PHILOSOPHICAL
TRANSACTIONS
—— OF ——

THE ROYAL
SOCIETY

B

BIOLOGICAL
SCIENCES

Historical linguistics in Australia: trees, networks and their implications

Claire Bowern

References	**This article cites 37 articles, 2 of which can be accessed free** http://rstb.royalsocietypublishing.org/content/365/1559/3845.full.html#ref-list-1
Rapid response	Respond to this article http://rstb.royalsocietypublishing.org/letters/submit/royptb;365/1559/3845
Subject collections	Articles on similar topics can be found in the following collections evolution (2143 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

Historical linguistics in Australia: trees, networks and their implications

Claire Bowern*

Department of Linguistics, 370 Temple Street, Yale University, New Haven, CT 06511 USA

This paper presents an overview of the current state of historical linguistics in Australian languages. Australian languages have been important in theoretical debates about the nature of language change and the possibilities for reconstruction and classification in areas of intensive diffusion. Here are summarized the most important outstanding questions for Australian linguistic prehistory; I also present a case study of the Karnic subgroup of Pama–Nyungan, which illustrates the problems for classification in Australian languages and potential approaches using phylogenetic methods.

Keywords: Australian languages; phylogenetics; historical linguistics, reconstruction

1. INTRODUCTION

In the historical linguistics literature, Australian languages stand out as unusual in more than one way. For example, the Pama–Nyungan family appears to be an exception to generalizations regarding language family size and hunter–gatherer communities (Wichmann *et al.* 2008). The family itself has been the subject of debate, although the consensus view within Australian linguistics has been, for some time, that alternative models of change (such as punctuated equilibrium; see Dixon 1997) are problematic and premature before further work has been done using the comparative method (Bowern 2006).

Elsewhere (Bowern & Koch 2004*a*; Bowern 2006), I have argued forcefully against models of language change that highlight areal diffusion at the expense of other types of change. However, it is clear that a lot remains to be said about language change in Australia, particularly with respect to areal patterns. There has been a tendency in some recent work (e.g. Dench 2001) to privilege areal spread over genetic descent, or to argue that an areal explanation for any given change is just as probable as shared descent. However, a claim that makes areal diffusion the primary mechanism of language change contradicts what we know about more general processes of language change. It is a testable claim about how languages are acquired and spread, and how changes are spread through communities and beyond. Linguistic diffusion thus correlates with social interactions; however, to my knowledge no one who has made an areal claim of this type (Dench 2001; Dixon 2001, 2002; Clendon 2006) has supported that claim with appropriately detailed sociolinguistic and anthropological data. Furthermore, many of these arguments confuse the family tree and comparative method. The comparative method can be used to reconstruct extensive diffusion as well as shared innovation through other types of descent (pace Thomason & Kaufman 1988; Labov 2007). An areal reconstruction does not represent a failure of the comparative method. Finally, there is a need for greater transparency in our assumptions which relate area, linguistic diffusion, shared innovations and the reconstruction of language history. For example, in Bowern (2008), I argue in response to Breen (2007) that subgrouping failures (or subgrouping difficulties) may themselves be indicative of certain types of change. That is, our failure to find neat subgrouping is not necessarily a failure of our methods; it is indicative of a type of language splitting produced by certain types of population prehistory.[1] Australia is thus an important region for historical reconstruction theory, especially as it relates to small populations.

There are currently four outstanding issues in the study of Australian linguistic prehistory:

— What is the subgrouping of Pama–Nyungan languages? (What does it mean for our interpretation of prehistory when we cannot draw a neat tree?)
— Where did Pama–Nyungan spread from?
— How do we account for the spread of Pama–Nyungan?
— What is the relationship between the Pama–Nyungan family and other languages in Australia?

We do not have good answers to any of these questions at present, although we have hypotheses for all of them. In this article, I discuss previous work on the above four questions and outline a programme for research in this area.

(a) *Background to the languages of Australia*

Some basic information about the languages in question is in order. At the time of European settlement in Australia in 1788, there were approximately 250 distinct languages spoken by people who lived in social units varying in size from fewer than 100 to several

*claire.bowern@yale.edu

thousand people. Aboriginal people lived in all parts of Australia, including the arid central desert regions. The languages have been grouped into approximately 28 families (O'Grady *et al.* 1966a; Wurm 1972; Wurm & Hattori 1981; Bowern & Koch 2004a). Initial classifications were completed using lexicostatistics (O'Grady *et al.* 1966b) and these classifications provided us with approximately 20 primary subgroups of the Pama–Nyungan family, along with the remaining 27 non-Pama–Nyungan families, which are clustered in the far north of the country.

Currently, more than 90 per cent of Australia's indigenous languages are endangered, 60 per cent of aboriginal people live in urban or regional centres, and fewer than 10 per cent of aboriginal people speak a traditional language. The largest languages have about 5000 speakers, and only 20 languages are being acquired by children. For those languages without detailed 20th Century records, primary data collection is therefore either extremely urgent or too late in most cases. However, there is a considerable amount of primary material from the 19th century, as well as unpublished fieldnotes (and increasingly these are being published). Although Australian speech populations are currently small, there is no reason to assume that languages would have been larger in the precolonial period; in fact, some community lingua francas, such as Dhuwal and Burarra, probably have more speakers now than they did before European settlement. Multi-lingualism was widespread in precontact times but not universal by any means. The picture from work such as Heath (1978, 1981) suggests linguistically diverse communities where speakers were all fluent in each other's languages; while that is accurate for Arnhem Land, other parts of the country show a variety of patterns, including monolingualism, asymmetrical bilingualism and multi-lectalism.

Australia is the only continent where agriculture did not develop before the colonial period. Several different subsistence methods which broadly fall under the label 'hunter–gatherer' were practised. Social organization also varied, from nomadic groups of perhaps 50 people to sedentary clan groups comprising several hundred individuals. Some groups were monolingual; in others exogamy was the norm and societywide multi-lingualism in several unrelated languages was found. (For an overview, see Hiscock (2008) and for detailed case studies, see Keen (2002, 2004).) I stress this because it is important to remember that Australia is not a homogeneous area, either geographically, socially or linguistically.

(b) *Pama–Nyungan languages and subgrouping*
While earlier work identified differences between northern and southern languages in Australia (e.g. Schmidt 1919; Kroeber 1923), the identification of the Pama–Nyungan family is due to work by Ken Hale and colleagues (Hale 1964, 1966; O'Grady *et al.* 1966a,b).[2] The work of Schmidt (1919) was the first large-scale classification attempt using all available materials (much of Schmidt's data came from the wordlists in Curr (1886)). He used a list of

44 vocabulary items, personal pronouns, interrogatives and some phonotactic and syntactic information. From this, he identified a set of languages which he called 'South Australian languages'. He also posited some intermediate-level major groupings within Southern Australian and some of these had further groups within them. Schmidt misclassifies the Pama–Nyungan languages that have undergone extensive sound change.

