map 10, which outlines the areas in which Austronesian languages are described. It is true that there are areas on the fringes of the Austronesian world, such as around the coasts of mainland Southeast Asia and New Guinea, where similar complexities can be found. What is striking is that all areas of mainland Southeast Asia where there are hills that could offer refuge from lowland states show the sort of diversity depicted in Map 9. It is clear that ‘the Austronesian world’ overwhelmingly consists of islands, which are highly accessible to a maritime culture. It must, however, be remembered that islands, too, have hinterlands, which have the potential to be ‘refuge areas’, where societies can escape the control of a large state; indeed, some of these islands are major land masses, such as Borneo, with extensive interiors and major ethnographic divisions between coastal and interior peoples. Further, since travel between two islands is most economical without stopping at intervening islands, even small islands have the potential to remain outside the social control of a surrounding polity, if they wish (see section 4 for more discussion). As with the mainland, the geography of Island Southeast Asia offers the opportunity for different socio-political entities that overlap in space.
Seeing the differences between the two maps, we must ask how such a large area of the world was so radically ‘Austronesianised’, without wholesale genetic replacement (see section 1, and also Donohue and Denham 2011, Denham and Donohue 2012b).
4. Accounting for language distributions

Mainland SEA exhibits many levels of linguistic integration: while there are many small, loosely-structured minority groups, there is also clear evidence for small elites achieving dominance in most of the lowland states (e.g., Munoz 2006, Osborne 2010 and others).

The spread of polities, or trading circles, is advantageous in many ways to many of the people who live within the bounds of the newly stable social unit. That does not, however, mean that the spread of a new social domain will be uniformly influential. Scott (2010) describes the decision to live on the margins of mainland Southeast Asian empires by numerous small groups that chose to escape the control of the plains’ polities. In the Indo-Malaysian archipelago the decision to not be included is an easy one: simply living on an island that is not connected to the others in a chain is enough to be left alone, given the way maritime travel can bypass islands more easily than stopping at all possible waypoints. This means that ‘controlled’ areas can be interspersed with regions that are outside the ‘control’ of a polity. Unlike the mainland, where nonconformist groups had to physically relocate to less accessible locations in highland regions (typically on the fringes of the lowland states), it is possible for a much more diverse, and less bounded, group of societies to occupy the same time and space in an archipelago region. In a typical mainland Southeast Asian setting (e.g., Lebar 1964 and others), emerging polities expand along river valleys and flat terrain. Ethnic groups that become refugees or discontents from these polities frequently come to occupy altitudinally differentiated zones, as shown schematically in Figure 1 (similar behaviour is also described for the Caucasus; Nichols 1998). While these zones lack the ease of movement of the flat lowlands, they allow the hill-dwellers to exploit a wider range of ecological zones, and so have access to resources that are denied to the valley-dwellers. From the valley-dwellers’ perspective travel between points in valleys, even when that requires a considerable detour from a straight-line path, is easier than negotiating the social and geographic hurdles imposed by the hills that lie between valleys. This continued contact makes for yet more cohesion along the valleys and reinforcement of the contiguous, though highly concave, geography of the lowland polity.
In an archipelagic society the expansion of a new polity is not constrained by the need to avoid particular terrain types when travelling. An emerging society (or social network) in a maritime environment can easily by-pass islands that are not included within its sphere, due to the ease of water-travel (the same is also true in other environments; see McAndrew et al. 1997, Stanish 2003, Shillington 2005 for examples in the Andes and the Sahara). In an archipelagic society it is quite possible, even likely, that at different times the expanding polity (or social network) would expand past areas that were not incorporated, due to the ease of travelling across the sea, and the ease of avoiding non-incorporated societies. Expressed from the alternative perspective, a non-incorporated society, the social equivalent of the hill-dwellers described earlier, can exist on an island that lies directly on the path between two archipelagic-societies settlements. On larger islands there may well be a hinterland that is not part of the coastal society, just as in the scenario summarised in Figure 1, but importantly travel between economically powerful incorporated coastal communities on different islands does not reinforce the archipelagic society, since the terrain traversed in such travel is sea, and does not have to involve passing through like communities. In Island Southeast Asia the polities of Srivijaya and Majapahit, and later Malacca, were this sort of polity (Munoz 2006, Borschberg 2010).

