

This article was downloaded by: [MPI Max-Planck-Institute Evolutionaere Anthropologie]

On: 22 May 2011

Access details: Access Details: [subscription number 931315694]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Australian Journal of Linguistics

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713404403>

A New Approach to Australian Lexicostatistics

Gavan Breen^a

^a Institute for Aboriginal Development,

Online publication date: 29 April 2011

To cite this Article Breen, Gavan(2011) 'A New Approach to Australian Lexicostatistics', Australian Journal of Linguistics, 31: 2, 233 — 268

To link to this Article: DOI: 10.1080/07268602.2011.560830

URL: <http://dx.doi.org/10.1080/07268602.2011.560830>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A New Approach to Australian Lexicostatistics*

GAVAN BREEN

Institute for Aboriginal Development

It seems that Australian languages can borrow nouns more readily than verbs. If this is so, you would expect that if the speakers of a language divide into two groups which lose contact with one another, they will start borrowing from their new neighbours so that their percentages of common vocabulary will decrease. However, since they borrow nouns more readily, the percentage of nouns they have in common decreases faster than the percentage of verbs, so that the percentage of verbs remains always higher than the percentage of nouns. Conversely, since they are borrowing from their new neighbours, the vocabulary they have in common with them will increase, but it will increase faster for nouns than for verbs. Therefore, the percentage of common nouns will remain always higher than the percentage of common verbs. Using some well-established genetically related groups of languages and some of their well-established genetically unrelated or distantly related neighbours I will show that the figures conform to this hypothesis (notwithstanding that languages may really have much more complex histories than I am assuming). I then apply this method to some less clear-cut situations. The results in some cases point to a classification different from that generally accepted.

Keywords: Australian Languages; Pama-Nyungan; Comparative Linguistics; Lexicostatistics; Maric; Karnic; Arandic; Wati; Mayi

1. Introduction

Breen (1990: ch. 7) points out that while Dixon (1980: 254) seems to be correct in claiming that there is no ‘basic vocabulary’ which is less susceptible to change than

*It will be obvious to Australianists that this title plays on the title of Arthur Capell’s famous 1956 book. I am grateful to Belinda Keillor for drawing the map and compiling the Warumungu and Warlpiri comparative vocabularies (during a period of voluntary work with me in July 2010), and to Jane Simpson for sending me her Warumungu vocabulary. Comments from two referees have led to significant improvements in the paper. Remaining defects are gremlins’ doing but, alas, my responsibility.

less 'basic' vocabulary,¹ certain classes of words are more likely to undergo change than others. Counts on numerous pairs of languages in western Queensland and neighbouring areas seemed to show that verbs were less likely than nouns to undergo loss and replacement. (The situation in languages in which verbs are a small closed class had not, and has not, been investigated.) Furthermore, some semantic classes of nouns were more stable than others; one not considered by Breen (1990) but perhaps the one most impervious to borrowing (as suggested also by Barry Blake, p.c.) is the one I call 'qualities'—words translated by English adjectives.² All of my counts here are for nouns versus verbs (I previously used overall totals versus verbs).

Before I explain the method I remind the reader that the original O'Grady *et al.* (1966) classification used a 100-word list and classified lects according to how many words were shared: lects sharing more than 70 were dialects of a single language, 51–70 languages within a subgroup, 31–50 languages in different subgroups of a group, 16–30 languages in different groups of a family, 15 or less languages in different families. (These figures are for contiguous languages or dialects; if they are separated the figures would be lower.) Although I do not use overall cognate counts, I do think of the upper and lower figures especially, 70% and 15%, as still having the same significance.

The theory behind the method is simple (and simplified). Consider first a language which is divided by the emigration of a large part of its population of speakers. The migrating group acquires new neighbours and starts interchanging vocabulary with them. The languages of the two separating communities, initially almost identical, start to grow apart as one of them (if not both) acquires vocabulary from new sources. But since nouns (on the whole—not necessarily every semantic field) are lost and replaced more readily than verbs, the percentage of nouns that the two have in common decreases more rapidly than the percentage of verbs that they have in common. When the noun percentage reaches 50, the verb percentage may be, say, 70. (The overall percentage in such a case would be about 54, but this does not have much significance.³) Such a relationship between the two categories indicates that they are genetically closely-related languages that are moving apart.

Now thinking of the migrating group and its new neighbours, we assume that the relationships are distant. The migrating group will borrow from and lend to its neighbours so that their vocabularies will converge. Nouns will be borrowed more readily than verbs, and so the percentage of nouns in common will be higher than the percentage of verbs in common. So if the noun percentage reaches 50 and the verb percentage is only 20, we have a strong indication that these are genetically unrelated or more-distantly related languages that have come to this situation from a low base.

¹ A reviewer questions the correctness of Dixon's claim, mentioning 'Common Australian' (Capell 1956) forms, and these certainly are special; perhaps we can call them 'very basic vocabulary'. I refer to these again in Sections 3(d) and 4. They are too few to form the basis for counting. Some evidence bearing on the correctness of Dixon's claim is discussed in Section 4.

² Most adjectives in most Australian languages belong to the same morphological class as nouns. See, for example, Blake (1987: 3, 8).

³ Dixon (1972: 330–337, 1980: 254–255) has argued that such a figure would have no significance even in the O'Grady *et al.* system; see also Alpher and Nash (1999).

Irrespective of the noun–verb ratio, very high cognate figures (70% or more) would always indicate close relationship, while very low figures (15% and less) indicate distant relationship but not necessarily absence of a genetic link [see Section 2(b)]. When the figures are very high or very low the statistical significance of the noun–verb ratio is almost always very low, which means that the probability that the ratio could have that value by chance is high. The primary purpose of the method is to determine whether there is a genetic relationship or not, and, unless the cognate percentages are very high or very low, the noun–verb ratio rather than the actual percentages is used for this. Having determined that there is a genetic relationship, the actual percentages can give us some idea of how close this is, and perhaps some insight into the history of the relationship between the languages.

The most striking example to date of the use of this method of comparison dates from before Breen (1990) had appeared in print. At a workshop held at the 1989 Australian Linguistic Society Annual Conference at Monash University, Rebecca Green launched her excellent paper demonstrating that three small language families of western Arnhem Land postulated by O'Grady *et al.* (1966: 30–31) were actually a single family by showing that the four languages concerned had low percentages of non-verbal vocabulary in common, but strikingly higher percentages of verbal vocabulary.⁴ A substantially revised version of this paper was eventually published as Green (2003). My own use of the method in Breen (1990) was not convincing because of the meagreness and poor quality of the data available. I used the method for discussing the relationships of Garrwa and Wanyi to their neighbours in my own paper from the 1989 workshop (see Breen 2003: 430). The fact that Yalarnnga and Kalkutungu share 41% of vocabulary overall but only 25% for verbs formed part of the demonstration that these two languages cannot be grouped at any level below Pama-Nyungan or even Proto-Australian (Breen & Blake 2007, ch. 5).