The next major classification is due to Capell (1941, 1956, 1979) and is based largely on typology (i.e. it is a phonetic classification based on shared structural features rather than shared innovations). His groups are partially areal and partially typological. The lexicostatistical classification of O'Grady *et al.* (1966a) used a modified form of the Swadesh word list. They aimed to cover the whole continent, and the classification project included fieldwork as well as existing materials. The process for the lexicostatistics classification was described in O'Grady & Klokeid (1969). Since then there have been a number of classifications based broadly on O'Grady *et al.*'s (1966a), including Wurm & Hattori (1981). The classification was never meant to stand as anything other than a first effort, to be refined as our knowledge of the languages grew. There has been subsequent comparative work by O'Grady and his students (Hendrie 1990; O'Grady 1990, 1998; O'Grady & Fitzgerald 1997) at the level of Pama–Nyungan.

Dixon (2002) was a new subgrouping of a rather different type. It is not a genetic subgrouping: that is, Dixon's approach is not cladistic. It is partly genetic and partly areal. He combines claimed linguistic areas and families and subgroups in the same classification.[3] The data on which certain groups are decided as areal or cladistic have not been published.

Bowern & Koch (2004a) is a collection of subgrouping studies, including Alpher (2004), which is a first principles demonstration of Pama–Nyungan as a family. The papers in this volume demonstrate the comparative method for nine subgroups of Pama–Nyungan; the remaining papers discuss wider relations among non-Pama–Nyungan languages (e.g. Baker 2004; Bowern 2004; Green & Nordlinger 2004).

There are problems in the reconstruction of Pama–Nyungan subgrouping, some of which I will mention briefly here. First is that there are not many people working in this area. There are no full-time historical linguists working on Australian languages: there are few active scholars in this area and everyone has an alternative speciality, such as typology or language documentation. While it is an active and close knit group of scholars, there is only so much that can be done, and this is relevant when we compare it with the number of people working on, for example, the history of French. Second is a data problem. There is a great deal more data for Australian languages than there was 30 years ago, but so many languages have disappeared that classification data for some areas is extremely degraded. In the Karnic example we will see below, for example, the northeastern fringe is represented entirely by 19th-century wordlist sources.

The third problem is that we do not have a very good picture of the language contact situation for

much of the country, and so we assume that it was equivalent everywhere to the best-studied cases in Arnhem land. However, from the data that I have been able to collect opportunistically, it is clear that there were multiple patterns. Not only was there widespread community-wide multi-lingualism in multiple languages as we find in current Arnhem land communities (see Bowern 2008), there were also asymmetric bilingual interactions (where one community learnt the language of its neighbours, but the neighbours would not learn the other language[4]), and there were monolingual populations. In some communities, they seem to have been key language people who knew many surrounding languages and would have acted as interpreters for their communities. Until we get a better idea of the social interaction among groups we will not have an adequate idea about the role of language contact and how big a role it should play in our theories.

(c) *The spread of Pama–Nyungan*

There are two primary competing theories regarding the origin of the current distribution of Pama–Nyungan languages. The first is that they spread in the early Holocene, probably from somewhere south of the modern Gulf of Carpentaria. Variants of this model have been proposed by Sutton (1990, 1997), McConvell & Evans (1997) and Evans & Jones (1997) and others. The location of the putative homeland is based primarily on methods in linguistic geography, such as the area of greatest diversity within Pama–Nyungan. It is also noteworthy that the current distribution of Pama–Nyungan languages and a spread from a north-eastern homeland correlates well with the distribution of backed artefacts (Hiscock 2002).

The second theory is that Pama–Nyungan is not a clade, but a remnant diffusion area created from the initial (Pleistocene era) colonization of Australia, with subsequent intense diffusion. This is essentially the view of Dixon (1997) and subsequent work; a modified view (Clendon 2006) has Pama–Nyungan as a bottleneck linguistic area which expanded from the South following climatic amelioration in the arid centre after the last glacial maximum. As appealing as these ideas might be to archaeology, they are implausible linguistically. A linguistic area maintained over such a large area for 40 000 years is highly implausible. It equires types of changes which have not been attested elsewhere, or which are rare elsewhere but would have to be exceedingly common in Australia (such as the borrowing of pronouns). I have written extensively on problems with the idea of Pama–Nyungan as a diffusion area, and will not repeat those arguments here (see Bowern 2006, 2007).

The Holocene expansion theory is itself also not without problems, however. It is unclear what the trigger for the spread of Pama–Nyungan would have been. There is no punctuation or other event in the archaeological record which we can associate with significantly better technology or superior warfare. However, as noted above there is some evidence for significant small tool expansion at the relevant time
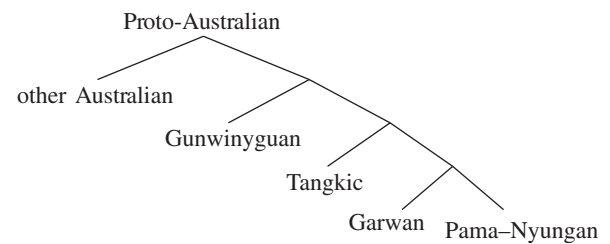

Figure 1. Putative Proto-Australian family tree (after Evans 2005).

period (see also Evans & Jones 1997). Large-scale conquests are unknown in hunter–gatherer communities for obvious demographic reasons. The archaeological record is patchy but does show habitation of the southern regions from the Pleistocene period (e.g. Devil's Lair in South Western Australia from 36 000 BP and Willandra Lakes in New South Wales; see Mulvaney & Kamminga 1999; Hiscock 2008).

A late Holocene expansion from the north would most probably have meant the acquisition by Pama–Nyungan speakers of lands belonging to non-Pama–Nyungan speakers. Why would speakers have shifted languages? It is well known that newcomers to communities tend to adopt the languages of the hosts, and not vice versa; moreover, people do not generally switch languages without good reason. Exceptions to this pattern of language shift seem to be largely confined to the colonial and post-colonial period (although see McConvell 2001; McConvell & Alpher 2002 for some situations where migration may lead to language shift towards immigrant languages). Finally, given that there is an obvious climatic change in the late Pleistocene, it is tempting to link the expansion of the family to it and assume a scenario of spread with climatic amelioration and greater access to resources driving population increase and therefore expansion. We therefore need a principled reason why this would not be appropriate.

In summary, Pama–Nyungan is obviously a challenging problem for those who maintain that the only causes of widespread expansion are either technological or expansion into uninhabited territory (e.g. Renfrew 1989; Bellwood 2001).