---

4 Depending on the winds at different times of year, it can easily be the case that it is simpler to travel a longer distance past unincorporated societies than a short distance to a fully incorporated group. This is shown in Figure 2.
Within Island SEA, we frequently talk of the widespread Austronesian family, and it is indeed incredibly widely spread, much more so than any other pre-modern language family. On the other hand, the very atypicality of Austronesian, with an incredibly diverse typological profile and a very conservative basic lexicon, leads us to examine what it means to be a member of this family, and whether membership in ‘Austronesian’ is of a similar nature to membership in (for instance) ‘Indo-European’ or ‘Nakh-Daghestanian’.

The field of historical linguistics recognises two kinds of criteria in establishing language relations (elaborating on Noonan 2010):

- Correspondences (in the phonological system, or in morphological paradigms); (importantly, correspondences in the lexicon are not of themselves sufficient criteria, as the lexicon is too borrowable; for recent discussion, see Donohue, Denham and Oppenheimer 2012a, 2012b; Greenhill and Gray 2012)

- The typological ‘nature of the language’ is the same kind (while this is neither a sufficient or necessary feature to establish genealogical relatedness, it is cited to exclude creoles from inclusion in the family of their lexifying parents – e.g., see discussion in Thomason and Kaufman 1988, Sebba 1997)

The nature of the sound correspondences in Austronesian languages has been discussed in Donohue (2013). To summarise the results presented there, the Austronesian languages of Indo-Malaysia sometimes show the same kind of profile for regular sound correspondences that characterises language families like Indo-European. Other individual languages, however, especially those found east of Sulawesi, show significantly lower levels of regularity. Maps 11 and 12, cited from Donohue (2013), shows the location of the extreme low-regularity languages, languages with such a lack of regularity in sound correspondences (and in most cases also a lack of reflexes of reconstructable morphological forms).\(^5\) (Donohue and Denham

\(^5\) The languages marked on the maps are, from least regular to most are: Bima (48%), Soboyo (62%), Ngadha (63%), Chru (67%), Kei (68%), Wandamen (68%), Biak (68%), Muna (69%) and Kambera (69%). Plenty of Austronesian languages in the Philippines and western Indonesian
2010 show the distribution of languages coded for the level of retention of Proto-Malayo-
Polynesian etyma.)

Map 11. Below 50% regular  Map 12. Below 70% regular

Two possibilities suggest themselves as explanations for this failure to achieve regular
sound correspondences (the two possibilities may be combined). The first solution is that
much of the basic vocabulary of these Austronesian languages should be considered to be
loanwords (from unidentified Austronesian sources). This acknowledgement of lower levels
of regularity is in keeping with the tradition of detailed studies of ‘speech strata’ described in
Austronesian languages such as Ngaju Dayak (Dyen 1956), Rotuman (Biggs 1965) and
Tiruray and Thao (Blust 1992 and Blust 2009, respectively). This would reduce the levels of
retained Austronesian lexicon even further than reported in Donohue and Denham (2010), and
would suggest that ‘lexical innovation’ (as viewed from the perspective of Proto-
Austronesian) was (erratically) rife in some parts of the Austronesian world. Given that one of
the characteristics of the Austronesian family is the extreme conservatism of the lexicon
(Wichmann forthcoming, Donohue and Denham 2010), this is a surprising finding. The
alternative possibility is that the languages in question have undergone large amounts of
erratic and unconditioned sound change. Either option suggests a language history that is not
the ideal one envisioned by advocates of the Neogrammarian hypothesis of regular sound
change.

Objectively assessing the ‘nature of the language family’ is not trivial, but neither is it an
insurmountable challenge. At the risk of missing many interesting and noteworthy quirks, we
can use the WALS database (Haspelmath et al. 2005) and the World phonotactics database
(Donohue et al. 2012) to test for features that occur with a significantly higher (or lower)
frequency in the family compared to its neighbours. For example, Indo-European can be
shown to be characterised by a number of typological traits which differ significantly in their
frequency when compared to other languages of Eurasia, the continental-scale comparison
(Table 3).6

---

6 Similar lists of defining traits can be assembled for other language families as well, though they are not presented here.
Given that it is possible to characterise a language family as sharing certain broad typological tendencies, what does this method yield when applied to the Austronesian languages, compared to their region (Southeast Asia and the Pacific)? Without needing to resort to a table, we can summarise the distinguishing characteristics quite simply: Austronesian languages show a tendency for head-initial syntax (first pointed out in Foley 1998). This is manifested by VO order at the clause level, prepositions rather than postpositions, and a N-modifier order in NPs. In contrast to the thirteen morphosyntactic traits that are distinctive for Indo-European, there is only one for Austronesian, head-initiality. In contrast to the nine phonological traits that are typical for Indo-European languages, not a single phonological trait characterises the Austronesian family as a whole.