Several years earlier, Dixon, in his grammar of Wargamay (1981, p. 4) says:

Giramay and Wargamay may have around 48% common vocabulary, squarely within the 'equilibrium figures' predicted for languages that have been in contiguity for a substantial period (Dixon 1972: 331–7, 1980: 254–60); a comparison of verb forms shows only 32% being completely or almost completely identical (differing only as regards vowel length, etc.) suggesting that the languages may not be closely genetically related.

[See also Dixon (1997: 20, 27).]⁵

⁴ It will be interesting to see how applicable this method would be in some other Top End languages like those referred to by Breen (1997: 91) and those discussed by Black (2006) in which borrowing seems to be almost non-existent. The list of words given in Harvey (1991) [which does not appear in the published version (Harvey 1997)] suggests that it might be applicable for Gunwinjguan, however. (I am grateful to a reviewer for informing me of the Black reference.)

⁵ A reviewer refers to Black (1997), suggesting that this proposes an approach 'grossly similar' to mine. While this is an excellent paper, the approach differs from mine in that it refers to basic versus less basic vocabulary (as the reviewer notes), it deals with overall cognate percentages while I deal primarily with the ratio of percentages for verbs to those for nouns, and it seeks to eliminate borrowings from the count, whereas I do not.

I don't propose this as a replacement for any other method except ordinary lexicostatistics, which, however, is still the only method for any language for which we have no information other than a small vocabulary dominated by nouns. Lexicostatistics can be modified or discounted in other ways too; for example, little or no weight can be attached to comparisons of vocabularies of neighbouring languages if a large proportion are fauna terms (see Breen 1990: 160–161, 2009: 246). However, a highly significant result with this method can lead to conclusions that what might have been thought of as innovations arising in a parent language are actually innovations arising in one language and borrowed by unrelated or distantly related neighbours.

The application of this method will be demonstrated in more detail below, using four more or less well-established groups of genetically related languages—Mari (or Maric) (Barrett 2005), Ngarna (Breen 2004a), Wati (O'Grady *et al.* 1966) and Mayi (Breen 1981b). I assess the data according to whether the languages or dialects concerned are contiguous or not, and whether they are accepted as being genetically related at some level lower than Pama-Nyungan (assuming its reality) or not. If two languages are contiguous it does not follow that the figures I quote are for contiguous dialects; for example, the unnamed language of which Pitta-Pitta is a dialect is contiguous with Kungkari (I think), but the dialect Pitta-Pitta is separated from Kungkari by Mayawarli, which is perhaps the most deviant of all the dialects of that language. I give figures in the form xx–yy, where xx is the percentage of cognate nouns in the sample, and yy is the percentage of cognate verbs.⁶ I use the term 'cognate' here for pairs of words that look as if they are related, irrespective of whether this is through descent or through borrowing.⁷ I am arbitrarily counting as nouns those (rare) items for which one language uses a noun and the other uses a verb. Nouns, in these languages, include most words translated into English as adjectives. Personal pronouns comprise a nominal list of eight (all the nominative pronouns in a minimal set except third person singular) but this is expanded if there are two words in a slot—either if a pronoun has separate nominative and ergative forms, or there is an inclusive/exclusive distinction. There are also one demonstrative adverb and four interrogative words in the list. I do not include these grammatical words in my counts but may refer to them. This differs from my earlier use of this method (Breen 1990, 2003, 2007, Breen & Blake 2007), where I compared overall (including pronouns as well as verbs) with verb figures. The difference is rather small when the noun/overall figure is high but larger when it is low.

At least half of the wordlists I have used are from my own fieldwork, being all I could ascertain for the language concerned from a list of about 250 words. Major semantic groups of nouns in my 250-word lists are body parts (46), fauna (34), inanimate nature (19), flora (18), people (16), descriptive words (16), time/place/

⁶ I generally give numbers to two significant figures, or three if the first is 1 or 2.

⁷ Most counts have been done more than once, and I have tightened my criteria for cognacy over the years. Although there are certainly doubtful cases where I might make one decision one day and another the next day, there are hardly any cases where I have declined to make a decision at all.

direction (15), culture (14). There are 46 verbs (or maybe one or two more, depending on the language). Most of the words in the list could be described as basic, in the Australian context; a few of them (such as 'black duck' and 'wood duck') I included only because they appear in the vocabularies published in Curr (1886–1887). In general I have available about 200 or more of these words for the languages I am using, but for a handful of them it is lower, down to about 130. Other major sources are wordlists compiled by the late Ken Hale, and wordlists published in Menning and Nash (1981). The Appendix gives details of all sources. The number of items that could be compared for any pair of languages would, of course, be equal to or (usually) somewhat lower than the number of words available for the lesser-known member of the pair.

2. Genetically Related Languages, and Distantly Related or Unrelated Neighbours

Languages that are genetically related—languages that have resulted from division of a parent language in the past, and dialects of a language—have been gradually losing common vocabulary, by borrowing from other languages or in other ways. If the verb inventory has indeed been changing more slowly than the nominal (lexical) inventory it would follow that the percentage of verbs cognate would be higher than that of nouns cognate. If the languages from which they borrow words are not closely related this imbalance would continue. If they are related the borrowing would perhaps tend to reverse the effect of the faster loss of nominals, and nominal and verb figures would become more nearly equal. One would expect that figures, in general, would be higher with contiguous languages or dialects, and with more closely-related languages.

Data are tabulated, with the following information given:

- names of languages/dialects compared;
- details related to geographical separation and, if well-known and not stated in the text, linguistic relationship;
- percentage of nouns cognate and percentage of verbs cognate, separated by a hyphen;
- ratio of noun percentage to verb percentage;
- number of noun pairs and verb pairs counted, separated by a comma;
- chi square figure, calculated, with no corrections applied, on the applet in <http://math.hws.edu/javamath/ryan/ChiSquare.html>;
- an estimate of the statistical significance of the calculated figure.

In most cases the statistical significance is low, but it is clear that sets of figures can have much higher significance than single pairs; for example, in (a) below where, in counts of seven pairs of languages/dialects that are known to be closely related, the noun cognate percentage is substantially lower than the verb cognate percentage (as shown by the ratio in the second column) in six cases, with rather low significance individually, and in the seventh case, in which both percentages are high and so there

is little prospect of obtaining a high significance, the noun percentage is slightly higher than the verb and the significance is far lower.

