

The role of contact and language shift in the spread of Austronesian languages across Island Southeast Asia

Mark Donohue and Tim Denham
The Australian National University

Abstract

The spread of modern humans into and across Indo-Malaysia and the Pacific represents the earliest confirmed dispersal of humans across a marine environment, and involved numerous associated technologies that indicate sophisticated societies on the move. The later spread of Austronesian over the region shows language replacement on a scale that is more reminiscent of the period of state-sponsored European colonisation than of a social landscape more than three millennia old, and yet the Austronesian languages present a typological profile that is more diverse than any other large language family. These facts require investigation. This chapter examines the disciplinary dimensions that offer separate, but intertwined, histories of the region. We point out a number of ways in which the dispersal of Austronesian languages, originating in Taiwan, should not be portrayed as a technological and demographic steamroller. This involves discussion of the nature of pre-Austronesian society and language in the south-west Pacific, and the degree to which it has and has not changed following ‘Austronesianisation’.

0. Introduction

In this chapter we set out to show that simple explanations in linguistics must be treated cautiously. We will examine the linguistic diversities that can be found, along more than one dimension depending on which kind of data is under scrutiny (e.g., lexical, phonological, morphological,...), in Indo-Malaysia, and show that independent lines of evidence about language and population dispersals across the region lead to different conclusions, depending on which discipline we examine. Not only is there little congruence between different disciplines, but different lines of argumentation within the one discipline can lead to different conclusions (e.g., Oppenheimer 2004, Donohue and Denham 2010, Denham and Donohue - LDC 2012).

As a result of the argumentation presented in section 1, confirmed by the case study shown in section 2, and based on previous inter-disciplinary surveys of alternative views on culture history in Island Southeast Asia (ISEA), we extrapolate some social scenarios for the settlement events that can be detected for ISEA and the Pacific (particularly Near Oceania) in section 3. We note that, methodologically, human histories are complicated, and human history in ISEA (and elsewhere) is far from monolithic. Rather than revolutions that swept earlier ‘stages’ of settlement history away, we see layers building on earlier traditions and being combined to produce multi-stratal linguistic and social histories for the region. Rather than assuming that data from different disciplines, or different kinds of data within the one discipline, should *a priori* be reconciled, we suggest that a more complex view of human history is more likely.

1. Linguistic divergence

The ‘Austronesian world’ extends from Madagascar, through ISEA to Taiwan, across New Guinea and to the corners of the Pacific. In Remote Oceania, Austronesian languages exist primarily in isolation from other language families (although see Donohue and Denham 2008, Blust 2008). However, in other regions, languages of the Austronesian family have been in contact with numerous other language families, some big, some small. Some of the families with which Austronesian has come into contact are listed in (1)

- (1) Tibeto-Burman (mainland south/southeast Asia);
- Kradai (mainland southeast/east Asia);
- Austroasiatic (mainland southeast Asia);
- Tambora (Sumbawa, south-central Indonesia);
- Timor-Alor-Pantar (Timor and nearby islands, south-east Indonesia and East Timor);
- Trans New Guinea (mainland New Guinea and off-shore islands);
- West Papuan (Halmahera and Bird’s Head of western New Guinea);
- Skou (north-central coast of New Guinea);
- Torricelli (north-east coast of New Guinea).

The (partial) list in (1) begs the question of what it takes to be a language family, that is, what are the linguistic criteria for assigning languages to a particular family, rather than another, or declaring a new family to exist?

Languages classified as members of the same language family are ‘... speech varieties related through a common ancestor’ (Noonan 2010). Linguistically, the proof for this is through the existence of shared sound correspondences, or morphological paradigms. Without these constraints we can easily fall into the trap of allocating languages to a family on the basis of traits shared by diffusion. For instance, it is widely acknowledged that similarities in vocabulary or gross typological traits (without sound-to-meaning correspondences) cannot be taken as proof of genealogical relatedness (e.g., Campbell and Poser 2008, Donohue and Musgrave 2007, Donohue et al. 2008, Donohue et al. 2011, Joseph and Janda 2003, Wichmann and Saunders 2007, amongst others).

Equally, the question of language affiliation when ‘regular intergenerational transmission’ does not apply, must be addressed. Numerous cases of ‘irregular’ language spread or language development have been subsumed under the rubric of ‘pidginisation’ or ‘creolisation’ (e.g., Ansaldo et al. 2007, DeGraff 2003, Donohue and Denham 2010, Holm 1988, Mufwene 2002, Weinreich 1953). These involve what turn out to be regular sound correspondences, sometimes some paradigmatic correspondences in the morphology, but a massive simplification of the phonological and grammatical structure of the language to

which the pidgin/creole might potentially be related. Noonan (2010), citing Croft (2000) and Thomason and Kaufman (1988), asserts that

‘certain linguistic features ... must be transmitted along a genetic line for a language to be considered a member of a given taxonomic unit. ... language relatedness is assessed along chains of transmission of these features from mother language to daughter language.’

Indeed, in a discussion relevant to the ISEA world, Campbell and Poser (2008: 61) note that ‘Humboldt emphasized grammatical structure as the criterion of linguistic affinity and determined that Kawi was ‘Malayan’ (Austronesian) in spite of the vocabulary of Sanskrit origin it contains’ (2008: 61, referring to Humboldt 1836-1839). Namely, the ‘essential character’ of the language in the Kawi inscriptions was not in keeping with the clues towards Sanskrit affiliation provided by the vocabulary, but matched closely a pairing with (Classical) Malay and related languages.

Working with these two evaluative tools, ‘regular correspondences’ and ‘essential nature’, we can construct a matrix that neatly represents a number of types of scenarios for language similarity. Examined like this, it is clear that there are more options than simply ‘related/unrelated’, and that neither regular correspondences nor typological traits alone can be proof of the kind of uninterrupted intergenerational transmission that lies behind normal language family relations.

Table 1. Kinds of language relationships

	+ regular correspondences	– regular correspondences
+ ‘essential character’	(normal genealogical relatedness)	Heavy contact scenario
– ‘essential character’	Pidgin/Creole scenario	(unrelated)

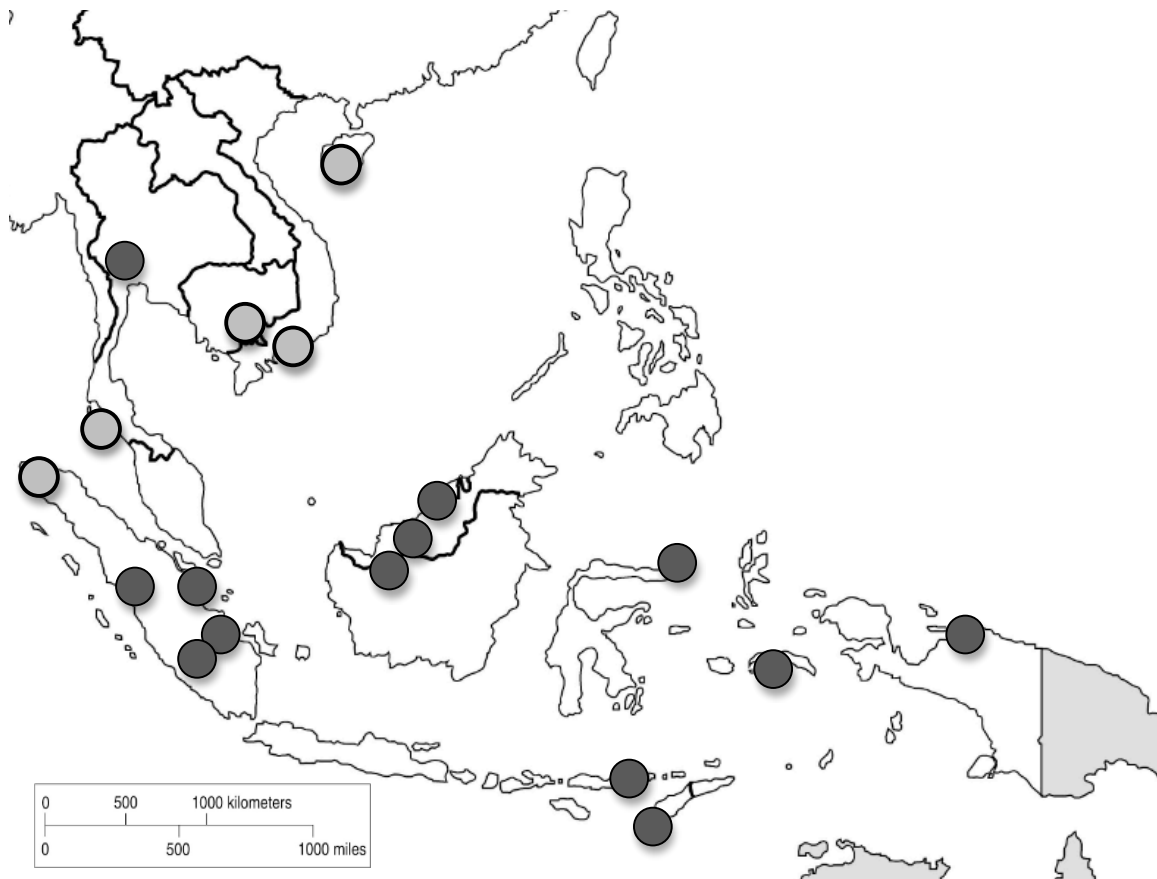
While Noonan’s presentation is unconventional, the essential relationships that he raises, which we have summarised in Table 1, have been widely employed. The process of examining regular correspondences in sound systems or morphological paradigms is well known, and need not be described here (again, see papers such as Joseph and Janda 2003, Campbell and Poser 2008, and all other reputable work on historical linguistics since the Neogrammarians, such as Brugmann and Delbrück 1897-1916). The notion of examining the ‘essential character’ is less conventional, and begs the question of what degree of variance can be considered ‘normal’ in a language family. Investigating norms in the area of grammatical variation within a family is research that is still waiting to be carried out, though initial work (Donohue 2012, 2013) suggests that it is a tractable problem. Importantly, we can examine languages along a dimension (or dimensions) of typological variance, and then compare the results relatively, and with respect to the subgrouping that we arrive at through the application of the comparative method.