However we choose to think about it, it is clear that the Austronesian family does not constitute a language family that coheres, lexically, structurally or typologically, in the same way that a well-established language family like Indo-European does. At the same time, the fact that ‘well-behaved’ languages are interspersed with erratic languages (from the perspectives examined here), without regard to geographic or subgrouping continuity, implies that there have been multiple pathways leading to language communities becoming Austronesian. In some cases we see the spread of language along with language communities; in others it is more parsimonious to think of the dispersal of Austronesian language traits to existing communities, which have acquired ‘Austronesianness’ to different degrees and in different ways. In some cases there are clear traces of a pre-Austronesian stratum, such as is seen in the distribution of agreement in Austronesian languages in and near New Guinea. For instance, 90% of Austronesian languages on or near New Guinea show verbal agreement, compared to 29% outside this region; this represents a highly significant difference (p < 0.0001) in the sample (n = 227). In other cases the absence of inflectional morphology, as discussed in section 2, combined with the failure of the language to show regular sound
correspondences at an ‘adequate’ level (Donohue 2013), suggests that creolisation played a significant role in the dispersal of Austronesian traits.

Where creolisation is not obviously implicated in the formation of the modern languages, we can still observe that language shift was a common process involved in the dispersal of Austronesian languages. Far from being exceptional, language shift was normal, though sporadic, involving at times (and places) imperfect learning of the new Austronesian languages depending on local social circumstances. Evidence of this putative substrate varies from language to language, and can be (not exhaustively) illustrated by examining three very crude dimensions of variation: the lexicon, the phonology and the morphosyntax. We have mentioned that head-initiality is the only trait that can be used to identify Austronesian languages in their geographical context. If we are to examine just those features that are typical of (reconstructed) Proto-Austronesian and Proto-Malayo-Polynesian, and of several languages ‘high’ in the genealogical tree, we can obtain an idea of what is and what is not typologically ‘typical’ of an Austronesian language. The different dimensions of the cube shown in Figure 3 are defined as follows:

Lexicon
- Examining a basic wordlist gives us an idea of the degree to which the basic lexicon was retained in any given language (eg., Donohue and Denham 2010), combined with the regularity of sound correspondences (as described in Donohue 2013), tells us how directly the lexicon of the language reflects Austronesian etymological sources.7

Phonology
- The phonological system of a language can be compared to the phonological system of Proto-Malayo-Polynesian, and any unexpected (or diachronically irregular) phoneme series (such as the voiceless prenasalised series /mp nt ntʃ ŋk/, or in other languages the loss of manner contrasts in stops) can be considered to be evidence of a non-Austronesian character. We examined ~200 phonological traits per language (see Donohue et al. 2012).

Morphosyntax
- Given what is known of the syntax of those languages ‘high’ in the Austronesian tree, in Taiwan, and the syntax of the languages in question, various traits (for example, OV word order, agreement, possessive classes, or an isolating nature (as per section 2) are traits of a non-Austronesian character; just as the loss of the Austronesian voice

---

7 Under ‘lexicon’ we include the lexical entries for productive, bound morphology, the retention of which is at least as indicative of origins as are free lexemes. In the case of non-productive, or fossilised, forms that show forms similar to productive morphemes in other languages, the analysis is less certain because we cannot assume that non-productive morpheme look-alikes have been directly inherited, rather than borrowed. Many languages of eastern Indonesia and Timor Leste show few, if any, retentions of reconstructed Austronesian bound morphemes, or else exhibit these morphemes with non-productive functions. The large number of languages with agreement prefixes the form and function of which are likely to be cognate with Austronesian genitive morphemes (see Wolff 1996 and others) are witness of a strong and enduring Austronesian influence, but the fact that these prefixes are not reconstructed in Austronesian makes their evidence equivocal.
system evidences a less Austronesian character. We used the ~150 morphosyntactic traits found in the WALS database (Haspelmath et al. 2005) as a basis for comparison.