(d) *Relationships between Pama–Nyungan and other Australian languages*

Given that there is little work in the reconstruction of the non-Pama–Nyungan families, and that reconstruction of Pama–Nyungan itself is at an early stage, discussion of more remote relationships is premature. Evans (2005) and Evans & Jones (1997) and other work has Pama–Nyungan as one of a series of families, as schematized in figure 1. This is not the only hypothesis concerning Pama–Nyungan relationships but it is the most widely accepted currently.

This tree is based on very little evidence, however. A single shared innovation defines each of these nodes. Moreover, the status of Gunwinyguan is unclear at present; the composition of the family is disputed and is not established by the traditional methods used in linguistic reconstruction. On the basis of

shared irregularities in verb morphology, Green (2003) suggests that Gunwinyguan in fact belongs to a macro-family containing a number of languages in Arnhem Land that have so far not figured in close relationships to Pama–Nyungan. This must remain an open question for the present.

An alternative tree is presented by Heath (1990), and used implicitly by Clendon (2006). In this tree, proto-Australian has two daughters: Proto-Pama–Nyungan and Proto-Non-Pama–Nyungan. Evidence here is scanty and inconclusive, and based mostly on pronominal data.

It is tempting to compare Australian languages to those of Papua New Guinea. After all, the two countries were joined by a land bridge until the end of the Pleistocene and Australia must have been settled via New Guinea. Comparisons have so far failed to reveal any relatives.

2. CASE STUDY: THE LAKE EYRE LANGUAGES

In this case study, I present work in progress using a combination of established historical methods and computational phylogenetic analysis. Karnic languages are an excellent case study for the problems in Pama–Nyungan, simply because the same problems we see at a larger scale with 150 languages are also found in the subgroup. We find a similar set of problems with the interplay between areal and non-areal features, difficulty in distinguishing archaic shared features from shared innovations, and little evidence for higher-order structure.

(a) *Overview*
(i) *Geographical area*
The languages which form the basis of this case study are those formally spoken in the Lake Eyre Basin of Eastern Central Australia, straddling the Queensland, Northern Territory, New South Wales and South Australian borders. The area where Karnic languages are spoken broadly comprises the Lake Eyre drainage basin; mostly rather flat semi-arid country, subject to occasional seasonal inundation. A map of the area can be seen in the electronic supplementary material, figure S1.

(ii) *Prior classification*
The classification of the Lake Eyre languages has not been stable. Researchers have vacillated between recognizing a series of low-level groups with no closer higher relations between them, and grouping the languages into a larger family; the composition of this family, however, has also varied over time. This section briefly surveys the most widely known classifications.

Schmidt (1919, pp. 43–44) defines a Karna group, which refers to Pitta-Pitta, Mithaka, Kunggari, and related dialects, as part of his *Süd-Zentral-gruppe*. He also defines two *Untergruppen* 'subgroups'; the Nulla-Untergruppe (Arabana–Wangkangurru) and the Dieri-Yarrawurka-Wonkamarra-Evelyn Creek-Untergruppe, the name of which is self-explanatory. O'Grady *et al.* (1966a) also recognized Karnic, although their Karnic was considerably smaller than

the current Karnic classifications. Other publications identify several independent groups, implicationally no more closely related to each than to any other Pama–Nyungan subgroup. Breen (1971) recognized the wider relations of O'Grady, Voegelin and Voegelin's 'Karnic' group and related it to 'Mitakudic' and others in the Lake Eyre Basin. Almost all the subgroupings have been based primarily on lexico-statistical data, whether as part of a wider preliminary survey of languages (O'Grady *et al.* 1966a; Wurm 1972) or a more detailed comparison (Breen 1971). Two of the later classifications, Austin (1990a) and Bowern (2001), also take morphological and lexical reconstruction into account, but because the reconstructions of the two authors are different they come to rather different conclusions regarding subgrouping. Trees of these previous classifications are provided in the electronic supplementary material, figure S2. The number of classification claims for the subgroup makes it one of the better studied in the family. However, there is little agreement beyond the lowest level groups. A summary is given in Breen (2007) and further discussion in Bowern (2009). I see the following issues as being most important for the study of the subgroup and its theoretical historical implications:

— Are there any higher-level groupings beyond the lower-level ones identified by early lexicostatistics, and about which all classifications are in agreement?
— If so, what are they?
— How far do the borders of the family extend? Does the family include Arabana–Wangkangurru? The eastern languages Garlali and Badjiri? The northern languages Yanda and Guwa?
— Is there a northern Karnic subgroup comprising Arabana and Pitta-Pitta (and associated dialects)?
— Which subgroup does Mithaka belong to?

What are the innovations which would characterize each of these groups?[5] As I showed in Bowern (1998), there are innovations in morphology which provide conflicting evidence for Karnic subgrouping. Evidence for a Northern subgroup includes shared vocabulary, innovations in pronouns (e.g. Proto-Karnic *ngantya 1sg dative > nominative and *nhuka 3sg nominative > *uka* (Arabana), *nhuwa* (Pitta-Pitta)) and the use of one of the allomorphs of the locative case marker as a causal. However, there are also innovations that group Pitta-Pitta and associated dialects with other Karnic languages, implying that there is no common northern Karnic clade. These include lexical items, a change of *-nga locative allomorph > dative (+ more general locative > dative change) and a second person singular accusative *nyuna (although this may be shared archaism and therefore useless for subgrouping).

In other work (e.g. Bowern 2009), I have argued that the existence of contradictory subgroupings such as this may have more than one explanation. In much work in historical linguistics, it is assumed implicitly (and sometimes stated explicitly) that conflicting subgrouping is primarily (only?) owing to

language contact that has obscured clear tree branching (Thomason & Kaufman 1988; Labov 2007). However, it seems clear that there are cases where the same processes of language change that produce tree-like splits may also produce networks. In particular, a large area that was settled fairly quickly and where speech communities retained alliances with one another for some time is highly likely to produce a complex dialect area, and relics of the conflicting isoglosses in the earlier dialect area may persist after the languages lose mutual intelligibility (and cease to be regarded as dialects of one another). Arguments of this type rely on relative chronology of shared innovations. In the Karnic case, they are further complicated by subsequent diffusion.[6]

The conflicting claims for subgrouping could have several other explanations, however. It could be that the authors' reliance on different types of data has resulted in different trees because of unidentified shared retention or borrowing in different areas of grammar and lexicon. In the following section, I consider a computational phylogenetic analysis of the problem.

(b) *Method*

Information on the language sources used is given in Bowern (2001, table 1, p. 248) and is summarized in the electronic supplementary material, S3. Data for 770 mostly lexical character sets were coded as multi-state and then converted to 5487 binary characters. The data were then analysed with SPLITSTREE 4.0 (Huson & Bryant 2006) using the NeighborNet algorithm (Bryant *et al.* 2005; Huson & Bryant 2006).[7] When using the comparative method in linguistics, loanwords are normally excluded. In this coding, however, I did not treat identified loans differently because of the high likelihood of substantial undetected borrowing. Moreover, in §2c, I use variable borrowing rates to identify potential loan paths, and this would not be possible if borrowings were filtered.