The important question is how the different linguistic data that emerge through this typology correspond to different possible social scenarios. The answer must come from case studies calibrated against known histories or known historical events. Here we present one such case study, examining the spread of the Malayic and Chamic languages.

2. Malayo-Sumbawan

The Malayo-Sumbawan subgroup of Malayo-Polynesian offers us a good opportunity to test hypotheses about the nature of language spread and demographic spread in ISEA. Within Malayo-Sumbawan the existence of the Malayic and Chamic subgroups is beyond question, and it is on these that we concentrate here. Both of these subgroups have been established by traditional methods (e.g., Adelaar 1992, Thurgood 1999), and both subgroups have written historical traditions extending back many centuries. For Malay we have clear inscriptional evidence dating to the 7th century AD (e.g., Coedès 1930, Haaksma 1933, Humboldt 1836-9, de Casparis 1975, Munoz 2006, Shaffer 1996), and for Chamic we can see a similar time period (e.g., Morrison 1975, Thurgood 1999, Stuart-Fox 2003). The distribution of the 18 languages (13 Malayic, 5 Chamic) examined here is shown in Map 1.

Map 1. The Malayic and Chamic languages included in this study



Key: filled circle represent Malayic languages, hollow circles are Chamic languages. The Chamic languages are, west to east: Acehnese, Urak Lawoi', Western Cham, Eastern Cham, Tsat. The Malayic languages are: Nonthaburi Malay (on mainland Asia), Minangkabau, Indonesian, Kerinci, Jambi Malay, Mualang, Iban, Melanau, Larantuka Malay, Kupang Malay, Manadonese, Ambonese Malay, Papuan Malay.

These languages represent the modern descendants of the expansion of a trading network that was centred in the Malay peninsula, and which was driven by the demand for

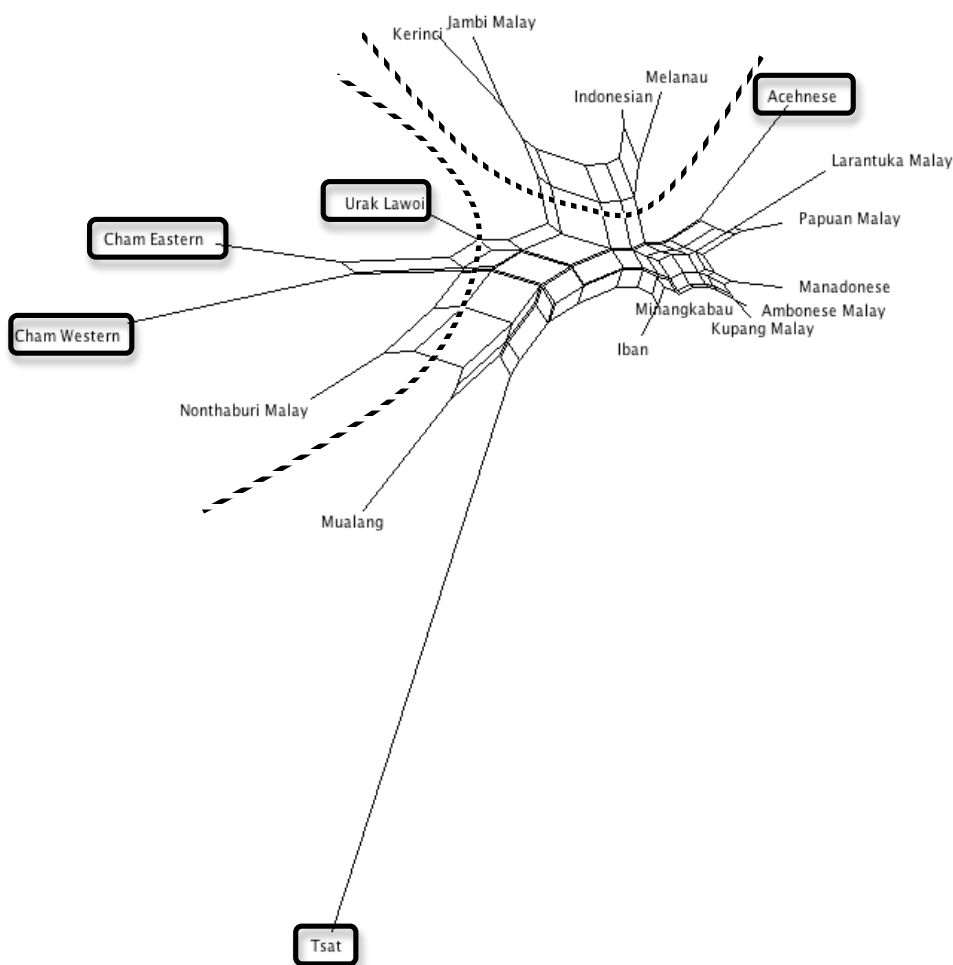
goods, particularly spices, that could be extracted from eastern Indonesia and exported to South Asia, the Middle East, and eventually Europe (see Munoz 2006).

2.1 A typological comparison of Malayic and Chamic

For each of these languages we were able to compile complete phonological inventories, including all segmental and suprasegmental contrasts, information on gross syllable structure and on constraints on the kinds of segments that can occupy different positions in the syllable and word. When these data were coded in a way to be compliant with the clustering algorithm in Splitstree (Huson and Bryant 2006), we ran the 191 contrastive phonological and phonotactic features, with the resulting dendrogram shown as Figure 1.

In Figure 1 the Chamic languages are shown in boxes. The first thing to note is that, when clustering by phonological traits is considered, the Chamic languages are not clustered separately from the Malayic languages. Indeed, the only cluster that includes all of the Chamic languages will include all of the languages in the dataset. We do, however, find other clusters emerging: two Chamic languages, Urak Lawoi' and Nonthaburi Malay, form a cluster, loosely affiliated with Mualang, Kerinci, Jambi Malay, Indonesian and Mualang form a separate cluster, and, the eastern Indonesian Malay varieties, plus Acehnese, form another distinct cluster (with Iban and Minangkabau loosely affiliated).

Figure 1. Malayic and Chamic languages: Phonological clustering



It is clear that genealogical groups (Malayic vs. Chamic) established through the comparative method techniques of sound and morphological correspondences, are not reflected in the phonologically-defined clusters shown in Figure 1. We will return to the question of what underlies this difference at the end of this section, but present an interim explanation in the form of Map 2. In this map the clusters shown in Figure 1 have been shown against the locations of the different language communities from Map 1. It is clear that several of the clusters that emerged in Figure 1 reflect geographic, and not genealogical, similarities. The cluster consisting of the Chamic languages, Urak Lawoi' and Nonthaburi Malay, represents a geographically and socially well-defined group, all being found in the mainland Southeast Asian linguistic area (Enfield 2005), all have been subject to the normative influence of other languages with similar phonological typologies in that area. The Malay varieties of eastern Indonesia all show similar traits, traits that are not shared with their western Indonesian cousins.

Tsat is the most isolated Chamic language in Figure 1. It is geographically isolated on Hainan Island, to the north of the rest of the family, and importantly in the middle of a different set of social networks involving different languages and different normative criteria. The other 'errant' Chamic language in Figure 1, Acehnese, is clustered at a distance from other Chamic languages, but very close to the other Sumatran languages, with which it has interacted for the last millennium.