(a) *Mari Languages*

Relevant examples with Mari languages, with notes on classification based on Barrett's (2005) analysis as well as my own knowledge, are:

Bidyara/Gunggari, contiguous and dialects of the same language:

%N%V	%N/%V	nN, nV	chi ²	significance ⁸
67–79	0.85	(122, 40)	1.91	low, about 80%

The low statistical significance does not matter; no unrelated languages would share such a high proportion of verbs. They are dialects of a language.

Bidyara/Gunya, contiguous and closely-related languages,

53–65	0.82	(161, 42)	2.82	moderate, about 90%
-------	------	-----------	------	---------------------

The same comment applies, but the lower percentage figures suggest that these are either dialects of a language or very closely-related languages.

Bidyara/Margany, not contiguous, closely-related languages,

47–57	0.82	(145, 44)	1.03	very low
-------	------	-----------	------	----------

Bidyara/Biri, not contiguous, closely-related languages,

37–48	0.77	(100, 32)	1.45	low, about 75%
-------	------	-----------	------	----------------

Bidyara/Gugu Badhun, distant, less closely related,

26–39	0.67	(135, 37)	2.34	moderate, < 90%
-------	------	-----------	------	-----------------

Margany/Gugu Badhun, distant, less closely related,

24–36	0.67	(145, 44)	2.31	moderate, < 90%
-------	------	-----------	------	-----------------

These results are all as expected according to the theory: the languages/dialects are all related and are diverging from one another, so the percentage of nouns cognate is

⁸ The abbreviations on this line are %N percentage of cognate nouns; %V percentage of cognate verbs; %N/%V the ratio of these two percentages; nN and nV the number of nouns and verbs, respectively, in the sample; chi² the chi-square figure; significance is the statistical significance implied by the chi-squared figure, which is inversely related to the figure: a high figure means that the significance is high, i.e. there is a low probability of the figure arising by chance from the null situation, and a low figure means that there is a high probability of the figure arising by chance.

lower than the percentage of verbs cognate. The actual percentages decrease as the geographical distance between the languages increases, as one would expect. The ratio of noun percentage to verb percentage also decreases with increasing distance, and the statistical significance of this figure increases. The individual pairs of figures are not highly significant statistically, but of course when we have a group of six (as here) all going the same way it becomes more significant. (We do not, however, have a situation where the separate figures can be multiplied together to give an overall probability, since there is some interdependence between them.)

One unexpected (but hardly surprising) result is for Margany/Gunya, contiguous and regarded as dialects of the same language. The figures are:

%N%V	%N/%V	nN, nV	chi ²	significance
75–70	1.07	(175, 45)	0.0032	extremely low

suggesting (but very weakly) that they might not be as closely related as thought. The main thing that must be said about these figures is that they are highly insignificant statistically. I therefore did a supplementary count, using all the words that were available in the vocabularies for both of them and were not in the original 250-word lists. This amounted to 152 nouns and 71 verbs, and the percentages were almost identical to those from the first count: 76–70. This is a situation where the two probabilities could be multiplied together (since the two samples are completely independent, having no words in common), but nevertheless the significance is still extremely low and it could very well be a chance result. Grammar suggests that they are dialects of the same language; it was possible to combine them in a single description—Breen (1981a)—although it proved advisable to separate the two descriptions of verb morphology. The major verbal inflections are similar, however. The previous analysis is maintained [see also Section 3(c)].

Most of the apparently unrelated neighbours of Mari languages are very poorly attested, while some neighbours, such as Guwa, seem likely to be more distant relatives of Mari [see Section 3(c)]. The following counts compare Mari languages with languages that seem clearly unrelated.

Margany/Wangkumara (contiguous)

%N%V	%N/%V	nN, nV	chi ²	significance
15–13	1.15	(174, 46)	0.138	extremely low

Margany/Kalali (contiguous)

26–15	1.73	(147, 39)	1.98	low, > 80%
-------	------	-----------	------	------------

Margany/Badjidi (contiguous)

37–9	4.1	(141, 35)	10.5	about 99.9%
------	-----	-----------	------	-------------

Margany/Paakantyi (not contiguous; a Darling River language)

7.4–1.2	6.2	(156, 41)	9.4	about 99.8%
---------	-----	-----------	-----	-------------

Bidyara/Dharumbal (not contiguous but separated by a number of Mari dialects)

16.7–16.1	1.04	(126, 28)	0.0059	extremely low
-----------	------	-----------	--------	---------------

Gugu-Badhun/Wargamay (not contiguous, but separated only by Warrungu, a dialect of the same language as Gugu-Badhun)

31–23	1.35	(136, 37)	0.81	very low
-------	------	-----------	------	----------

These figures are all in the expected direction but most of them are statistically quite insignificant or only moderately significant. In those cases where both percentage figures are in the teens or lower and the languages are contiguous we can conclude that there is only a distant relationship, if any, and there has been no significant borrowing. However, the Margany/Badjidi figures are highly significant; there is only one chance in a thousand that such a preponderance of noun figures over verb figures could occur if the actual cognate percentages for nouns and for verbs in the two languages as a whole were equal. This looks like a clear case of unrelated languages that have been in contact and borrowing for a long while. The Gugu-Badhun/Wargamay figure (at least) could indicate a genetic relationship and a long period of separation followed by a return to near neighbourhood with borrowing, mostly of nouns. This, suggested by the fairly high verb figure, will be proposed for some other pairs of languages, with rather more justification, below. The Margany/Kalali figures are enigmatic.

(b) Ngarna Languages

Examples with Ngarna languages are:

Warluwarra/Bularnu, dialects of contiguous languages but separated by one or two intermediate dialects,

%N%V	%N/%V	nN, nV	chi ²	significance
40–63	0.63	(164, 47)	7.7	about 99.5%

Bularnu/Wakaya, contiguous languages, not as closely related as the previous pair, nothing known about dialects within Wakaya. There was another language, Yinjilanji, intermediate between Wakaya and the pair Bularnu/Warluwarra but perhaps not geographically between them.

27–56	0.48	(165, 45)	13.4	>99.95%
-------	------	-----------	------	---------

Warluwarra/Wakaya, non-contiguous languages, separated by Bularnu,

15–52	0.29	(170, 48)	29.6	> 99.99%
-------	------	-----------	------	----------

Wakaya/Yanyuwa, non-contiguous but related languages separated by unrelated languages,

16–20	0.80	(158, 38)	0.34	very low
-------	------	-----------	------	----------

Bularnu/Yanyuwa, non-contiguous but related languages separated by unrelated languages,

8.4–20	0.42	(154, 44)	5.0	97.5%
--------	------	-----------	-----	-------

Warluwarra/Yanyuwa, non-contiguous but related languages separated by both related and unrelated languages,

11–19	0.58	(159, 37)	1.89	low, about 80%
-------	------	-----------	------	----------------

Some of these figures are startling, comparable with the figures Green (2003: 369) gives for the Maningrida area languages, with verb figure more than twice the noun figure. Most of the figures are highly significant; two of the figures involving Yanyuwa are not significant, but all the Yanyuwa figures show that genetically related languages, if geographically separated, can have very low percentages of common vocabulary but with verb figures clearly higher than noun figures.