Originally, 40 taxa were sampled, including dialects of the better attested languages and languages outside the Lake Eyre Basin. Some of these languages are very poorly attested and lack of data resulted in such a small number of informative characters.[8] The Thura–Yura language Adnyamathanha (Simpson & Hercus 2004) was used as an outgroup. Others were excluded because of clear data contamination.[9] The results were then compared with those obtained in previous classification studies.

It has long been known that certain types of words are more susceptible to borrowing than others. Flora, fauna and artefact terms are universally borrowed at higher rates than body part terms, for example. McConvell (2010) finds that more than 50 per cent of animal terms are loans in the Pama–Nyungan language Gurindji, for example.[10] Other semantic fields are more variable. For example, kinship terms are often treated as basic vocabulary in Indo-European studies, but in groups that practise exogamy kinship terms are often subject to borrowing; more generally, certain kinship terms are subject to replacement by

baby-talk terms and are therefore perhaps less reliable than other basic vocabulary items. See further Haspelmath (2008) and Haspelmath & Tadmor (2009).

(c) *Results and discussion*

The NNet for 25 well-attested taxa, using all data points, is given in figure 2. All the lower-level groupings identified by previous classification studies also appear in this network: Eastern Karnic (Wangkumara dialects and Punthamara), Western Karnic (Diyari, Ngamini and Yarluyandi), and Central Karnic (those languages plus Yandruwandha, Yawarrawarrka and Nhirrpi). Mithaka and Karuwali are in Central Karnic but not Western Karnic, and some splits group Arabana and Pitta-Pitta (and associated dialects) together.

Since NeighborNets are known to be somewhat sensitive to missing data, a subset of 1211 characters from 23 taxa with the best attestations was studied. The network from this dataset is given in the electronic supplementary material, S4. Missing data here are slightly over 30 per cent. The structure is consistent with figure 2 except that splits are ambiguous as to Mithaka's placement with Western Karnic or with Yandruwandha's group. There is also less network-like splitting in the Eastern Karnic groups. These varieties approach the similarity of dialects, and since it is very common to find overlapping isoglosses among mutually intelligible varieties (since at any given point changes may have differing ranges) this should be unsurprising. Since the full dataset and the well-attested data give the same groupings, missing data are unlikely to be an issue for the languages under consideration here.

We find some ambiguity in the placement of the Western Lake Eyre languages Arabana and Wangkangurru. One set of splits groups the languages with Pitta-Pitta and Wangkayutyuru; this is consistent with Hercus's (1994) 'northern Karnic' group. A second set of splits groups these languages with those immediately to the east of Lake Eyre, in particular, Western Karnic. This has not been proposed in any of the previous classifications of Karnic languages, although it has long been noted that languages on either side of Lake Eyre exhibit loanwords and grammatical borrowings, such as the use of pronouns inflected for kinship information. Since the Western Karnic languages have undergone changes that Arabana–Wangkangurru has not (Hercus 1994; Bowern 1998), these splits are likely to reflect loans.

NeighborNets allow us to schematize ambiguity in classification, but they do not by themselves allow us to pinpoint the source of the ambiguity. For example, if some varieties share extensive archaic features, this will place them closer together, even though shared archaisms are unrevealing for subgrouping. (This is true, of course, for all distance-based classification methods.) However, we can try to identify potential borrowings and shared archaisms by considering sub-domains of vocabulary. The Karnic character sets were coded for semantic domain and then further divided into borrowability hierarchies.[11] Figure 3 shows the NNet diagram for characters for flora and
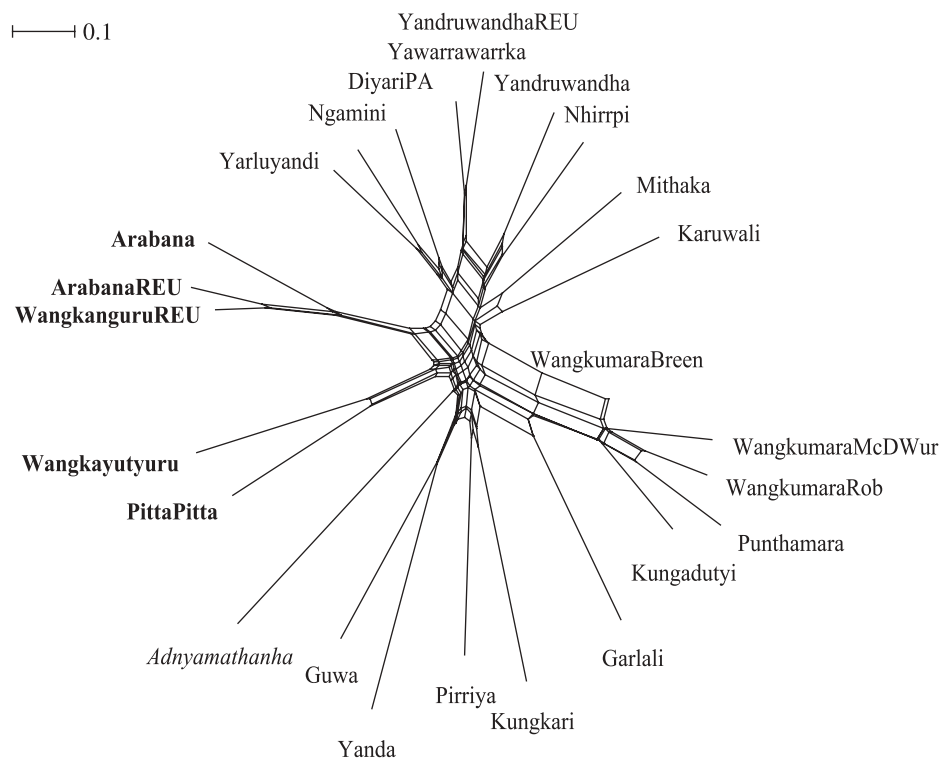
Figure 2. Karnic NeighborNet: 25 taxa, 5487 binary characters. Boldface indicates languages whose classification we are particularly considering.
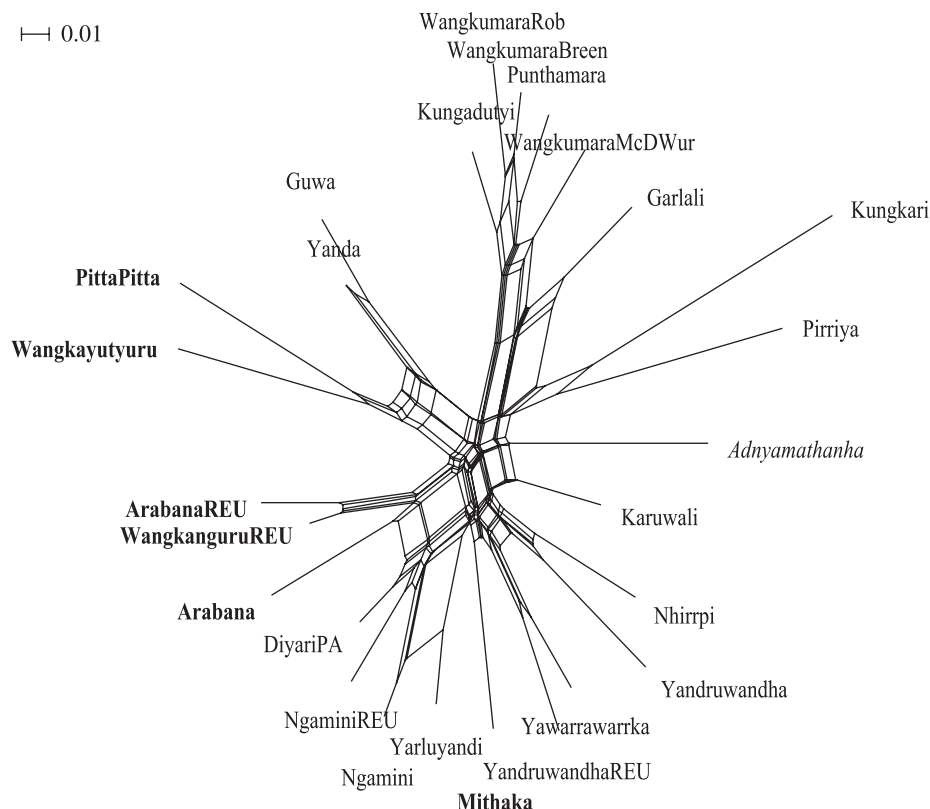