Map 2. Phonological clusterings among the Malayic and Chamic languages

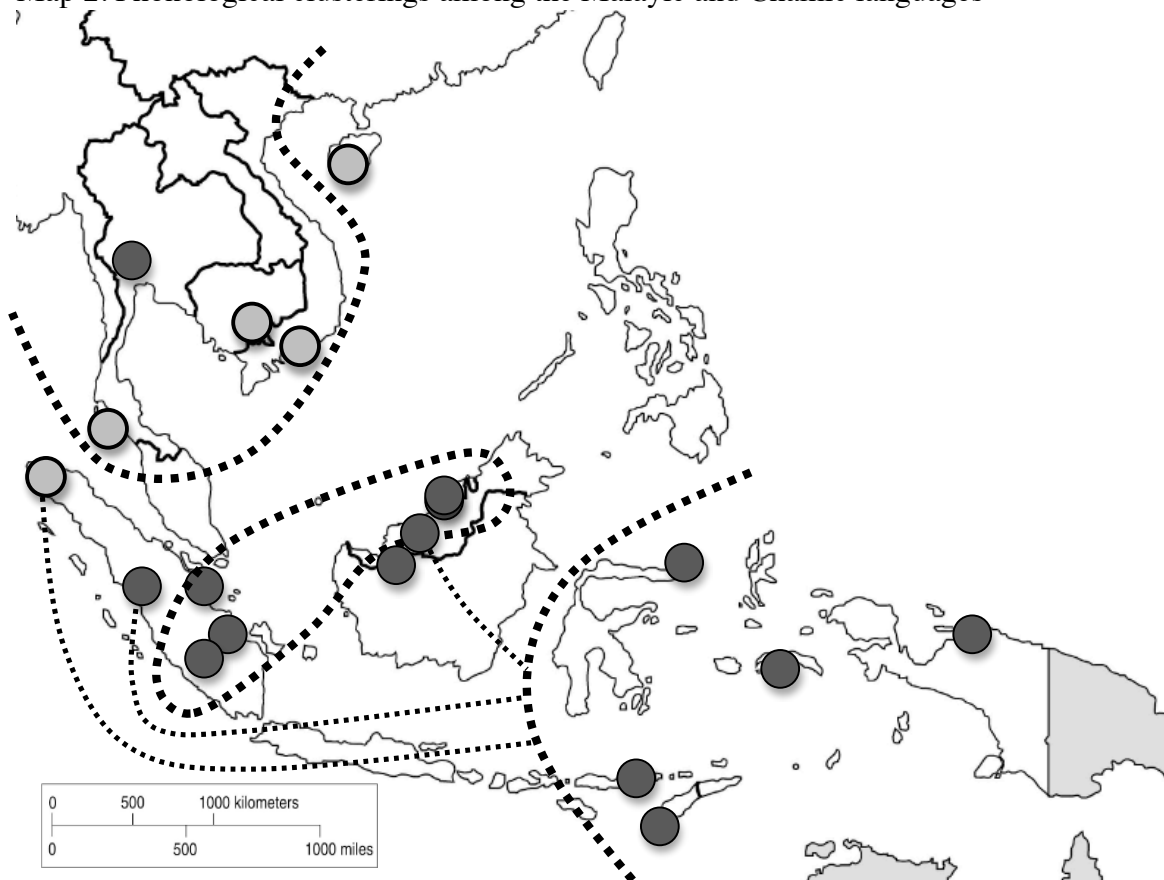
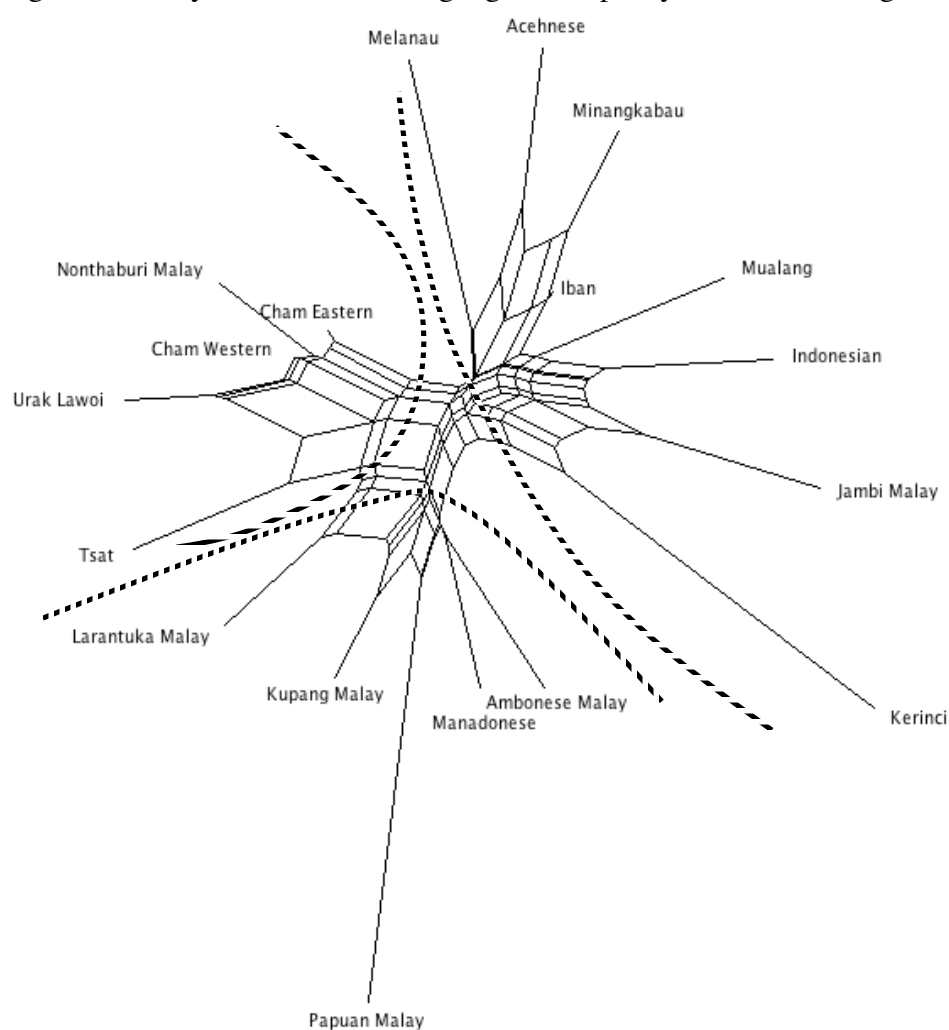


Figure 1 presented the results of clustering with as complete a phonological summary of the different languages as could be assembled. To examine morphosyntax we face a more open-ended question, and we have chosen to limit the data examined to the features coded for the *World Atlas of Language Structures*. We are not claiming that the features included in *WALS* are in some way ‘optimised’ for typological comparison, but we do note that they have been selected for typological breadth, and even more importantly they are a set of features that has been chosen independently of this study (and so cannot be accused of selectional bias – Donohue et al. 2011). They can be taken as representing an incomplete, but unbiased, set of perspectives on typological diversity. When we examine the clusters that derive from these features, we find Figure 2.

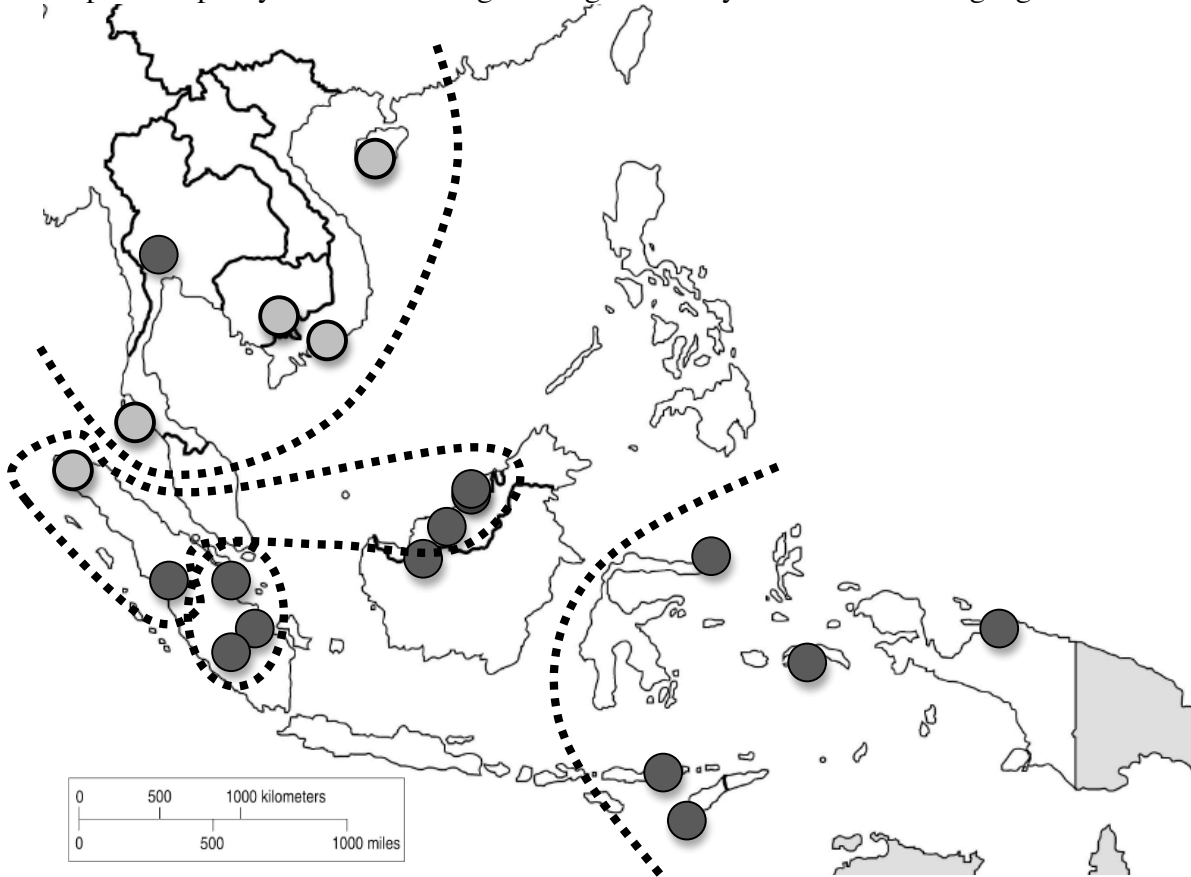
In Figure 2 we can first note that again the Chamic languages are not separated from the Malayic languages. Acehnese appears most closely related to Minangkabau, its southern sister in Sumatra, and Nonthaburi Malay appears more typologically similar to the other Chamic languages. The eastern Indonesian languages form a cluster on their own, and a fairly well-defined cluster brings together the languages of Borneo and Sumatra. The clusters are not only different from the genealogical clusters, but they also differ in numerous ways from the clusters that resulted from examining the phonological data (summarised in Figure 1).

Figure 2. Malayic and Chamic languages: Morphosyntactic clustering



As with Figure 1, we can readily find geographic explanations for the major clusters that emerge in Figure 2. It is important to note that it is not Euclidean space or physical geography that is responsible for similarities between languages. Rather the clusters seem to correspond to social space, namely, the extension of social networks across space.