Figures for Ngarna languages with unrelated neighbours include:

Yanyuwa/Garrwa (status unclear; see next paragraph), contiguous

%N%V	%N/%V	nN, nV	chi ²	significance
28–8.8	3.1	(110, 40)	6.2	> 98%

Yanyuwa/Wanyi (closely related to Garrwa), not contiguous

19–2.6	6.3	(109, 39)	5.9	about 98%
--------	-----	-----------	-----	-----------

Wakaya/Wanyi, contiguous

11.7–8.8	1.33	(115, 40)	0.27	extremely low
----------	------	-----------	------	---------------

Wakaya/Warumungu (regarded as an isolate within Pama-Nyungan), contiguous

11.5–5.9	2.0	(135, 34)	0.93	very low
----------	-----	-----------	------	----------

Wakaya/Kaytetye (regarded as Arandic), contiguous

7.6–6.5	1.17	(144, 31)	0.12	extremely low
---------	------	-----------	------	---------------

Warluwarra/Wangka-Yutjuru (Karnic), contiguous

6.8–0	∞	(176, 47)	5.8	about 98%
-------	---	-----------	-----	-----------

Warluwarra/Yalarnnga [classified as an isolate within Pama-Nyungan by Breen and Blake (2007)], contiguous

5.0–6.25	0.80	(160, 48)	0.115	extremely low
----------	------	-----------	-------	---------------

Warluwarra/Kalkutungu (as for Yalarnnga)

1.5–4.4	0.34	(166, 45)	1.46	low, <80%
---------	------	-----------	------	-----------

Warluwarra/Antekerrepenh (Arandic), contiguous

5.2–6.7	0.78	(172, 45)	0.140	extremely low
---------	------	-----------	-------	---------------

I regard the Yanyuwa/Garrwa pair as an excellent example of two unrelated languages that have been contiguous for a long time; a fairly large number of nouns have been borrowed, but not many verbs. All figures involving Garrwa or Wanyi point to their belonging to a different family from the Ngarna languages. Blake (1988, see especially 23–25) was unable to decide whether these languages were Pama-Nyungan or not (see also Blake 1990). Harvey (2009) argues that they are Pama-Nyungan. Breen (2003: 430) noted significant similarities between them and West Barkly (Ngurlun) languages Wambaya and Gudanji, but for the lexical comparisons the verb figures are very close to the overall figures and do not provide clear information on a possible genetic relationship. There may have been some big tribal movements in this area (see Breen 2003: 430–431).

All other figures in this group also point to Ngarna being a separate family. Some have a verb figure higher than the noun figure, but this has no significance when the figures are so low. The most striking pair of figures here is for Warluwarra and Wangka-Yutjuru: of 176 nouns, 13 were judged cognate (but two of these were counted as halves), and 10 of the 13 pairs were fauna names, which are the category most likely to be borrowed across unrelated-language boundaries and so be areal. Of the 47 pairs of items in the verb count not one pair was cognate. The main informant for both of these languages was the same person.

The Ngarna languages were originally classified as a group (Wakayic) within Pama-Nyungan plus, some distance away with non-Pama-Nyungan languages separating them, the Yanyuwa language, a family-level isolate within non-Pama-Nyungan. They share few features with their neighbours, and I had surmised that they

were non-Pama-Nyungan when I observed their apparent relationship (Breen 1971a: 278, 1971b: 41). However, Blake (1988) showed that the pronouns and case markers of these languages (his Warluwaric) are of what he demonstrated to be a 'Pama-Nyungan type'. I accepted his argument then; however, I would now classify these languages as the Ngarna family.⁹

(c) *Western Desert Languages*

A few counts were made for Western Desert dialects/languages and neighbours, using wordlists from vocabularies in Menning and Nash (1981).

Yankunytjatjara/Pitjantjatjara, contiguous,

%N%V	%N/%V	nN, nV	chi ²	significance
71–81	0.88	(114, 32)	1.43	low, about 80%

Yankunytjatjara/Kukatja, not contiguous

37–74	0.50	(111, 25)	11.4	>99.9%
-------	------	-----------	------	--------

Yankunytjatjara/Wangkangurru (Karnic), not contiguous but separated only by the Western Desert dialect Antikirinya

9.9–7.1	1.39	(101, 28)	0.198	extremely low
---------	------	-----------	-------	---------------

Yankunytjatjara/Walmatjarri [Ngumpin-Yapa (see McConvell & Laughren 2004)], not contiguous

11.1–18.3	0.61	(113, 30)	1.14	low, about 75%
-----------	------	-----------	------	----------------

Kukatja/Walmatjarri, contiguous or close

26–20	1.30	(110, 27)	0.36	very low
-------	------	-----------	------	----------

The highly significant result for Yankunytjatjara with Kukatja confirms the well-known fact that they are part of a large group of genetically closely-related languages. The figures for Yankunytjatjara with Walmatjarri, similar to the figures for Yanyuwa with Southern Ngarna languages, indicates that they are genetically related but have been separated for a long time. The figures for Kukatja with Walmatjarri are consistent with their being related in the same way as Yankunytjatjara and Walmatjarri are (as they must be), but having come into proximity, perhaps after a long separation, so that they have for a time been able to borrow from one another,

⁹ I have not made any comparisons with the West Barkly languages to the north of Wakaya, or with Marra and Alawa west of Yanyuwa, but these have always been regarded as different at the family level.

and the noun figure has become higher than the verb figure. The Yankunytjatjara/Wangkangurru figures suggest that these two languages do not belong to the same family.

(d) *Mayi Languages*

The classification of these languages is not so well established because of the meagre data available on them. In only two dialects, Ngawun and Mayi-Thakurti, both belonging to the eastern language, do we have a large enough number of verbs to apply this lexicostatistical method. The Mayi languages comprise two languages, Mayi-Kutuna to the west and a roughly C-shaped dialect chain, Ngawun, Mayi-Kulan, Mayi-Yapi, Mayi-Thakurti and Wunumara, forming an eastern language. Mayi-Yapi is closest, both geographically and lexically, to the western language. Ngawun is quite close geographically to Mayi-Thakurti and Wunumara, but they do not have a great deal of vocabulary in common [see Breen (1981: 14) for cognate counts].