Figure 3. Karnic NNet: characters that are more susceptible to borrowing. Boldface indicates languages whose classification we are particularly considering.

fauna items and other items that are commonly borrowed. Figure 4, in contrast, shows items that are less likely to be borrowed.[12] Boldface indicates languages whose classification we are particularly considering. The high-borrowing category includes 1336 characters, while the low-borrowing category includes 3027 characters.

Let us first consider the similarities between the two networks. In both cases, a group of Eastern languages is identified, although the internal structure of the
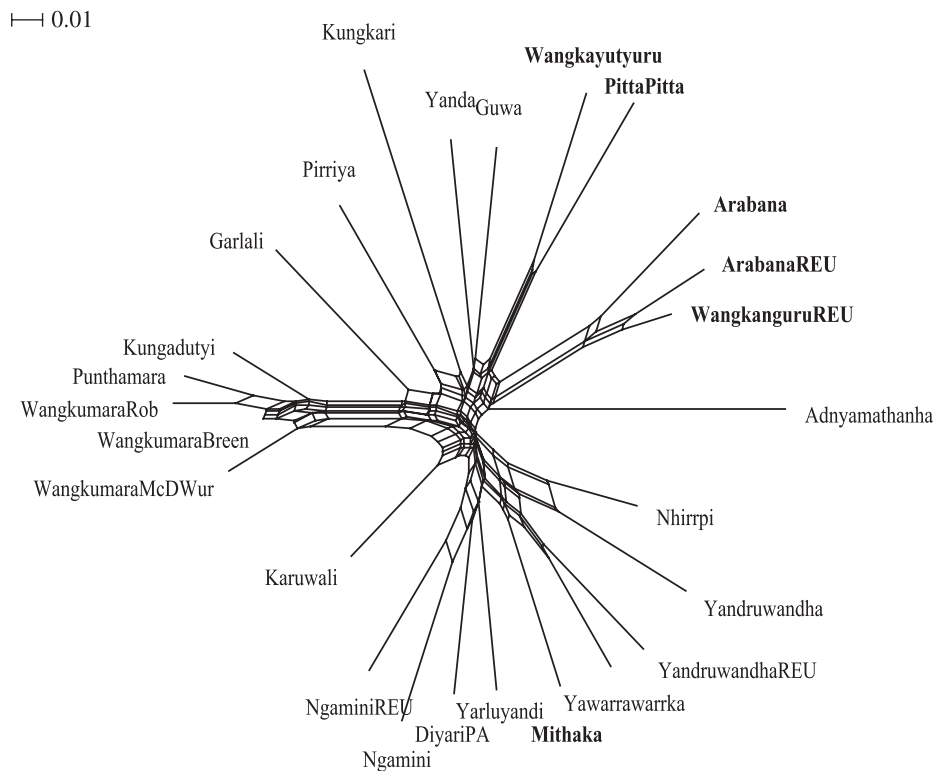
Figure 4. Karnic NNet: characters that are less susceptible to borrowing. Boldface indicates languages whose classification we are particularly considering.

group is different. We find a certain amount of network-like structure in Central Karnic in both the high-borrowing and low-borrowing datasets. This is unsurprising since speakers of the languages were in contact and a thriving trade network in grind-stone tools and other items existed in precontact times (McBryde 1987) and linked these groups closely together.

Let us now consider points of difference between the two networks. As one would expect, figure 3 is considerably messier than the low-borrowing network in figure 4. This is presumably because loans are affected by geography, with languages borrowing from more than one neighbour.[13] In the low-borrowing network, as in the aggregate data, we find ambiguous clustering. The Western Karnic languages in the high-borrowing ability network ambiguously group with the other central Karnic languages (as identified by Austin 1990*b*; Bowern 1998) and with the languages to the west of Lake Eyre, Arabana and Wangkangurru. Although in the aggregate data there were conflicting splits grouping Arabana and Wangkangurru variably with Pitta-Pitta and Wangkayutyuru and with Western Karnic (although not with Central Karnic), the signal for grouping with Pitta-Pitta and Wangkayutyuru is much weaker in the high-borrowing network, and weaker in the highly stable vocabulary. This implies that a Northern Karnic grouping is neither an artefact of loans, nor purely of shared retentions. While it does not conclusively demonstrate the existence of a Northern Karnic family, it is suggestive of such a grouping.[14] Central and Eastern Karnic are differentiated in both networks, although in the high-borrowing network Ngamini is ambiguously grouped with both Diyari and Yarluyandi.