Map 3. Morphosyntactic clusterings among the Malayic and Chamic languages



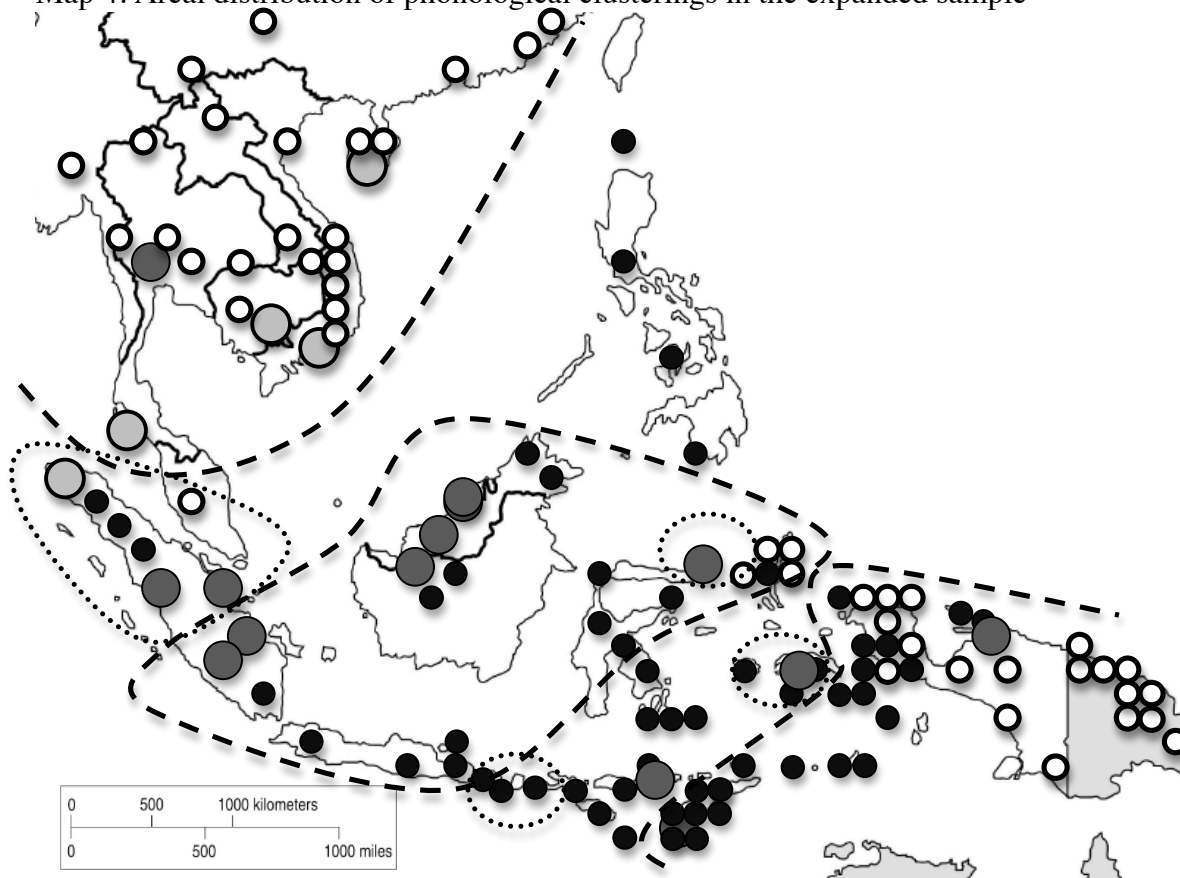
2.2 Malayic and Chamic in their areal context

Examining the correspondences between typological clusters and geographic proximity, we come to the strongly supported conclusion that areality rather than genealogy is behind the clusterings. The question remains, why are the mainland languages distinct from the Sumatran-Borneo languages, and the latter different from the eastern Indonesian languages. To answer this, we increased the sample of languages we were examining, to allow us to observe surrounding languages, both related and unrelated.¹ We shall just summarise the results on two maps, showing the clusters based on phonological and morphosyntactic traits.

¹ See Appendix 2 for a list of languages in the expanded sample. Clustering was performed with the same method and the same datasets that was described in 2.1; in this chapter the actual clustering dendrograms for Maps 4 and 5 are not shown here, as they involve a very large number of languages and it is the areal patterns that we are most interested in here.

In Map 4 the clusters for phonological traits are mapped out. The different clusters are shown with dashed lines, delineating clearly areal groupings, with the exception of the dotted lines. These four regions represent a single cluster in terms of typological features, one that is dispersed across different physical regions. We can still describe this discontinuous cluster in terms of social geography: the cluster is dispersed across a number of regions that are all on the fringe of the core Malay-Javanese cultural zone, centred around the Java sea and including southern Sumatra, Borneo, and western Sulawesi.

Map 4. Areal distribution of phonological clusterings in the expanded sample



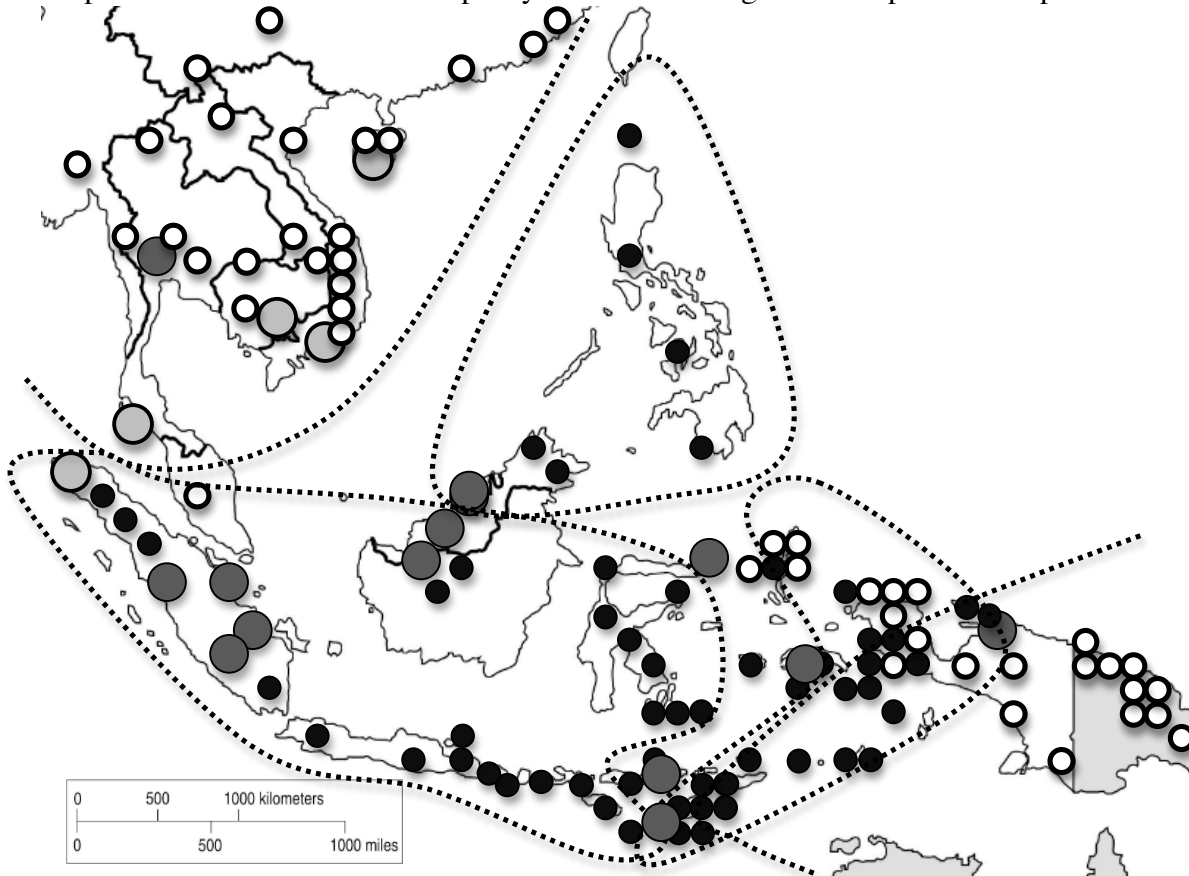
Key: large circles represent Malayic and Chamic languages, following the key to Map 1. Small filled circles are non-Malayic and non-Chamic Austronesian languages. Small white circles with borders show non-Austronesian languages. Not all languages included in the expanded sample are shown on the maps.

Map 4, like Map 2, shows clearly areal influences in the clustering of languages. In Map 4 the presence of languages that are not only non-Malayic and non-Chamic languages, but also non-Austronesian languages, underscores the areal rather than genealogical basis for the typological clusters (only the four central clusters consist solely of Austronesian languages). The Malayic languages do not form a distinct clade in the network, but rather are distributed across all of the large typological clusters. The Chamic languages are more discrete, but we still find the same split between Acehnese and the mainland languages that emerged in Figures 1 and 2. In all cases the Malayic and Chamic languages form a typological ‘fit’ with their geographically proximal languages: the Chamic languages conform to a Southeast Asian profile, with the exception of Acehnese on Sumatra, which conforms to a

Sumatran profile. Papuan Malay in the far east clusters with the Papuan and Austronesian languages of New Guinea, and not with the other Malayic languages.

When we examine the map displaying clusterings based on morphosyntactic traits we find very similar patterns (Map 5). Malayic and Chamic are not respected in the clusters, and Austronesian and non-Austronesian languages are mixed together in clusters that reflect areal distributions, and not genealogical ones.

Map 5. Areal distribution of morphosyntactic clusterings in the expanded sample



In sum, we can see that the different typological profiles that we find in the Malayic and Chamic languages simply reflect the different areas in which they are found.

How, then, does this inform our understanding of the relationship between social histories and linguistic data? Recall that we are attempting to understand the historically recent spread of the Malayic and Chamic languages. Two (kinds of) logical scenarios present themselves, shown in outline in Table 2.

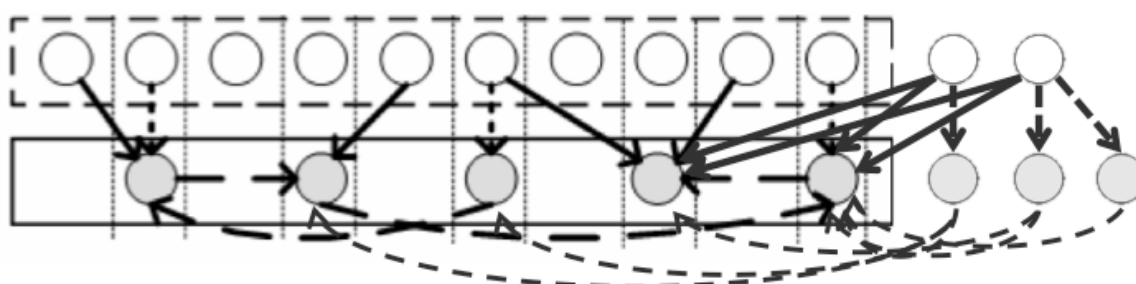
Table 2. Scenarios accounting for the diversity within Malayic

	Scenario A	Scenario B
Motivation for spread of the language subgroup	Malayic languages spread with the spread of Malayic socio-economic influence.	Malayic languages spread with the spread of Malayic socio-economic influence.
Demographics of the language spread	Speakers of Malayic languages moved to new areas with that growing socio-economic influence.	Speakers of local languages relexified to participate in the new Malayic socio-economic sphere.
Motivation for the observed typological diversity	Malayic language speakers acquired certain linguistic traits from the local languages of the areas they moved to.	As language shift to Malayic proceeded, adult learning led to a Malay-lexified language with the typological characteristics of the local language(s) of the new areas.