The following figures are for Ngawun vs Mayi-Thakurti, and for each of these with western neighbours Kalkutungu and Yalarnga. A comparison of Mayi-Thakurti with the southern neighbour Guwa is also given; the number of comparable verbs is quite small, but it hardly matters in this case.

Ngawun/Mayi-Thakurti (not contiguous, part of a dialect chain)

%N%V	%N/%V	nN, nV	chi ²	significance
32–69	0.46	(113, 27)	11.1	> 99.9%

Ngawun/Kalkutungu (separated by another Mayi dialect)

8.0–4.8	1.67	(131, 31)	0.37	very low
---------	------	-----------	------	----------

Mayi-Thakurti/Kalkutungu (contiguous)

12.5–3.3	3.8	(140, 30)	2.14	low, about 85%
----------	-----	-----------	------	----------------

Ngawun/Yalarnga (Yalarnga possibly contiguous with or close to a Mayi dialect)

4.0–0	∞	(123, 33)	1.39	low, < 80%
-------	---	-----------	------	------------

Mayi-Thakurti/Yalarnga (not contiguous)

6.3–0	∞	(127, 29)	1.93	low, > 80%
-------	---	-----------	------	------------

Mayi-Thakurti/Guwa (separated by the Mayi dialect Wunumara)

15.0–0	∞	(110, 16)	2.76	moderate, about 90%
--------	---	-----------	------	---------------------

Ngawun and Mayi-Thakurti share overall (including pronouns) about 42% of vocabulary, which in itself does not suggest a close relationship. However, the finding (Breen 1981b) that they are links in a dialect chain is strongly supported by the substantially higher figure for verbs than for nouns in common. The higher figure for nouns than verbs for the other pairs simply backs up the obvious fact that they have only a few percent of vocabulary in common and so would seem to belong to different families. As one would expect, there has been some borrowing between the contiguous pair, Mayi-Thakurti and Kalkutungu, mainly in the fauna and material culture fields. Mayi-Thakurti and Guwa also have a number of fauna names and some other nouns in common.

(e) Summary

The figures assembled above show with a very high statistical significance that languages that are closely genetically related (as determined by traditional comparative methods) share a higher proportion of verbs than they do of nouns. Languages that are clearly not closely or not at all related, and that have had the chance to borrow from one another, share a higher proportion of nouns than of verbs. Only when the figures are very high or very low is this not clear, and these are the situations where, in the nature of the statistical methodology, the significance of the results is necessarily low. However, even though there may not be a reliable noun/verb ratio, very high cognate percentages always indicate that the languages are closely related and very low percentages that they are unrelated. For example, for the Margany/Gunya counts discussed above, the very high figures for both nouns and verbs show that they are closely related and the fact that the noun percentage is higher than the verb percentage is a matter of chance.

Another problem with very low numbers is that a significant proportion of the words that are common to the two languages are Capell's (1956) 'Common Australian' vocabulary or other very widespread words (about which more will be said in the Discussion).

The fact that all counts of pairs of Mari languages (with one statistically highly insignificant exception), all pairs of Ngarna languages, the two pairs of Wati languages (and certainly many more could be added) and the two Mayi dialects give results with the same orientation is striking. It is not permissible to multiply together the odds calculated for the various Mari counts to get a probability of them all pointing the same way, because there is a certain amount of interdependence between them. They all draw from more or less the same 'vocabulary pool', to use Steve Johnson's (1990: 423) term; the most distantly related pair share less than 30% overall. It is permissible, however, to multiply together average figures for different groups, and with the four groups that I have discussed here this would give us almost astronomical odds against all of these figures being the result of chance.

3. Some Case Studies

I present now some brief reports on studies using this modified lexicostatistic method, including some in which the results are inconclusive.

(a) *Yambina*

Yambina is a Mari dialect which Breen (2009), following Terrill (1998: 77), assumed to belong to the language called (after its best-known dialect) Biri. It is known only from a vocabulary published in Curr (1886–1887). Terrill gives cognate counts with other supposed Biri dialects, some of them also known only from Curr vocabularies, which normally contain very few verbs, to support her decision. I made counts with other Curr lists and got figures of 62% and 63% with Biri dialects. However, I counted 67% cognate with Yagalingu, a Bidyara dialect (i.e. belonging to a neighbouring and closely-related language). I disregarded this figure because there were only 29 words in the Yagalingu dialect list (which was not a Curr list), compared with 100 or more for the others.

However, the contributors of the Yambina list, Sir Samuel Wilson and T. Murray (1887), sent both the standard list Curr had circulated to his correspondents (Wilson), and a supplementary list (Murray) which contained 41 verbs.¹⁰ I therefore did nominal and verb counts with vocabularies of some of the best-attested dialects of four languages of this group: Biri, Bidyara, Margany and Gugu-Badhun. Yambina is contiguous with dialects of the first two of these languages, and presumably is itself a dialect of one of them, but the last two are more remote geographically. The vocabularies were taken from Beale (1974), Breen (1973, 1981a) and Sutton (1973), respectively. The cognate figures were as follows:

Yambina/Biri				
%N%V	%N/%V	nN, nV	chi ²	significance
71–52	1.30	(94, 27)	3.36	moderate, about 92%
Yambina/Bidyara				
36–46	0.77	(118, 32)	1.28	low, <80%
Yambina/Margany				
24–34	0.68	(117, 32)	1.72	low, about 80%

¹⁰ Breen (2009) said 45, but two were nominals glossed as verbs and two others were adverb + verb phrases. Since a number of these verbs are not in the 250-word list I included in the count all items in the Yambina list for which correspondents in the other lects could be found.

Yambina/Gugu-Badhun

30.3–30.0 1.01 (109, 30) 0.072 extremely low

Although the highest percentages cognate were obtained for the Biri count, the fact that the verb figure is substantially lower than the noun figure suggests that the relationship between Yambina and Biri is not as close as a total figure would suggest. On the other hand, the figures for the Bidyara count suggest that the relationship between Yambina and Bidyara is closer than a total figure would suggest. (Overall cognate percentages are 66% for Yambina/Biri and 39% for Yambina/Bidyara.) Yambina is contiguous with a Bidyara dialect (Yagalingu) as well as with Biri dialects, and I suggest now that it may be a Bidyara dialect that has switched ‘allegiance’ to the Biri group of dialects and has been borrowing (mostly nouns) from them. The significance of the figures is not high, but is encouraging.

(b) Kalkutungu and Yalarnnga

Blake (1979: 128) considered whether Kalkutungu was really Pama-Nyungan or not, and concluded that it was, but only because of its relationship with Yalarnnga, and Yalarnnga was because of its relationship with Yanda—a very tenuous link, since Yanda is known only from one amateurish wordlist of about 100 words. However, Blake (1988) defined a set of pronouns characteristic of Pama-Nyungan and showed that Kalkutungu pronouns conformed to this set. Since then its membership of Pama-Nyungan (given its existence) has not been challenged.