These three dimensions, crude as they are, allow us to visualise the different extents to which different languages can be said to be Austronesian. In Figure 3, along with Map 13, we can see how these different dimensions of ‘Austronesian-ness’ are occupied by a selection of different languages from across southern Island Southeast Asia. Bima and Kei are both low on all three of the scales examined, while Sangir (and languages to the north) are high on all scales. Languages such as Rote and Sika have ‘well-behaved’ lexicons, but show very non-Austronesian phonological and morphosyntactic traits. Gayo, in the top left rear corner, is as exemplary as Sangir structurally, but has a lexicon in which substantial elements have been acquired by indirect inheritance, or borrowing. In Kowiai the phonology and the lexicon are ‘well-behaved’, but the morphosyntax is highly aberrant, from a conservative Austronesian perspective.

Figure 3 and Map 13. The many dimensions of Austronesian-ness.

Key:
1: Bima, Kei (W, E)
2: Sika, Rote (W, E)
3: Buru, Kowiai (W, E)
4: Sangir, Philippine lgs
5: (Old) Javanese
6: Banggai
7: Gayo
8: Muna
9: Wolio
10: Bugis

---

The languages of the Philippines fill the same position as Sangir in this cube, being exemplary in most ways.
Importantly for the model we discuss, the ‘aberrant’ and ‘exemplary’ Austronesian languages in Island Southeast Asia are not separated discretely by geography. In the Philippines and in much of western Island Southeast Asia we generally find ‘exemplary’ languages, but within Wallacea we find an apparently random distribution of exemplary and aberrant languages, the two (non-discrete) categories mixing and intermingling on the contemporary map. In Figure 4 and Map 12 we can see that the languages discussed in other chapters of this volume, all showing low levels of ‘Austronesian-ness’ in at least one dimension, are widely dispersed.

Figure 4 and Map 14. Austronesian languages of this volume plotted in the cube.

Key:
1: Colloquial Javanese
2: Kéo
3: Riau Indonesian, Papuan Malay (W,E)
4: Minangkabau

5. Conclusions
We conclude that social admixture was the norm, not the exception, in the dispersal of Austronesian languages across Indo-Malaysia. Rather than a ‘wave’ of Austronesianisation rolling out over a simple and unsophisticated pre-Austronesian social milieu, an already integrated region saw the appearance of a new and attractive linguistic code, and that code was variously adopted or not adopted by different societies, as they saw fit. As with all social innovations, the spread was not uniform, and did not follow uniform pathways. In some areas
there was wholesale language replacement, presumably where the social pressure or social incentives for acquiring the new linguistic code were strong; in other areas the Austronesian content was adopted only much later, and much more haphazardly. We can infer that in some communities the process of ‘Austronesianisation’ took generations, and followed an almost reluctant path that did not see the wholesale loss of local structural or typological characteristics, nor the adoption of Austronesian forms from a single Austronesian parent.

We propose that languages listed as Austronesian form a spectrum of different non-discrete types, and different characteristics can be used as evidence of some of the different social processes that led to ‘becoming Austronesian’. The logical conclusions are that the languages classified as Austronesian (LCA) deserve re-examination, with a number of different outcomes (vertex references refer to the vertices found in Figure 3).  

- Some LCAs are members of the Austronesian family in the sense normally accepted by practitioners of the comparative method (vertex 4);
- Some LCAs show all the structural characteristics of creole languages, and should be considered to be creoles, rather than the outcome of regular, uninterrupted inter-generational transmission (vertices 1, 2 and 3);
- Some LCAs can be considered to be members of the Austronesian family, but with such substantial pre-Austronesian substratal structural properties that they can perhaps better be considered to be mixed languages, or non-Austronesian languages that have been remodelled via metatypy (e.g., Ross 2006) (vertices 1, 2, 5 and 6);
- Some LCAs show the right structural characteristics for a member of the Austronesian family, but with clear evidence that the vocabulary was not derived from a single source, implying that regular inter-generational transmission was not the process of the formation of the modern language (vertices 5 and 7);
- Some LCAs show so few characteristics of Austronesian languages, with such irregular sound and morphological correspondences, that they should be considered to not belong to the Austronesian family, but rather should be thought of as being the vestiges of pre-Austronesian families once dispersed about Island Southeast Asia that have been affected by the inflood of Austronesian lexical items, and, in some cases, Austronesian structural characteristics (vertex 1).