Both of these scenarios offer plausible sketches that account for the observed data. Written records tell us that B is correct (e.g., Adelaar 1992, Adelaar and Himmelmann 2005, and other works): the Malayic languages achieved their greatest diffusion in space by a process of language shift from the pre-Malayic languages of different regions. The shift happened via a process of adults (imperfectly) learning the new linguistic code, but applying a largely Malay lexicon, and then transferring this mixed language to the children of the community.

We envision a state of affairs that can be modelled as in Figure 3 (adapted from Gong 2010). In this figure three kinds of language transmission are shown: Vertical Parent-to-Child transmission, with a vertical dotted line, Horizontal Child-to-Child transmission with a dashed line, and Oblique Adult-to-Child transmission, with a solid line. The existing language community (in boxes) has regular intergenerational transmission from adult to child, and from adults to other children, innovation spreads through horizontal transmission between children. When a socially dominant outsider group arrives (the two white circles on the right of the figure, their role in the spread of change language outstrips that of other language users. There is more oblique transmission, but also a greater number of children who acquire linguistic patterns directly. Since these children grow up in a more prestigious way, their language too influences that of their own generation. In this way language shift can be rapid and dramatic, but, equally, it will not necessarily be pervasive across the entire range of modules of language. The most easily mutable elements will change the fastest: the lexicon, the order of words, for instance. Given the small number of outsiders arriving, many elements of the original language of the community will be preserved (Donohue and Denham in press).

Figure 3. The Oblique disruption of Horizontal and Vertical language transmission.



Given that observed occurrence of this form of language spread in historical times, involving elite dominance and a cultural, but not demographic, change from without, we must conclude that this is an equally plausible scenario for the diffusion of Austronesian languages across ISEA, as suggested in various works (e.g., Donohue and Denham 2010, Tumonggor et al. 2013).

We are again face to face with diversity: just as there are diverse ways of measuring or assessing relations between languages, so too there are diverse ways in which languages can be, and have been, counted as related. We know that ‘culture’ is not monolithic, and that social histories cannot be reduced to sound bites. Similarly, languages should not be assigned in simplistic terms to single-line histories, but need to be examined in terms of as many different kinds of linguistic data as can be amassed.

3. Interdisciplinary perspectives

Multidisciplinary research provides temporal depth and broader social contexts for the processes of language dispersal within ISEA. The weight of linguistic data and interpretation casts Taiwan as the homeland for the subsequent dispersal of Austronesian languages across ISEA (Blust 2009). Although the timing, cultural associations and linguistic structure of this dispersal are much contested (for example, see debates in Donohue and Denham 2010), a source location on Taiwan is not.

The timing of Austronesian language dispersal is associated with the movement of material cultural traits, often portrayed as farmer-voyagers (Bellwood 2005, 2007), across the Batanes Strait from Taiwan to the Philippines around 4000 years ago (eg, Piper et al. 2009). Although there was certainly contact between Taiwan and the Philippines from at least 4000 years ago onwards, with likely colonisation of the Mariana Islands from the Philippines around 3500 years ago (Hung et al. 2011, Carson et al. 2013), the processes for the dispersal of Austronesian languages across the rest of ISEA are unclear.

Re-evaluations of the migratory routes (Carson et al 2013) and the establishment of the Lapita culture in the Bismarck Archipelago at 3470-3250 cal BP (Denham et al. 2012, Specht et al. 2013) open up new vistas for the interpretation of ISEA history during c.4000-3000 years ago. Conventional portrayals have depicted the dispersal of Austronesian languages southwards to the Philippines, with a subsequent rapid radiation across ISEA, movement northwards of New Guinea to found the Lapita culture in the Bismarck Archipelago and later colonisation of Remote Oceania by Lapita-bearing populations (Kirch 2000, Spriggs 1997, Bellwood 2007). Today, a more fragmented picture is emerging. It now seems likely that initial Taiwanese cultural influences, and potentially associated languages, only skirted the majority of ISEA before c. 3500 years ago, indeed, it is now being suggested that Lapita pottery may have originated from the Mariana Islands rather than ISEA (Carson et al. 2012).

At this time, Taiwanese material cultural influences across southern ISEA (namely, from Sumatra to western New Guinea) are negligible. For example, the oft-cited cultural

marker of Taiwanese cultural influence - red-slipped pottery (Kirch 2000, Spriggs 2003, Bellwood 2007) – has only a minimal occurrence within ISEA before 3000 years ago (Bulbeck 2008: Fig. 4). A different pottery tradition originating on mainland SEA occurred across the majority of southern ISEA, which like pigs (*Sus scrofa*, Larson et al. 2007), indicates a different cultural affiliation, namely from the Malay Peninsula rather than from Taiwan (Donohue and Denham 2010). Additionally, many of the plants cultivated within ISEA around this time are likely to have been domesticated in the New Guinea region or ISEA, they were not imported from Taiwan and there is very limited trace of East Asian farming within the region at this time (Denham 2010, 2011, 2013).

Consequently, there is an emerging asynchrony between archaeological and linguistic models for ISEA around 4000-3000 years ago. Archaeologically, the overwhelming impression is of limited Taiwanese cultural influences within ISEA before 3000 years ago, as well as subsequently (with mainland Asian influences potentially coming across the South China Sea rather than predominantly from Taiwan). Linguistically, Taiwanese influences came to dominate ISEA before c.2000 years ago, when written sources attest their widespread presence (eg., Coedès 1930, Thurgood 1999). Corresponding human genetic research does not provide a clear or consistent impression of Taiwanese signatures among populations in ISEA, or adjacent regions (reviewed in, e.g., Tumonggor et al. 2013) rather they are ‘fickle’ (Spriggs 2010) and it is unclear which time-depths are represented by different genetic markers (Denham and Donohue 2012a, b).

In sum, there is a seeming paradox. Taiwanese material cultural and human genetic signatures are variably present within ISEA from 4000 years ago to the present, yet Austronesian languages are ubiquitous (though see Donohue and Denham in press). What historical processes can account for these asynchronies?

Foremost, it is becoming clear that the spread of Austronesian languages within ISEA needs to be decoupled from the dispersal of farmer-voyagers from Taiwan, there was no demic diffusion across ISEA from Taiwan via the Philippines (Donohue and Denham 2010, in press, Carson et al. 2012). Consequently, Austronesian languages did not spread across ISEA as a result of colonisation and displacement. In turn, models that suggest phylogenetic structure to Austronesian language dispersal within ISEA, largely assuming the spread of farmer-voyagers from Taiwan, are erroneous or at best misplaced (eg, Grey et al. 2009). Historical linguistics has repeatedly shown that there is no high-level internal structure to Malayo-Polynesian languages within ISEA (e.g., Blust 2009, plus debate in Donohue and Grimes 2008). The absence of phylogenetic structure based on the comparative method (following Blust 2009, rather than method used in Gray et al. 2009) is not surprising, because the revisions in archaeological thinking regarding the dispersal of material cultural traits across ISEA, or lack thereof, suggests that the dispersal of these languages is not associated with a clear historically and geographically defined migration.

In contrast to previous authors, we do not consider the lack of phylogenetic structure to the historical dispersal of Malayo-Polynesian languages within ISEA as a hindrance. Rather it provides a valuable clue. It immediately suggests that a process other than a

‘migration producing phylogeographic structure’ occurred in the past. Consequently, alternative processes need to be invoked.

Although the timing and historical processes for the dispersal of M-P languages across ISEA are unclear for the period 4000-3000 years ago, they are clearer for the last 2500 years. Within the last 2500 years various processes can be invoked to have enabled Austronesian language dispersal – these include multiple trade routes into and across ISEA (e.g., Bellwood 2007, Bulbeck 2008, Carson et al. 2013, Munoz 2006, Shaffer 1996 and others) that eventually become formalised through the influence of Indic cultures, the spread of Islam, the formation of polities, colonial rule and independent state formation in the modern period. To summarise, Austronesian languages did not disperse 4000-3000 years ago, rather they have continued to disperse from 4000 years ago to the present (also see Lansing et al. 2011). In this regard, the spread of Malayic and Chamic languages may provide an analogue for the spread of Austronesian languages across ISEA, and also may account for the geographical structuring of Austronesian and non-Austronesian languages in Southeast Asia.

4. Building a demographic model

In section 2 we discussed the linguistics of the Malayic and Chamic subgroups, which both form a closely related set of languages and also a pair of geographically dispersed clades within Malayo-Polynesian. This study was presented as an example of what can be historically calibrated: we know that Malay spread rapidly, and that the spread was of a desirable social norm, rather than extensive movements of Malayic-speaking peoples.² Further, it is not suggested here, nor has it been suggested in histories of the Malayic-speaking peoples (Hall 1981, Munoz 2006, Ryan 1976, Shaffer 1996 and others), that all of the dispersals occurred at the same time. In many parts of what is now eastern Indonesia we have written records concerning the first use of ‘Malay’ in the area. In some cases we have written evidence about the graded spread of Malay across different regions (Hall 1981, Munoz 2006, Ricklefs 2001), while in others we can judge by the innovations in lexemes and in grammatical forms. We also have written accounts of the spread of the Chamic languages (eg., Thurgood 1999), and proof from both Malayic and Chamic that the languages dispersed at different rates in different times and places. There was no single dramatic Malayic or Chamic expansion.

What is clear is that the spread of Malayic and Chamic involved the expansion of lexical memes into new areas, and that the ‘essential character’ of the languages observed in these new areas is that of the non-Malayic (or non-Chamic) languages of those areas. The clusters formed from an examination of typological traits without respect for phonological or morphological realisation do not respect subgroup boundaries within Austronesian, nor in many cases the division between Austronesian and non-Austronesian, whether in the west on mainland Southeast Asia or in the east on and near the New Guinea mainland. As shown in

² The earlier dispersal of Malayic-speaking peoples, between Borneo and Sumatra, most likely did involve demic diffusion. This is a different time period to that being discussed here, and involved much shorter distances than those attested between the Malayic languages today.

Donohue (2013), the patterns we observe are those that reflect language shift in the face of a socially dominant intruder language – Malayic or Chamic, as the case may be.