Breen and Blake (2007: ch. 5) conclude that, while these two languages may be more closely related to one another than either is to any other language, they cannot be shown to have a common parent. A recent count, based on the 250-word lists, gave figures of 32–19, which is a fairly substantial difference, but the chi-squared calculation gives 2.61, which implies about one chance in 10 that the result arises by chance. I therefore did a second count using all vocabulary items attested for both languages and not in the 250-word list. This count gave figures of 49–18. With a chi-squared figure of 12.3, there is less than one chance in a thousand that this is a chance result. This suggests that these are unrelated languages that have been together for a very long time. However, the verb figures are higher than we get in most cases for languages that seem to be clearly unrelated or very distantly related, and I cannot exclude the possibility of a distant relationship.

Kalkutungu/Yalarnnga

%N%V	%N/%V	nN, nV	chi ²	significance
32–19	1.68	(154, 39)	2.61	moderate, about 90%
49–18	2.72	(120, 42)	12.3	about 99.95%

The difference between these results will be discussed further in Section 4.

(c) Possible Mari Outliers

Some of these languages are part of what Dixon (2002: xxxiii) called the Greater Maric Group, being (geographically) outliers of his Maric proper subgroup but not relatable to it. Barrett (2005: 169) thinks Guwa seems to be the closest relative to Maric. He does not regard Kungkari and Pirriya as related to Maric, despite their sharing a considerable amount of vocabulary.

My figures for comparison of Guwa to Mari languages, all involving small numbers of verbs, are as follows:

Guwa/Gugu Badhun

%N%V	%N/%V	nN, nV	chi ²	significance
18–46	0.39	(108, 14)	6.2	nearly 99%

Guwa/Biri

19–31	0.61	(106, 16)	1.19	low, about 75%
-------	------	-----------	------	----------------

Guwa/Bidyara

22–31	0.71	(116, 18)	0.54	very low
-------	------	-----------	------	----------

Guwa/Margany

25–39	0.64	(120, 18)	1.54	low, about 80%
-------	------	-----------	------	----------------

Guwa/Gunya

25–41	0.61	(123, 17)	1.93	low, about 85%
-------	------	-----------	------	----------------

Only one of these figures is highly significant, but I hesitate to attach as much significance to it as it would seem to deserve, since it relies on only seven pair of verbs. However, the consistency of the figures, with the verb percentage always higher than the noun percentage, does suggest a genetic relationship between Guwa and Mari.

Possible relationships between Mari and languages to the north have hardly been touched in my counts, and will not be discussed here.

West of the Mari area lie Kungkari and Pirriya and related dialects, described on the basis of scanty data by Breen (1990). Figures published then (1990: 160) indicated that their vocabularies shared 49% overall (not just nouns), and 43% of verbs, suggesting (weakly) that they were not genetically related. A new count gives 44% for nouns and 50% for verbs, suggesting, just as weakly, that they are. Statistically, the figures are quite insignificant, and furthermore there are only 16 pairs of verbs to compare. I will not say more about Pirriya. Kungkari is a little better known.

Breen (1990: 162) obtained figures of 28% (overall) and 37% (verbs),¹¹ for Kungkari and its Mari neighbour Margany, suggesting that Kungkari was rather closely related to Mari. I have repeated this count and done others with more distant Mari languages. The number of verb pairs that could be counted were, again, rather low, ranging from 25 to 28. The figures I have obtained are as follows:

Kungkari/Margany

%N%V	%N/%V	nN, nV	chi ²	significance
24–31	0.77	(122, 27)	0.62	very low

Kungkari/Gunya

23–25	0.92	(110, 28)	0.0097	extremely low
-------	------	-----------	--------	---------------

Kungkari/Gugu-Badhun

19–24	0.80	(96, 25)	0.34	very low
-------	------	----------	------	----------

The degree of apparent relationship (as measured by cognate percentages) decreases gradually from the contiguous Margany, through the next dialect, Gunya, and to the very distant Gugu-Badhun. The statistical significance of the figures is again low. An additional problem with these figures, and others involving similarly poorly attested languages, is that the verb percentages, with verbs numbering less than 30, are likely to be much influenced by a few very widespread verbs which may be related only at a proto-Australian level, such as *nyina* 'sit', *thana* 'stand', *nguna* 'lie', *patha* 'bite'. This detracts from the significance of the rather high verb percentages. Also, Kungkari personal pronouns and interrogatives, as Barrett (2005: 170–171) points out, are very different from those of the Mari languages.

Some figures comparing Kungkari with Kalkutungu and Yalarnnga and hinting at a possible relationship also appear in Breen (1990: 160). Revised figures are:

Kungkari/Kalkutungu (not contiguous)

%N%V	%N/%V	nN, nV	chi ²	significance
9.1–15.4	0.59	(127, 26)	0.95	very low

Kungkari/Yalarnnga (not contiguous)

12.1–17.3	0.70	(107, 26)	0.49	extremely low
-----------	------	-----------	------	---------------

¹¹ The figures of 34% (overall) and 52% for verbs given in Breen (2007) seem to be a mistake. The figures given in Breen (1990) are equivalent to 26% for nouns and 37% for verbs. The recent recount (see below) gives figures closer to the 1990 ones.

Figures for Kalkutungu are particularly untrustworthy because of the sound changes that it has undergone. For example, Kalkutungu *itja* 'bite' could be related to Kungkari (and widespread) *patja* but might also be related to Yalarnga *tatja* (and I would not want to relate it to both); *ari* 'eat' seems to be related to the geographically more remote Kungkari *kari* but could be cognate with Yalarnga *ngarri*. Other Kungkari/Kalkutungu pairs regarded as cognate include *nyina/ini* 'sit', *thana/naa* 'stand' and *nguna/nuu* 'lie', but not *nhaka/naa*, *nanyi* 'see' (because of the different points of articulation of the initial nasals).

Given the substantial distances separating these pairs of languages, we cannot exclude the possibility that they are related. If it is the case, then, that Kalkutungu and Yalarnga are related to Kungkari (and, of course, to one another), and Kungkari is related to Mari, it would follow that Kalkutungu and Yalarnga should be related to Mari languages. I therefore did counts for each of them with Bidyara. The results were as follows:

Bidyara /Kalkutungu

%N%V	%N/%V	nN, nV	chi ²	significance
1.2–10.7	0.11	(170, 42)	10.3	> 99.8%

[correspondences, with Bidyara first, were 'faeces' *guna/unu*, 'breast' *ngamun/mimi* ~ *kanthangamangama* (counted as a half; the latter glossed as '? nipple'; note also *ngamarna* one of several words given for 'chest' in Kalkutungu), 'vegetable food' *mandha* ~ *ngamany* ~ *mayi/maa* (counted as a half), 'stand' *dhana/naa* ~ *narnpinarnpi* (counted as a half), 'lie' *wuna/nuu*, 'bite' *badha/itja*, 'die' *wula/ulhi* (doubtful, but note that Kalkutungu has *ngalhi* 'we two' corresponding to *ngali* in numerous languages), 'chop' *bandju/intji*].