Included in the dataset are five languages that are not usually discussed in Karnic classification. Karuwali was suggested as Karnic as far back as Breen (1971) but did not feature much in subsequent classifications that dealt mostly with grammatical data, as there are only wordlist data for the language. Aggregate data place the language as a sister to Mithaka; however, both the high- and low-borrowing networks in figures 3 and 4 have Karuwali showing conflicting splits, between the out-group Adnyamathanha and Central Karnic in the first instance, and the Eastern and Karnic groups in the second. More research is required. Two other doubtfully Karnic languages, Yanda and Guwa, group with Pitta-Pitta and Wangkayutyuru in all cases. Pirriya and Kungkari group with Eastern Karnic in the high-borrowing data, but not clearly with any particular Karnic group in the aggregate and low-borrowing data. This implies either that they are a further primary subgroup within Karnic or that they are not Karnic.

The final language that warrants discussion is Garlali. Breen (2007) argues against Bowern's (2001) classification of Garlali as Karnic. Data here are from Breen's fieldnotes; Garlali is grouped more or less strongly with Eastern Karnic in all networks. However, given the problematic history of the language description and the likelihood of borrowings in the data, it is possible that if detailed Maric subgroup data were included Garlali would not group with Eastern Karnic in the same way.

Thus in summary, there was weak evidence for a Northern Karnic group. The grouping of Central and Western Karnic is solid, although within Western Karnic the placement of Mithaka is still ambiguous. Regarding the borders of the family, we have evidence

for including Arabana–Wangkangurru as Karnic, some evidence for Garlali, good evidence for Karuwali and some evidence for the Northern languages Yanda and Guwa and the Eastern fringe languages Pirriya and Kungkari. The ambiguous grouping of languages such as Mithaka in all networks lends further support to the analysis of Karnic as being an area of long-standing dialect breakup, supporting analysis using the comparative method.

3. CONCLUSIONS AND FUTURE DIRECTIONS

I see several areas where work is urgent for Australian prehistory and reconstruction. One is the compilation of site stratigraphy in archaeology. Data from many parts of the country show a pattern of site settlement and abandonment. It would be extremely useful to have this data tabulated and plotted using GIS. To my knowledge there is no amalgamation of this data. Within linguistics there are some urgent tasks. Perhaps the most urgent is lexical reconstruction of low-level subgroups. Some areas are better studied than others but all are in need of more work.

Australia would benefit from less interdisciplinary work, particularly where linguistics is concerned. At the same time that Australia has been a leader in interdisciplinary work in linguistics, anthropology and archaeology, we have lagged in the detailed linguistic work required to build solid interdisciplinary hypotheses. For example, we cannot at present use data from flora and fauna to shed light on the Pama–Nyungan homeland because we do not have the relevant linguistic reconstructions. Linguistics has also suffered from an attempt to fit as much of the archaeological data as possible into a model, where it is not clear what is relevant. For example, the length of settlement in Australia is irrelevant to the reconstruction of Pama–Nyungan, just as the original settlement of Europe by *Homo sapiens* is irrelevant to the reconstruction of Proto-Indo-European.

In this article, I have presented a method for comparing subgrouping hypotheses by considering character sets broken down by semantic domain. This method relies on compensating for our inability to distinguish shared innovation from shared archaism by considering subnetworks. In areas where the history is not well known and relative chronology cannot be inferred because of high rates of lexical replacement, this method provides an alternative that does not rely solely on the linguists' impression of the most common groupings. This method was tested on the Karnic subgroup of Pama–Nyungan. This revealed splits that differed between frequently borrowed lexical items and infrequently borrowed ones, and both of these produced different networks from an aggregate character set. This implies both that borrowing has occurred, and that shared archaisms may be affecting subgrouping. Removing those characters produced a network in which Karnic languages fell into three primary groups: a northern group (confirming the work of Hercus 1994), a Central group (confirming Austin 1990a and Bowern 1998) and an Eastern group, which is recognized by all the relevant previous work.

ENDNOTES

[1]For example, Gray *et al.* (2009) find that the areas of least tree-like signal in Austronesian coincides with 'pauses' in migrations where groups are diverging *in situ*.

[2]This section is based on information in Koch (2004) and the reader is referred to this paper and Bowern & Koch (2004b) for more information.

[3]Note that this is not the same as the social network model used for Oceanic languages by Ross (1988): these are not linkages in the sense that Ross uses the term.

[4]This is similar to the current situation in Belgium, where native Flemish speakers usually also speak French, but native speakers of French almost never speak Flemish.

[5]From early on in historical work, linguists have avoided distance-based methods relying on similarity of forms in favour of innovation-based parsimony methods using data from both the lexicon and grammar.

[6]This model has certain elements in common with Schmidt's (1872) wave theory, although there are other aspects of that model which are inappropriate here, such as the lack of differentiation between changes between dialects and those that affect languages.

[7]For the use in linguistics of NNet and NJ techniques, see amongst others Gray *et al.* (2007), McMahon & McMahon (2006) and Bryant *et al.* (2005).

[8]In the case of the Yardli languages Wadikali, for example, less than 10 per cent of the character data could be supplied because the language is known only from a wordlist of 70 items (Hercus & Austin 2004).

[9]The sources in Reuther (1883), for example, show extensive influence from the languages to the west of Lake Eyre. Comparison with varieties recorded later (such as Austin 1981) shows that this contamination is not likely to be borrowing into that variety alone.

[10]McConvell (2010) suggests that flora/fauna borrowings might be particularly associated with territorial expansions of speakers.

[11]Note that my procedure here is not identical to that used in McMahon & McMahon (2006): they divided the Swadesh word list into degrees of borrowability: my dataset is much larger; the high-borrowing set, for example, has 1336 binary characters.

[12]I do not know of any work that quantifies the relative probability at which certain semantic classes are borrowed apart from the preliminary findings in Haspelmath (2008). The statement is based on the fact that in many parts of the world speakers borrow words from neighbouring languages when they come into contact with new items, and these new items tend to be artefacts, plants and animals, and technology, and tend not to be items such as body parts. In areas where there is a high degree of exogamy, kinship terms show instability. I assume this is because in such areas speakers have several language terms to choose from when referring to different relatives.

[13]A small supporting point is that the arrangement of taxa in the high borrowing NeighborNet is very close to the geographical distribution of the languages, with ambiguous splits often reflecting geographically adjacent languages.

[14]Remember that the grammatical reconstruction and limited innovation-based lexical data were ambiguous in this respect.

REFERENCES

Alpher, B. 2004 Pama–Nyungan: phonological reconstruction and status as a phylogenetic group. In *Australian languages: classification and the comparative method*

(eds C. Bowern & H. Koch), ch. 5, pp. 105–142. Amsterdam, The Netherlands: John Benjamins.

Austin, P. 1981 *A grammar of Diyari, South Australia*. Cambridge, UK: Cambridge University Press.

Austin, P. 1990a Classification of Lake Eyre languages. *La Trobe Univ. Working Papers Linguist.* **3**, 171–201.

Austin, P. 1990b The Karangura. *Records of the South Australian Museum*, vol. 25, pp. 129–137.