On a greater level, in sections 2 and 3 we have shown that a non-demic and non-synchronous model for the diffusion of *Austronesian* languages across ISEA is not only plausible, but is recapitulated in the later epigraphically attested spread of two of the subgroups of Austronesian in both Island and mainland Southeast Asia. Not only is the model plausible, but it is demonstrably better at accounting for the tremendous variety of form found in Austronesian.

Donohue and Denham (in press) have shown that the number of unique identifying traits in Austronesian is far less than that found for, for example, Indo-European. In particular, there are no phonological traits that identify Austronesian languages, with a particular divide emerging between the languages of Taiwan and those found outside that home island. Parkvall (2008) confirms that the level of typological diversity in Austronesian is unprecedented amongst other language families. Other work has shown that the level of lexical conservatism in Austronesian is quite extraordinary for a family of its size, suggesting that a process of dispersal different from that found in most other families is responsible for its spread and subsequent domination of such a large, and previously inhabited, area (Wichmann to appear, Donohue and Denham 2010). In terms of human genetics, too, we find diversity, rather than affinity. Tumonggor et al. (2013: 172) note that ‘Austronesian populations are characterized more by their diversity than by any shared genetic inheritance’.

It is likely that similar processes held for the dispersal of Austronesian elsewhere. Increasingly, the academic consensus is that there was *not* a dramatic cultural dispersal from Taiwan into ISEA (Donohue and Denham 2010, Carson et al. 2013), nor was there a spread of human genetic material at a period that is synchronous with any alleged dispersal from Taiwan (Donohue and Denham 2011, Jinam et al. 2012, Denham and Donohue 2012). A recent paper (Tumonggor et al. 2013: 172) notes that ‘many aspects of culture—notably the widespread dispersal of Austronesian languages—are not obviously associated with genetics’, and that ‘the data would fit a model of rapid expansion from Taiwan to the Philippines and Indonesia, but are equally consistent with population movements in the opposite direction’. Similarly, the synchrony of the ‘Lapita expansion’ in Oceania has been shown to be a construct, rather than a single archaeological event (Specht et al. 2013), and that while subsequent dispersal into Oceania was seemingly based from the Bismarck Archipelago (Denham et al. 2012), it should not necessarily be associated with Austronesian languages (Donohue and Denham 2008, Donohue and Denham 2012).

The Austronesian languages were, prior to the period of European expansionism, the most widely dispersed linguistic phylogeny on the planet. Their origin in Taiwan is not in question, the means by which they dispersed outside Taiwan has long been driven by forced analogy with the spread of Polynesian languages into previously uninhabited islands in Remote Oceania. We have earlier argued that demic diffusion was not the process by which these languages replaced existing language ecologies in ISEA (Donohue and Denham 2010, and subsequent works, especially Donohue and Denham in press). Here we have shown that there are historical examples of Austronesian languages spreading in Southeast Asia without

large amounts of demic diffusion. We have shown that the linguistic data, when examined piece-by-piece, strongly support the hypothesis that language shift was a major driver in the spread of Austronesian languages across ISEA.

References

- Adelaar, Karl. A. 1992. *Proto-Malayic: the reconstruction of its phonology and parts of its morphology and lexicon*. Canberra: Pacific Linguistics C-119.
- Ansaldò, Umberto, Stephen Matthews, and Lisa Lim, eds. 2007. *Deconstructing Creole*. (Typological Studies in Language 73.) Amsterdam: John Benjamins.
- Bellwood, Peter. 2007. *Prehistory of the Indo-Malaysian Archipelago*. ANU E-press.
- Bellwood, Peter, James J. Fox, Darrell Tryon. 1995. *The Austronesians: Historical and Comparative Perspectives*. Canberra: The Australian National University.
- Blust, Robert A. 2009. *The Austronesian languages*. Canberra: Pacific Linguistics.
- Brugmann, Karl, und Bertold Delbrück. 1886, 1916. *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*. Strassburg: K. J. Trübner.
- Bulbeck, David. 2008. An integrated perspective on the Austronesian diaspora: the switch from cereal agriculture to maritime foraging in the colonisation of Island Southeast Asia. *Australian Archaeology* 67: 31–51.
- Campbell, Lyle & William J. Poser. 2008. *Language Classification: History and method*.
- Carson, Mike T. 2012. Evolution of an Austronesian landscape: The Ritidian site in Guam. *Journal of Austronesian Studies* 3:55–86.
- Carson, Mike T., Hsiao-chun Hung, Glenn Summerhayes, and Peter Bellwood. 2013. The Pottery Trail From Southeast Asia to Remote Oceania. *Journal of Island & Coastal Archaeology* 8: 17–36
- Claude, Jacques. 1991. The Use of Sanskrit in the Khmer and Cham Inscriptions. In *Sanskrit Outside India Vol. 7: 5-12*. Leiden: Panels of the VIIth World Sanskrit Conference. 1.
- Coedès, George. 1930. *Les inscriptions malaises de Çrivijaya*. *Bulletin de l'Ecole française d'Extrême-Orient* 30: 29–80.
- Croft, William 2000. *Explaining Language Change*. Harlow: Longman
- de Casparis, J.G. 1975. *Indonesian palaeography: a history of writing in Indonesia from the beginnings to c. A.D. 1500*. Leiden: Brill.
- De Graff, Michel. 2003. Against Creole Exceptionalism. *Language* 79 (2): 391–410.
- Denham, Tim. 2010. From Domestication Histories to Regional Prehistory: Using Plants to Re-evaluate Early and Mid-Holocene Interaction between New Guinea and Southeast Asia. *Food and History* 8 (1): 3-22.
- Denham, Tim. 2011. Early Agriculture and Plant Domestication in New Guinea and Island Southeast Asia. *Current Anthropology* 52 (Supp 4): S379-S395
- Denham, Tim. 2013. Early farming in Island Southeast Asia: an alternative hypothesis. *Antiquity* 87 (335): 250-257.
- Denham, Tim, Christopher Bronk-Ramsey, and Jim Specht. 2012. Dating the appearance of Lapita pottery in the Bismarck Archipelago and its dispersal to Remote Oceania. *Archaeology in Oceania* 47 (1): 39-46.

- Denham, Tim, and Mark Donohue. 2012a. Reconnecting genes, languages and material culture in Island Southeast Asia: aphorisms on geography and history. *Language Dynamics and Change* 2: 184-211.
- Denham, Tim, and Mark Donohue. 2012b. Lack of correspondence between Asian-Papuan genetic admixture and Austronesian dispersal in eastern Indonesia. *Proceedings of the National Academy of Sciences* 109 (39): E2577.
<http://www.pnas.org/cgi/doi/10.1073/pnas.1209616109>.
- Donohue, Mark. 2012. Typology and Areality. *Language Dynamics and Change* 2: 98-116.
- Donohue, Mark. 2013. Towards a Papuan history of languages. *Language and Linguistics in Melanesia* 31 (1): 24-41. Available online at
<http://www.langlxmelanesia.com/issues.htm>.
- Donohue, Mark. 2013. Who inherits what, when? contact, substrates and superimposition zones. In Balthasar Bickel, Lenore A. Grenoble David A. Peterson, and Alan Timberlake, eds., *Language Typology and Historical Contingency*: 219-240. *Typological Studies in Language* 104. Amsterdam: John Benjamins.
- Donohue, Mark, and Tim Denham. 2008. The Language of Lapita: Vanuatu and an early Papuan Presence in the Pacific. *Oceanic Linguistics* 47 (2): 365-376.
- 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 Tim Denham. 2012. Lapita and Proto-Oceanic. *Journal of Pacific History* 47 (4): 443-457
- Donohue, Mark, and Tim Denham. In press. Becoming Austronesian: mechanisms of language dispersal across southern Island Southeast Asia. In David Gil and John McWhorter, eds., *Austronesian Undressed*. Canberra: Pacific Linguistics.
- Donohue, Mark, Tim Denham and Stephen Oppenheimer. 2012. Uncoupling inheritance and diffusion: a lexical-based methodology detects social distance. *Diachronica* 29 (4): 502-522.
- Donohue, Mark, and Charles E. Grimes. 2008. Yet more on the position of the languages of eastern Indonesia and East Timor. *Oceanic Linguistics* 47 (1): 115-159.
- Donohue, Mark, and Simon Musgrave. 2007. Typology and the linguistic macro-history of island Melanesia. *Oceanic linguistics* 46 (2): 348-387.
- Donohue, Mark, Simon Musgrave, Bronwen Whiting and Søren Wichmann. 2011. Typological feature analysis models linguistic geography. *Language* 87 (2): 369-383.
- Donohue, Mark, Søren Wichmann and Mihai Albu. 2008. Typology, areality and diffusion. *Oceanic Linguistics* 47 (1): 223-232.
- Gil, David. 2001. Creoles, Complexity and Riau Indonesian. *Linguistic Typology* 5: 325-371.
- Gray, Russell, Alexei J. Drummond and Simon J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* 323: 479-483.
- Gong, Tao. 2010. Exploring the roles of Horizontal, Vertical and Oblique Transmissions in Language Evolution. *Adaptive Behaviour* 18 (3-4): 356-376.
- Haaksma, Rémy. 1933. *Inleiding tot de studie der vervoegde vormen in de Indonesische talen*. Leiden: Voorheen E.J. Brill.
- Hall, Daniel George Edward. 1981. *History of South East Asia*. London: Macmillan.
- Holm, John. 1988. *Pidgins and Creoles*. Cambridge: Cambridge University Press.