Bidyara/Yalarnga

%N%V	%N/%V	nN, nV	chi ²	significance
2.9–5.5	0.53	(154, 45)	0.71	very low

[correspondences 'tooth' *yira/yiya* ~ *yirrali*, 'breast' *ngamun/ngama* ~ *mimi* (counted as a half), 'faeces' *guna/kuna* ~ *wuna*, 'foot' *dhina/thina*, 'vegetable food' *mandha* ~ *ngamany* ~ *mayi/manhi* ~ *rantharru* ~ *mantha* (counted as a half), 'rotten' *gadja* ~ *gabura/kutja* (counted as a half), 'stand' *dhana/thana*, 'lie' *wudyayi* ~ *wuna/nguna* (counted as a half), 'die' *wula* / *wula*].

The verb figures are higher than the noun; nevertheless these figures are all so low that they suggest (notwithstanding that the language areas are a long way apart) that neither Kalkutungu nor Yalarnga is even in the same family as the Mari languages; overall cognate percentages are 3.1% for Bidyara/Kalkutungu and 3.5% for Bidyara/Yalarnga. Also, some of the Bidyara/Kalkutungu correspondences are quite dubious and the high verb cognate figure there can hardly be believed.

(d) *Karnic*

Breen (2007) recognized two groups of related languages as possible branches of a larger group called Karnic. These were Central Karnic (as defined by Austin 1990a) and what we could call North-West Karnic, comprising the northern languages represented by Pitta-Pitta and Wangka-Yutjurru and their dialects and the western language whose best-known dialects are Arabana and Wangkangurru. (This is to be distinguished from Austin's Western Karnic, which comprises the western and northern languages of Central Karnic.)

Some figures for Central Karnic languages are:

Ngamini/Yaluyandi, not contiguous but seem to be part of a dialect chain

%N%V	%N/%V	nN, nV	chi ²	significance
82–73	1.12	(103, 32)	0.99	very low

Ngamini/Diyari, contiguous

72–72	1.00	(162, 43)	0	extremely low
-------	------	-----------	---	---------------

Diyari/Yaluyandi, not contiguous

78–77	1.01	(105, 31)	0.012	extremely low
-------	------	-----------	-------	---------------

Diyari/Yandruwandha, contiguous languages, perhaps separated by a dialect of Yandruwandha

43–51	0.84	(164, 43)	1.32	low, about 75%
-------	------	-----------	------	----------------

Ngamini/Yandruwandha, contiguous languages

40–38	1.05	(176, 43)	0.041	extremely low
-------	------	-----------	-------	---------------

Yaluyandi/Mithaka, contiguous, both poorly attested

58–68	0.85	(90, 22)	0.72	very low
-------	------	----------	------	----------

Ngamini/Mithaka, contiguous languages (or possibly part of a dialect chain) separated by Yaluyandi

42–52	0.81	(135, 29)	0.95	very low
-------	------	-----------	------	----------

Diyari/Mithaka, contiguous languages (or possibly part of a dialect chain) separated by two or more related dialects

37–50	0.74	(129, 29)	1.725	low, about 80%
-------	------	-----------	-------	----------------

Yawarrawarrka/Mithaka, contiguous languages

51.5–59 0.87 (126, 28) 0.50 very low

Yandruwandha/Mithaka, contiguous languages separated by Yawarrawarrka (dialect of Yandruwandha)

43–46.5 0.92 (143, 29) 0.100 extremely low

Yawarrawarrka/Yandruwandha, contiguous dialects

79.6–80.2 0.99 (142, 43) 0.0088 extremely low

These figures are all of quite low to very low significance statistically. However, they are not uninteresting. The figures for Yaluyandi, Ngamini and Diyari with one another seem to suggest that they are not dialects recently deriving from a mother-language, but former dialects that have separated and diverged for a period and have then come back together and begun borrowing from one another. Comparison of the pronouns [Table 1 (see Breen 2011: 137–138)] and of the morphology [Tables 2 and 3, repeated from Breen (2007)] support this suggestion as far as Diyari is concerned; there are significant differences.¹² However, these grammatical comparisons do

Table 1 Central Karnic pronouns

	Diyari	Ngamini	Yaluyandi	Mithaka	Karru ^a	Yawa	Yandru
1sgNom	nganhi	nganyi	nganyi	nganyi	nganyi	nganyi	nganyi
1sgErg	ngathu	ngathi	ngathi	ngathu	ngathu	ngathu	ngathu
2sgNom	yini	yini	yini	yini	yini	yini	yini
2sgErg	yundru	yindi	yindi	yundu	yindu	yundru	yundru
1du.in	ngaldra	ngali	ngali	ngali		ngaldra	ngaldra
1du.ex	ngali	ngalku	ngalku		ngalu	ngali	ngali
2du	yula	yulku	yulku	yula		yula	yula
3du	pula	pulku	pulku	pula		pula	pula
1pl.in	ngayana	nganyudu	nganyudu			ngandra	ngandra
1pl.ex	ngayani	ngayini	ngayani			ngana	ngana
2pl	yuda ^b	yuda	yuda	yuda		yuda	yuda
3pl	thana	thana	thana	thana		thana	thana
who	wara	wara	wara	wara	wara	wara	wara
what	minha	minha	minha	minha	minha	minha	minha
where	wardayadi	wararda	warada	wardali	warda-	yilanggi	yilanggi
when	wintha	wintja		wintjala		wintjama	walpi

Note: ^aThis and the next two names are abbreviated here; the names are Karruwal, Yawarrawarka and Yandruwandha.

^bNote that intervocalic <d> in my orthography for these languages is an alveolar tap.

¹² Bownen (2006) suggests that Diyari has borrowed the first person dual inclusive pronoun *ngaldra* from Yandruwandha. I would suggest that Diyari has largely reshaped its pronoun set on the model of Yandruwandha, including changing the other first person dual *ngali* from inclusive to exclusive.