The creole (re-)classification of many AN languages means that we should reconsider what the ‘minimal requirements’ are to be counted as Austronesian, or that we should more readily consider creole languages as members of the language families of their (primary?) lexifiers. This would however miss the point that such languages have multiple antecedents, and that a classification which truly reflects a language’s social history should include information about the different sources of the different modules of that language. It also misses the point that showing some evidence for inclusion in a particular language family

---

9 Note that we are not proposing a 1:1 mapping relation: the cube representation conceals the different kinds of data that fit into the broad categories ‘lexicon’, ‘phonology’ and ‘morphosyntax’, each of which logically has its own history. Without more detailed work we cannot arrive at more precise set of hypotheses about the linguistic and social histories of the different communities.
does not imply that membership in that language family is the only possible classification of that language.

The more isolating languages present a particularly interesting perspective on the nature of the dispersal of Austronesian linguistic structures across Island Southeast Asia. We know that a strongly isolating morphosyntactic character is not an archaic Austronesian characteristic. We have described three possible broad social scenarios that might underly to the heavily isolating languages (vertices 1 and 2 in the cubes); these implicate creolisation, heavily substrate, and/or misclassification. If we are dealing with a creolisation scenario we can simply appeal to the well-known process of simplification due to imperfect second-language acquisition by adults. If we are discussing LCAs that are really’ (or, ‘better classified as’) non-Austronesian, or if we are supposing a heavy substrate, then we have to account for the repeated emergence of isolating characteristics in the Austronesian languages across Island Southeast Asia, but not outside this region.

Of course, the important point is not that this is a characteristic of the Austronesian languages of the region, but rather a characteristic of languages of the broader Southeast Asia region, and its preponderance for heavily isolating behaviour (as described in section 2). If we project the contemporary linguistic situation back, which seems reasonable to judge from the arguments presented in section 3, then we would have to assume that many of the languages of the region that were spoken in place prior to the dispersal of Austronesian languages also showed an isolating profile. The reasons for this can be speculated upon; it might be that sustained interaction over a long time led to continual creolisation scenarios, but that cannot be ascertained with certainty, nor is it particularly relevant to a discussion of the contemporary languages. What is relevant is the reasonable assumption that many, or at least some of the, pre-Austronesian substrata across much of Island Southeast Asia had a large number of isolating characteristics. This means that we can suppose that at least some of the more isolating characteristics of many of the contemporary Austronesian languages (some of which are described in this volume) can be attributed to contact with languages that were already strongly isolating (or, of course, that some of the contemporary languages in question are those strongly isolating, pre-Austronesian languages, with an Austronesian veneer applied). This in turn gives us a model that allows us to hypothesise about the observed trend towards an isolating profile in colloquial varieties of large languages in Island Southeast Asia. If there is already a (long-standing) linguistic ecology that promotes isolating characteristics, then multilingual speakers will see that as a model that might in some cases be explicitly targeted, and will certainly not be avoided.

The reclassification of many of the LCAs in Island Southeast Asia, particularly in eastern Indonesia (see Map 12), as not belonging to the Austronesian family will make the language maps of that part of the archipelago more closely resemble those drawn for mainland Southeast Asia (see Map 9). The same patterns of intermingled families, lack of contiguous territories, and extensive bilingualism (leading to much shared vocabulary and many shared structural characteristics) are common between mainland Southeast Asia and Island Southeast Asia. While the written traditions of the two regions are different, the other points of difference noted in section 3, namely those involving different written and political traditions,  

---

10 Delancey (2010, in press) offers perspectives on the genesis of many of the subgroups of Tibeto-Burman that are highly congruent with the conclusions reached here.
and differences in the description of the distribution of language families, can in part be the result of different traditions and methodologies in the practice of linguistic classification rather than reflecting deeply-rooted social differences between mainland and Island Southeast Asia.

References