- Humboldt, Wilhelm von. 1836–9. *Über die Kawi-Sprache auf der Insel Java, nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues* (3 volumes). Berlin: Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin.
- Jinam, Timothy A., Lih-Chun Hong, Maude E. Phipps, Mark Stoneking, Mahmood Ameen, Juli Edo, HUGO Pan-Asian SNP Consortium, and Naruya Saitou. 2011. Evolutionary History of Continental Southeast Asians: “Early Train” Hypothesis Based on Genetic Analysis of Mitochondrial and Autosomal DNA Data. *Molecular Biology and Evolution* 29 (11): 3513–3527.
- Joseph, Brian D. and Richard D. Janda, eds. 2003. *The handbook of historical linguistics*. Oxford: Blackwell.
- Kirch, Pat. 2000. *On The Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact*. Berkeley: University of California Press
- Lansing, J. Stephen, Murray P. Cox, Therese A. de Vet, Sean S. Downey, Brian Hallmark and Herawati Sudoyo. 2011. An ongoing Austronesian expansion in Island Southeast Asia. *Journal of Anthropological Archaeology* 30: 262–272.
- Morrison, George Ernest. 1975. The Early Cham language and its relation to Malay. *Journal of the Malaysian Branch of the Royal Asiatic Society* 48: 52–59.
- Mufwene, Salikoko. 2002. *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Munoz, Paul Michel. 2006. *Early Kingdoms of the Indonesian Archipelago and the Malay Peninsula*. Singapore: Editions Didier Millet.
- Noonan, Michael. 2010. Genetic classification and Language contact. In Raymond Hickey, ed., *The Handbook of Language contact*, 48–67. Wiley Blackwell.
- Parkvall, Mikael. 2008. Which parts of language are the most stable? *SprachTypologie und Universalien Forschung*. 61: 234–250.
- Ricklefs, Merle C. 2001. *A History of Modern Indonesia since c.1200. 3rd Edition*, Stanford: Stanford University Press.
- Ryan, Neil Joseph. 1976. *A History of Malaysia and Singapore*. London: Oxford University Press.
- Shaffer, Lynda Norene. 1996. *Maritime Southeast Asia to 1500*. London: ME Sharpe Armonk.
- Sneddon, James N. 2003. *The Indonesian language: its history and role in modern society*, University of New South Wales Press.
- Specht, Jim, Tim Denham, James Goff and John Edward Terrell. 2013. Deconstructing the Lapita Cultural Complex in the Bismarck Archipelago. *Journal of Archaeological Research*.
- Spriggs, Matthew. 1997. *The Island Melanesians. The Peoples of Southeast Asia and the Pacific*. Cambridge: Blackwell.
- Spriggs, Matthew. 2003. Chronology of the Neolithic transition in Island Southeast Asia and the western Pacific: a view from 2003. *The Review of Archaeology* 24 (2): 57–80.
- Spriggs, Matthew. 2010. ‘I was so much older then, I’m younger than that now’: why the dates keep changing for the spread of Austronesian languages, in John Bowden, Nikolaus Himmelmann and Malcolm Ross, eds, *A Journey through Austronesian and Papuan linguistic and cultural space: papers in honour of Andrew Pawley* (Pacific Linguistics 615). Canberra: Pacific Linguistics.

- Stuart-Fox, Martin. 2003. *A Short History of China and Southeast Asia: Tribute, Trade, and Influence*. London: Allen and Unwin.
- Thomason, Sarah Grey, and Terrence Kaufman. 1988. *Language contact, Creolization, and Genetic Linguistics*. Los Angeles: University of California Press.
- Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change: With an Appendix of Chamic Reconstructions and Loanwords*. Oceanic Linguistics Special Publications, No. 28.
- Tumonggor, Meryanne K., Tatiana M. Karafet, Brian Hallmark, J. Stephen Lansing, Herawati Sudoyo, Michael F. Hammer and Murray P. Cox. 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *Journal of Human Genetics* 58: 165–173.
- Weinreich, Uriel. 1953. *Languages in Contact*. New York: Mouton.
- Wichmann, Søren. to appear. Neolithic linguistics. In Barjamovic, Gojko, Irene Elmerot, Adam Hyllested, Benedicte Nielsen, and Bjørn Okholm Skaarup, eds, *Language and Prehistory of the Indo-European Peoples—A Cross-Disciplinary Perspective*. Budapest: Archaeolingua.
- Wichmann, Søren, and Arpiar Saunders. 2007. How to use typological databases in historical linguistic research. *Diachronica* 24 (2): 373–404.

Appendix 1.

Nexus files used for Figures 1 and 2.

Phonology:

```
#nexus
```

```
BEGIN Taxa;
DIMENSIONS ntax=18;
TAXLABELS
[1] 'Acehnese'
[2] 'Ambonese Malay'
[3] 'Cham Eastern'
[4] 'Cham Western'
[5] 'Iban'
[6] 'Indonesian'
[7] 'Jambi Malay'
[8] 'Kerinci'
[9] 'Kupang Malay'
[10] 'Larantuka Malay'
[11] 'Manadonese'
[12] 'Melanau'
[13] 'Minangkabau'
[14] 'Mualang'
[15] 'Nonthaburi Malay'
[16] 'Papuan Malay'
[17] 'Tsat'
[18] 'Urak Lawoi'
;
END; [Taxa]

BEGIN Characters;
DIMENSIONS nchar=191;
FORMAT
    datatype=STANDARD
    missing=?
    gap=-
    symbols="0 1 2 3 4 5 6 7 8 9"
    labels=left
    transpose=no
    interleave=yes
;
MATRIX
'Acehnese'          200002000020000200000000444000000000000000100010101004400000
'Ambonese Malay'   200002000020000200000000444000000000000000100010101004400000
'Cham Eastern'     3000030000000020200000001450300200000000000100010101004400000
'Cham Western'    3000030000000030200000001450300300000000000100010101004400000
'Iban'             2000020000200002000000001454000000000000000100010101004400000
'Indonesian'       200101000020000200000000544000000000000000100010101004400000
'Jambi Malay'     200101000000020200000001554000000000000000100010101004400000
'Kerinci'          200101000000020200000001554000000000000000100010101004000000
'Kupang Malay'    200002000020000200000000444000000000000000100010101004400000
'Larantuka Malay' 200002000000200200000000444000000000000000100010101004400000
'Manadonese'      200002000020000200000000444000000000000000100010101004400001
'Melanau'         2001010000200002000000001554000000000000000100010101004400000
'Minangkabau'    2000020000200002000000001454000000000000000100010101004400000
'Mualang'         30000300000000303000000014540000400000000000100010101004400000
'Nonthaburi Malay' 3000030000000030300000001454400000000000000100010101004400000
'Papuan Malay'    200002000020000200000000444000000000000000100010101004400000
'Tsat'            2000031000000002000000001450200200000000000100010001003300001
'Urak Lawoi'      3000030000200003000000001454300000000000000100010101004400000

'Acehnese'          000100100000000013300000000001001010000000020000213310404000
'Ambonese Malay'   000100000000000?12200000000001001010000000020000212200200000
'Cham Eastern'     000100000000000012200000000010010100000000100?0313320400000
'Cham Western'    000100000000000012200000000020010100000100000?0313320450090
'Iban'             000100000000000?122000000000100101000000020000212210200000
```

```

'Indonesian'      0001000000000000?12200000000001001010000000020000212210200000
'Jambi Malay'    000100000000000012200000000001001010000000020000213300400000
'Kerinci'        0001000000000000?12200000000001001010000000020000313320400000
'Kupang Malay'   0001000000000000?12200000000001001010000000020000212200200000
'Larantuka Malay' 000100000000000012200000000001001010000000020000212210200000
'Manadonese'     0001000000000000?13300000000001001010000000020000212200000000
'Melanau'        0001000000000000?12200000000002001010000000020000212210200000
'Minangkabau'   0001000000000000?122000000000010010100000000200?0212200200000
'Mualang'        000100000010000?1321000000000001010000000020000211110000000
'Nonthaburi Malay' 000200000010000?13220000000001001010000000020000313310400000
'Papuan Malay'   0001000000000000?12200000000001001010000000020000212210000000
'Tsat'           00020001001000000441000000000001010000000020000212210210000
'Urak Lawoi'     0001000000000000?12200000000001001010000000020000313310400000

'Acehnese'       0000000000000000000000000000000000010042100??000000000111?00??0??
'Ambonese Malay' 00000000000000000000000000000000000100311001100000000000000010010
'Cham Eastern'   000090000000000110011010000000100431????????000000000000??0??
'Cham Western'   000090000000000000001000000000100421100000??000000111?0000000
'Iban'           00000000000000000000000000000000000100321011100000000000000110010
'Indonesian'     000000000000000000000000000000000102003210011100000000000000??0??
'Jambi Malay'    000000000000000000000000000000000102004210111100010100000000010010
'Kerinci'        00000000000000000000000000000000010200421001110010000000000010010
'Kupang Malay'   0000000000000000000000000000000001003110001000000000000000010010
'Larantuka Malay' 0000000000000000000000000000000001003310011000000000011101010001
'Manadonese'     00000000000000000000000000000000010031100110000000000000000?0??
'Melanau'        000000000000000000000000000000000102003110111000000000000000000000
'Minangkabau'   000000000000000000000000000000000100311??1100??000000000000?????
'Mualang'        000000000000000000000000000000000100222011100000000000000100100
'Nonthaburi Malay' 000000000000000000000000000000000100421111110000000000000000101
'Papuan Malay'   000000000000000000000000000000000100321000100??0000001???100?000
'Tsat'           000000000000050011111110000001003221111000000011100000010110
'Urak Lawoi'     000000000000000000000000000000000100411??1100??000000000000?????

'Acehnese'       ??????000000
'Ambonese Malay' 11?1000001
'Cham Eastern'   ??????000000
'Cham Western'   00???000001
'Iban'           11?1000001
'Indonesian'     ???10000001
'Jambi Malay'    11?1000001
'Kerinci'        11?1010001
'Kupang Malay'   01?1000001
'Larantuka Malay' 11?1000001
'Manadonese'     ???1000001
'Melanau'        00?1000001
'Minangkabau'   ???1000001
'Mualang'        00?10001001
'Nonthaburi Malay' 00?1000001
'Papuan Malay'   ???1000001
'Tsat'           00110101001
'Urak Lawoi'     ???11000000
;
END; [Characters]
BEGIN st_Assumptions;
  extaxa= 'Balinese' 'Madurese' 'Sasak' 'Sumbawa' 'Sundanese';
  charttransform=Uncorrected_P HandleAmbiguousStates = Ignore;
  disttransform=NeighborNet;
  splitstransform=EqualAngle;
  SplitsPostProcess filter=dimension value=4;
  autolayoutnodelabels;
END; [st_Assumptions]