Baker, B. 2004 Stem forms and paradigm reshaping in Gunwinyguan. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 13, pp. 343–373. Amsterdam, The Netherlands: John Benjamins.

Bellwood, P. 2001 Archaeology and the historical determinants of punctuation in language-family origins. In (eds A. Aikhenvald & R. M. W. Dixon), *Areal diffusion and genetic inheritance: problems in comparative linguistics*, ch. 2, pp. 27–42. Oxford, UK: Oxford University Press.

Bowern, C. 1998 The case of Proto-Karnic. BA Honours thesis, Australian National University, Canberra, Australia.

Bowern, C. 2001 Karnic classification revisited. In *Forty years on: Ken Hale and Australian languages* (eds J. Simpson, D. Nash, M. Laughren, P. Austin & B. Alpher), ch. 17, pp. 245–261. Canberra, Australia: Pacific Linguistics.

Bowern, C. 2004 Diagnostic similarities and differences between Nyulnyulan and neighbouring languages. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 11, pp. 295–318. Amsterdam, The Netherlands: John Benjamins.

Bowern, C. 2006 Another look at Australia as a linguistic area. In *Linguistic areas* (eds Y. Matras, A. McMahon & N. Vincent), pp. 244–265. Basingstoke, UK: Palgrave Macmillan.

Bowern, C. 2007 Review of 'areal diffusion and genetic inheritance: problems in comparative linguistics'. *Anthropol. Linguist.* **49**, 429–434.

Bowern, C. 2008 *Linguistic fieldwork: a practical guide*. Basingstoke, UK: Palgrave Macmillan.

Bowern, C. 2009 Reassessing Karnic: a reply to Breen 2007. *Aust. J. Linguistics* **29**, 337–348. (doi:10.1080/07268600903232733)

Bowern, C. & Koch, H. (eds) 2004a *Australian languages: classification and the comparative method*. Amsterdam, The Netherlands: John Benjamins.

Bowern, C. & Koch, H. 2004b Introduction: subgrouping methodology in historical linguistics. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 1, pp. 1–16. Amsterdam, The Netherlands: John Benjamins.

Breen, G. 1971 Aboriginal languages of Western Queensland. *Linguist. Commun.* **5**, 1–88.

Breen, G. 2007 Reassessing Karnic. *Austral. J. Linguist.* **27**, 175–199. (doi:10.1080/07268600701522780)

Bryant, D., Filimon, F. & Gray, R. 2005 Untangling our past: languages, trees, splits and networks. In *The evolution of cultural diversity: phylogenetic approaches* (eds R. Mace & C. Holden & S. Shennan), pp. 69–85. London, UK: UCL Press.

Capell, A. 1941 The structure of Australian languages. In *Studies in Australian linguistics*, vol. 3 (ed A. P. Elkin). The Oceania Monographs, pp. 46–80. Sydney, Australia: AIAS.

Capell, A. 1956 *A new approach to Australian linguistics* (*Handbook of Australian languages part I*), Oceania Linguistic Monographs, vol. 1. Sydney, Australia: University of Sydney.

Capell, A. 1979 The history of Australian languages: a first approach. In *Australian linguistic studies*, vol. C54 (ed. Stephan A. Wurm), pp. 419–619. Canberra, Australia: Pacific Linguistics.

Clendon, M. 2006 Reassessing Australia's linguistic prehistory. *Curr. Anthropol.* **47**, 39–61. (doi:10.1086/497671)

Curr, E. M. (ed.) 1886 *The Australian race*, vol. 4. Melbourne, Australia: Government Printers.

Dench, A. 2001 Descent and diffusion: the complexity of the Pilbara situation. In *Areal diffusion and genetic inheritance: problems in comparative linguistics* (eds A. Aikhenvald & R. M. W. Dixon), ch. 5, pp. 105–133. Oxford, UK: Oxford University Press.

Dixon, R. M. W. 1997 *The rise and fall of languages*. Cambridge, UK: Cambridge University Press.

Dixon, R. M. W. 2001 The Australian linguistic area. In *Areal diffusion and genetic inheritance: problems in comparative linguistics* (eds A. Aikhenvald & R. M. W. Dixon), ch. 4, pp. 64–104. Oxford, UK: Oxford University Press.

Dixon, R. M. W. 2002 *Australian languages: their nature and development*. Cambridge, UK: Cambridge University Press.

Evans, N. 2005 Australian languages reconsidered: a review of Dixon (2002). *Oceanic Linguist.* **44**, 242–286.

Evans, N. & Jones, R. 1997 The cradle of the Pama–Nyungans: archaeological and linguistic speculations. In *Archaeology and linguistics* (eds P. McConvell & N. Evans), ch. 22, pp. 385–417. Melbourne, Australia: Melbourne University Press.

Gray, R. D., Greenhill, S. J. & Ross, R. M. 2007 The pleasures and perils of Darwinizing culture (with phylogenies). *Biol. Theory* **2**, 360–375. (doi:10.1162/biot.2007.2.4.360)

Gray, R. D., Drummond, A. J. & Greenhill, S. J. 2009 Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science* **323**, 479–483. (doi:10.1126/science.1166858)

Green, R. 2003 Proto Maningrida within Proto Arnhem: evidence from verbal inflectional suffixes. In *The Non-Pama–Nyungan languages of Northern Australia: comparative studies of the continent's most linguistically complex region*, 369. Canberra, Australia: Pacific Linguistics.

Green, I. & Nordlinger, R. 2004 Revisiting Proto-Mirndi. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 12, pp. 319–342. Amsterdam, The Netherlands: John Benjamins.

Hale, K. L. 1964 Classification of the Northern Paman languages, Cape York Peninsula, Australia: a research report. *Oceanic Linguist.* **3**, 248–265. (doi:10.2307/3622881)

Hale, K. L. 1966 The Paman group of the Pama–Nyungan family. Languages of the World: Indo-Pacific fascicle 6. *Anthropol. Linguist.* **8**, 162–197.

Haspelmath, M. 2008 Loanword typology: steps toward a systematic cross-linguistic study of lexical borrowability. *Empirical Approaches to Language Typology* **35**, 43.

Haspelmath, M. & Tadmor, U. 2009 *Loanwords in the world's languages: a comparative handbook*. Berlin, Germany: Mouton de Gruyter.

Heath, J. 1978 *Linguistic diffusion in Arnhem Land*. Canberra, Australia: Australian Institute of Aboriginal Studies.

Heath, J. 1981 A case of intensive lexical diffusion: Arnhem Land, Australia. *Language* **57**, 335–367.

Heath, J. 1990 Verbal inflection and macro-subgroupings of Australian languages: the search for conjugation markers in non-Pama–Nyungan. In *Linguistic change and*

reconstruction methodology, vol. 45 (ed. P. Baldi). Trends in Linguistics: Studies and Monographs, pp. 403–417. Berlin, Germany: Mouton.