```

Morphosyntax

#nexus

```

BEGIN Taxa;
DIMENSIONS ntax=18;
TAXLABELS
[1] 'Acehnese'
[2] 'Ambonese Malay'
[3] 'Cham Eastern'
[4] 'Cham Western'
[5] 'Iban'
[6] 'Indonesian'
[7] 'Jambi Malay'
[8] 'Kerinci'
[9] 'Kupang Malay'
[10] 'Larantuka Malay'
[11] 'Manadonese'
[12] 'Melanau'
[13] 'Minangkabau'
[14] 'Mualang'
[15] 'Nonthaburi Malay'
[16] 'Papuan Malay'
[17] 'Tsat'
[18] 'Urak Lawoi'
;
END; [Taxa]

BEGIN Characters;
DIMENSIONS nchar=247;
FORMAT
    datatype=STANDARD
    missing=?
    gap=-
    symbols="0 1 2 3 4 5 6 7 8 9"
    labels=left
    transpose=no
    interleave=yes
;
MATRIX
'Acehnese'          10000?210101000????0000000000001000???0000001?3100000???101?
'Ambonese Malay'   10000010001001020002010000001101000???0000001021010001011001
'Cham Eastern'     0100?????????????0?????????????????????????????0?3???0000??????
'Cham Western'     01000? ???????00??0?????????????1000???100000????????????????
'Iban'             10000?00000000?00000000000000??1000???0000001?0000000???1??
'Indonesian'       010000300000002100000000000101110000011010001020000001111001
'Jambi Malay'      11000020000001120000000000010111000?001110001021000001111001
'Kerinci'          100111?1101111121100000000010111000???1100001?210000011110?0
'Kupang Malay'     01000020000001020000000000001101010???1010001021000000111001
'Larantuka Malay'  01000010001000020000000000000001010???101000103100000???11001
'Manadonese'       10000?0001011120?00000000001101000001000000102100000???10?1
'Melanau'          1000001001000202000000000000000010000010000001121000000111001
'Minangkabau'     10000020000000020000000000010111000???0000001030000001111001
'Mualang'          110000300000002200000000000101110000010000001141000001111001
'Nonthaburi Malay' 01000000000000020000000000001111000???00000010111100001110?1
'Papuan Malay'     100010210111010201000000100011110100010001010010000000011021
'Tsat'             110000100010000200000000000000000001???000000???100000011011
'Urak Lawoi'       010000100000000200000000000000??1000???10000010210000000110?1

'Acehnese'          010000?001??000000?01??00????110001010000010???0101001?????
'Ambonese Malay'   0100001???1??0000001111000????11100?001001010010010100??????
'Cham Eastern'     ??0000????2??00?0001?????????????????0????????????010?0????00?
'Cham Western'     000000????2??00?????????????????????0?00000?????010?0????00?
'Iban'             000000?001??000000?01??00????0001010000010???01010????00?
'Indonesian'       000000100010101000001101000010110001010100010000110001000000
'Jambi Malay'      0000001???1??0100111011000????110001011000011??0110001000000
'Kerinci'          1110101????1??01000010????????????0001000001010???001001000000
'Kupang Malay'     0100001???1011100001011000????111001011100010010010001000000
'Larantuka Malay'  0100001???1????00001011000????111001010000010010010100?00000

```

```

'Manadonese' 0100001????0??01001000????????????????001010001010????010100????00?
'Melanau' 010000?0002??0100111011000????????0001001001010010010100000000
'Minangkabau' 00000001001010000000111000????010001010000010010010100?00000
'Mualang' 0000000??0010100000111000????100010?0000010000010100000000
'Nonthaburi Malay' 0100001???2??0000001011000????111000000001010?0010100???001
'Papuan Malay' 0100001???1??1000101011000????111001011001100010?0?000000000
'Tsat' 010000????1??0000111011000????00100101000001001?0101001??00?
'Urak Lawoi' 0100001???2??0000111011000??????100101000101000?010100????00?

'Acehnese' ?0000011111111000101010100010011100011100000000010010?11100
'Ambonese Malay' ?0000011010101001001110100010011????01100000000000000?00000
'Cham Eastern' ?00000100101????0101010????00??????01100000000000000?00000
'Cham Western' ?00000100101?1000101??0010010001101001100000000000000?00000
'Iban' ?0000010010101000101010100010011110011100000000000000000000
'Indonesian' 0000001001010100010101010001001110001110000000000100001000
'Jambi Malay' 0000001001010100010101110001000100001110100000000000?01000
'Kerinci' 00000010010101000101010100010011????11100010001000000?00000
'Kupang Malay' 000000100101010010011101000100111010011000000000000000000
'Larantuka Malay' 0000001001010100100101010001001110100110000000000000?00000
'Manadonese' ?0000010010101001001100100010010?0?01010000000000000?00000
'Melanau' 0000001111010100010101010001001100001010000000000000?00000
'Minangkabau' 0000001111111000101010100010001101111100000000001000001000
'Mualang' 0000001001010100010101001001000100001010000000000000?00000
'Nonthaburi Malay' 00000010010101000101010010010001?0?0111000000000000000500000
'Papuan Malay' 00000010010101001010110100010011101001100000000001000211100
'Tsat' ?000001101010100111111010011001010100110000000000000?00000
'Urak Lawoi' ?000001001010100010101011001000110100110000000000000500000

'Acehnese' 100111????10??????10010010000001000001001001?00?0001000110001
'Ambonese Malay' 000100100200000??010100100000?000101000100110000001000110001
'Cham Eastern' 000????????????????????10000????????????????1????????????????1
'Cham Western' 000?????????0000?0?01?010000?1000????????1????????????10001
'Iban' 000????00101001????011010000001000001001001?00?0001000110001
'Indonesian' 000111100101001101001101000000100000100100110000001000110001
'Jambi Malay' 000111100200000??1001101000010000100101100110001001001110001
'Kerinci' 000101????200000??010010000??00?0010010?1????100010001000?
'Kupang Malay' 000100100100000??1101101000000100101000100110000001000110001
'Larantuka Malay' 000010?????0000?01010010000001001010001011??000001000110000
'Manadonese' 000100100200000??010010000??10000100010?110000001000110001
'Melanau' 000011100201111101001001000010000110100101110001001100110001
'Minangkabau' 0001111001010011010011010000001000001001001?00?0001000110001
'Mualang' 000110100210000??1001001000010000101100101110000001000110001
'Nonthaburi Malay' 000??000200000??0101001000010?00?00100100101000001000110001
'Papuan Malay' 100100100100000??1110001000000100001000101?10000001000110001
'Tsat' 00010000000000??0?0100??????10000010010??00010001000110001
'Urak Lawoi' 0001000000?0000??010110??????10000010010010100100101001001?

'Acehnese' 0101010
'Ambonese Malay' 010101?
'Cham Eastern' ?1?1?1?
'Cham Western' ?1?1?1?
'Iban' 0101010
'Indonesian' 0101010
'Jambi Malay' 0101010
'Kerinci' ???????
'Kupang Malay' 0101010
'Larantuka Malay' 101011?
'Manadonese' 010101?
'Melanau' 010101?
'Minangkabau' 0101010
'Mualang' 010101?
'Nonthaburi Malay' 010101?
'Papuan Malay' 0101010
'Tsat' 010101?
'Urak Lawoi' 1101011
;
END; [Characters]
BEGIN st_Assumptions;
  extaxa='Balinese' 'Madurese' 'Sasak' 'Sumbawa' 'Sundanese';
  chartransform=Uncorrected_P HandleAmbiguousStates = Ignore;

```

```
disttransform=NeighborNet;  
splitstransform=EqualAngle;  
SplitsPostProcess filter=dimension value=4;  
autolayoutnodelabels;  
END; [st_Assumptions]
```


Appendix 2

Language considered

The expanded sample used for Maps 4 and 5 in section 2 consists of 155 languages. The locations of most of these languages is shown in Maps 4 and 5; some languages are not found in the confines of the maps used, and no attempt has been made to represent the (northern) Australian languages on the maps.

Austronesian (Malayo-Sumbawan): Acehnese, Ambonese Malay, Balinese, Cham Eastern, Cham Western, Iban, Indonesian, Jambi Malay, Kerinci, Kupang Malay, Larantuka Malay, Madurese, Manadonese, Melanau, Minangkabau, Mualang, Nonthaburi Malay, Papuan Malay, Sasak, Sumbawa, Sundanese, Tsat, Urak Lawoi.

Austronesian (other Malayo-Polynesian subgroups): Amarasi, Ambai, Arguni, Balantak, Banda, Begak, Biak, Bobongko, Buru, Cebuano, Dadua, Dhao, Dobel, Dohoi, Embaloh, Fordata, Galolen, Gayo, Helong, Javanese, Kambara, Karo Batak, Kei, Keo, Kesui, Koiwai, Komodo, Lampung, Maya, Muna, Nuaulu, Palue, Pendau, Proto-Malayo-Polynesian, Rote, Sawu, Selaru, Souw Amana Teru, Taba, Tagalog, Tatana, Tboli, Tengger, Tetun, Toba Batak, Tolaki, Tugun, Tukang Besi, Uab Meto, Uma, Uruangnirin, West Damar, Wolio, Wotu.

Austro-Asiatic: Chrau, Jahai, Khmer, Nyah Kur, Nyaheun, Pacoh, Semelai, Sre, U, Vietnamese. Tai-Kadai: Be, Hlai, Li, Maonan, Mulao, Thai, Zhuang.

Tibeto-Burman: Amoy, Bai, Cantonese, Fuzhou, Hangzhou, Karen, Kayah Li, Lahu, Shanghai, Tujia.

Trans New Guinea: Amele, Asmat, Dani, Dom, Ekagi, Fore, Fuyuge, Iha, Imonda, Kaugel, Kewa, Lani, Menya, Mian.

WestPapuan: Abun, MaiBrat, Mpur, Pagu, Ternate, Tidore, Tobelo.

Other Papuan families: Alamlak (Sepik), Kanum (South New Guinea), Arapesh, One (Torricelli), Skou (Skou), Mairasi (Tanah Merah).

Australian (Pama-Nyungan): Djapu, Martuthunira, Ngiyambaa, Nyawaygi, Uradhi, Wirangu, Tangkic.

Australian (not Pama-Nyungan): Kayardild, Ngarinyin, Tiwi, Yukulta.