Hendrie, T. R. 1990 Initial apicals in Nuclear Pama–Nyungan. In (eds N. G. O'Grady & T. D. Tryon) *Studies in comparative Pama–Nyungan*, vol. C-111 1990, pp. 15–77. Canberra, Australia: Pacific Linguistics.

Hercus, L. A. 1994 *A grammar of the Arabana–Wangkangurru Language. Lake Eyre Basin, South Australia*, vol. C-128. Canberra, Australia: Pacific Linguistics.

Hercus, L. & Austin, P. 2004 The Yarli languages. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 9, pp. 227–244. Amsterdam, The Netherlands: John Benjamins.

Hiscock, P. 2002 Pattern and context in the holocene proliferation of backed artifacts in australia. *Archeol. Papers Am. Anthropol. Assoc.* **12**, 163–177. (doi:10.1525/ap3a.2002.12.1.163)

Hiscock, P. 2008 *Archaeology of ancient Australia*. London, UK: Routledge.

Huson, D. H. & Bryant, D. 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267. (doi:10.1093/molbev/msj030)

Keen, I. 2002 Seven aboriginal marriage systems and their correlates. *Anthropol. Forum* **12**, 145–157. (doi:10.1080/0066467023206222770)

Keen, I. 2004 *Aboriginal economy and society: Australia at the threshold of colonisation*. Oxford, UK: Oxford University Press.

Koch, H. 2004 A methodological history of Australian linguistic classification. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 2, ppp. 17–66. Amsterdam, The Netherlands: John Benjamins.

Kroeber, A. L. 1923 Relationship of the Australian languages. *J. R. Soc. New South Wales* **57**, 101–117.

Labov, W. 2007 Transmission and diffusion. *Language* **83**, 344–387. (doi:10.1353/lan.2007.0082)

McBryde, I. 1987 Goods from another country: exchange networks in the Lake Eyre Basin. In *Australians to 1788*, Australians: a historical library vol. III, pp. 252–273. Sydney, Australia: Fairfax, Syme and Weldon Associates.

McConvell, P. 2001 Language shift and language spread among hunter–gatherers. In *Hunter–gatherers: social and biological perspectives* (eds C. Panter-Brick, P. Rowley-Conwy & R. Layton), pp. 143–169. Cambridge, UK: Cambridge University Press.

McConvell, P. 2010 Loanwords in Gurindji. *Loanword typology*. The Hague, The Netherlands: Mouton.

McConvell, P. & Alpher, B. 2002 On the Omaha trail in Australia: tracking skewing from east to west. *Anthropol. Forum* **12**, 159–175. (doi:10.1080/0066467023206222789)

McConvell, P. & Evans, N. 1997 *Archaeology and linguistics*. Melbourne, Australia: Melbourne University Press.

McMahon, A. & McMahon, R. 2006 *Language classification by numbers*. Oxford, UK: Oxford University Press.

Mulvaney, J. & Kamminga, J. 1999 *The prehistory of Australia*. Sydney, Australia: Allen and Unwin.

O'Grady, G. N. 1990 Umpila and Wadjuk: a long-slot approach to Pama–Nyungan. In (eds N. G. O'Grady &

T. D. Tryon) *Studies in comparative Pama–Nyungan*, vol. C-111 1990, pp. 1–17. Canberra, Australia: Pacific Linguistics.

O'Grady, G. N. 1998 Toward a proto-Pama–Nyungan stem list, part I: Sets j1-j25. *Oceanic Linguist.* **37**, 209–233. (doi:10.2307/3623409)

O'Grady, G. N. & Fitzgerald, S. 1997 Cognate search in the Pama–Nyungan language family. In *Archaeology and linguistics* (eds P. McConvell & N. Evans), ch. 19, pp. 341–355. Melbourne, Australia: Melbourne University Press.

O'Grady, G. N. & Klokeid, T. 1969 Australian linguistic classification: a plea for coordination of effort. *Oceania* **39**, 298–311.

O'Grady, G., Voegelin, C. & Voegelin, F. 1966a Languages of the world: Indo-Pacific fascicle 6. *Anthropol. Linguist.* **6**, 1–106.

O'Grady, G. N., Stephen, A. W. & Kenneth, L. H. 1966b *Map of Aboriginal languages of Australia*. Victoria, Canada: Department of Linguistics, University of Victoria.

Renfrew, C. 1989 Models of change in language and archaeology. *Trans. Philol. Soc.* **87**, 103–155. (doi:10.1111/j.1467-968X.1989.tb00622.x)

Reuther, J. G. 1883 Three central Australian grammars (diari, jandruwanta, wonkanguru) (eds L. A. Hercus & J. G. Breen). Manuscript translated by T. Schwarzschild, L.A. Hercus.

Ross, M. D. 1988 *Proto-Oceanic and the Austronesian languages of Western Melanesia*, vol. C, 98. Canberra, Australia: Pacific Linguistics.

Schmidt, J. 1872 *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar, Germany: Böhlau.

Schmidt, W. 1919 *Die Gliederung der Australischen Sprachen*. Vienna, Austria: Mechenaristen-Buchdruckerei.

Simpson, J. & Hercus, L. 2004 Thura-Yura as a subgroup. In *Australian languages: classification and the comparative method*, vol. 249 (eds C. Bowern & H. Koch). Current Issues in Linguistic Theory, ch. 8, pp. 197–226. Amsterdam, The Netherlands: John Benjamins.

Sutton, P. 1990 The pulsating heart: large-scale cultural and demographic processes in Aboriginal Australia. In *Hunter–gatherer demography*, vol. 39 (eds B. Meehan & N. White). Oceania Monographs, pp. 71–80. Sydney, Australia: University of Sydney.

Sutton, P 1997 Materialism, sacred myth and pluralism: competing theories of the origin of Australian languages. In *Scholar and sceptic: Australian Aboriginal studies in Honour of L.R. Hiatt*, pp. 211–242. Canberra, Australia: Aboriginal Studies Press.

Thomason, S. & Kaufman, T. 1988 *Language contact, creolization and genetic linguistics*. Berkeley, CA: University of California Press.

Wichmann, S., Stauffer, D., Schulze, C. & Holman, E. W. 2008 Do language change rates depend on population size? *Adv. Complex Systems* **11**, 357–369. (doi:10.1142/S0219525908001684)

Wurm, S. A. 1972 *Languages of Australia and Tasmania*. The Hague, The Netherlands: Mouton de Gruyter.

Wurm, S. A. & Hattori, S. (eds) 1981 *Language Atlas of the Pacific area part I: New Guinea Area, Oceania, Australia*. Canberra, Australia: Australian Academy of the Humanities (maps 20–23 compiled by M. Walsh).