Table 2 Central Karnic noun morphology

	Nominative	Ergative	Accusative	Dative	Locative	Ablative
Diyari						
Sg	-0	-li, -yali	-0	-ya	-nhi	-ndru
NSg	-0	-li	-nha	-rni	-ngu	-ngundru
Fem name	-ni	-ndru	-nha	-nhangka	-nhangu	-ngundru
Male name	-nha	-li	-nha	-rni	-ngu	-ngundru
Ngamini	-0	-nu	-0	-ngka	-mu	-ngundu
Yaluyandi	-0	-ndu	?	-ngka	-mu	-ndu
Mithaka	-0	-ndu, -lu	-0	-ngadi	-ngu	?
Yawarrawarrka	-0	-li	-0	-ma	-nyi, -ni	-nguda
Yandruwandha	-0	-li	-0	-ngadi	-yi	-nguda

strongly suggest that Ngamini and Yaluyandi (and presumably the intervening Karangura) are dialects of a language.

A striking feature of the personal pronoun inventory is the resemblance of the Diyari set to the Yawarrawarrka and Yandruwandha pronouns. Only the first person singular (slightly) and the two first person plural pronouns are different. On the other hand, the Ngamini and Yaluyandi pronouns, which differ from one another only slightly in one form, differ from the Diyari set in eight (Yaluyandi) or nine (Ngamini) of the 12 personal pronouns in the table. Note also the resemblances between Diyari and Yandruwandha/Yawarrawarrka and its differences from Ngamini/Yaluyandi in the ergative and locative suffixes in Table 2.

Other comparisons tend to support the present classification. I tentatively suggest that Central Karnic comprises four closely-related languages:

- Diyari, with dialects Thirrari (Austin 1981: 4–6, 1990a) and probably Pilardapa (Austin 1990a, b);
- Ngamini and Yaluyandi, presumably with Karangura (see Austin 1991);
- Mithaka with Marrulha and Karruwali;
- Yandruwandha and Yawarrawarrka with related dialects.

My figures for North-West Karnic are:

Table 3 Central Karnic verb morphology

	Imperative	Past	Present	Future/Purposive	Reflexive	Reciprocal
Diyari	-0, -ya	-ya	-yi	-lha	-tjarri	-mali
Ngamini	-0, -ya	-rna wadayi	-yi, -ya	-lha	-tjarri	-mali
Yaluyandi	?	-ndu, -nda	-yada	-lhangka	?	?
Mithaka	-iya	-ndadi	-nda(kuli)	-nanga	?	-pali
Yawarrawarrka	-0	-ni (imm.) -itha (near)	-rla	-iya	-yindri	-yindri
Yandruwandha	-0	-na (imm.) -nhana (near)	-rla	-nga	-yindri	-yindri

Note: ‘imm’ refers to immediate past tense and ‘near’ to near past tense (see Breen 2004b).

Pitta-Pitta/Wangka-Yutjurru, contiguous dialects of separate, but closely related, languages

%N%V	%N/%V	nN, nV	chi ²	significance
60–71	0.85	(178, 42)	1.85	low, > 80%

These figures suggest that they could actually be dialects of a single language. Each of these is regarded as one of a number of closely-related dialects and this does seem to be so, but the above figures perhaps argue for the two clusters of dialects not being sufficiently different to be classed as two languages. See Blake and Breen (1971: 16) for the original justification for this classification, which is supported by phonological and grammatical differences between the two dialect groups, and is maintained.

Wangka-Yutjurru/Wangkangurru, contiguous languages but separated by Wangkamanha, a dialect of the same language as Wangka-Yutjurru, which results in a long distance of separation. Pronouns are substantially different.

%N%V	%N/%V	nN, nV	chi ²	significance
25–39	0.64	(156, 37)	2.86	moderate, > 90%

Pitta-Pitta/Wangkangurru, not contiguous languages and with a long distance between them. Pronouns are substantially different.

%N%V	%N/%V	nN, nV	chi ²	significance
26–36	0.72	(153, 35)	1.20	low, about 75%

Wangkangurru (and its close relative Arabana) were excluded by Austin (1990a) from Karnic, mainly because of differences in pronouns and case forms. They were included by Bower (1998: 206–207).

Although not highly significant, the counts comparing the northern language(s) with the southern one, with verb percentage substantially higher than noun percentage, suggest fairly strongly that they are genetically related. This is supported by the apparent shared innovations listed by Breen (2007: 185–186).

The other language that has always been regarded as Karnic is the set of dialects represented by (modern) Wangkumara (see Breen 1971a: 12). Some counts for Wangkumara with the two Karnic branches are:

Wangkumara/Yandruwandha (Central Karnic), contiguous

%N%V	%N/%V	nN, nV	chi ²	significance
19–14	1.36	(178, 44)	0.098	extremely low

Wangkumara/Ngamini (Central Karnic), not contiguous

19–12	1.58	(171, 44)	0.37	very low
-------	------	-----------	------	----------

Wangkumara/Pitta-Pitta (North-West Karnic but separated by a long distance)

13.8–9.5	1.45	(174, 42)	0.55	very low
----------	------	-----------	------	----------

Wangkumara/Wangkangurru (North-West Karnic but separated by a long distance and by Central Karnic languages)

16.0–5.6	2.86	(162, 36)	2.67	moderate, about 90%
----------	------	-----------	------	---------------------

These figures, especially the first set for the contiguous pair, Wangkumara and Yandruwandha, suggest that Wangkumara may not even be in the same family as the others. For that pair, nominative pronouns are very similar but there are big differences in pronoun morphology. There is little in common in noun and verb morphology; see Table 4 [extracted from Tables 7, 8 and 10 in Breen (2007); for Wangkumara M = masculine singular, N = non-masculine singular and plural, D = dual]. The count for Wangkumara with Margany, contiguous and a Mari language, is 15–13 [see Section 2(a)], not very different from the Wangkumara/Yandruwandha figure.

Some Wangkumara words which have a resemblance to Karnic forms are regarded as not cognate with them but cognate with other languages; for example, Wangkumara *thalanya* ‘tongue’ is cognate with very similar Mari, Kungkari and Paakantyi forms and not with Karnic *tharli*; Wangkumara *thaltha* ‘eat’ is probably cognate with *thala/dhala* in Mari languages and not with Karnic *thayi/thatji* (Wangkangurru *thanhi*); Wangkumara *nhadja* ‘see’ is judged cognate with Kalali *nhantja* and perhaps with *nhaga/nhaka* and variants thereof in Mari languages and Kungkari and not with Karnic *nhayi/nhatji* (Wangkangurru *nhanhi*).¹³

Kalali and Badjidi, spoken to the east of Wangkumara, were originally included in Karnic by Breen (1971a). The following figures, some repeated from Section 2(a), supplement the notes on them in Breen (2007).

Wangkumara/Kalali

%N%V	%N/%V	nN, nV	chi ²	significance
45–24	1.87	(151, 35)	6.1	about 98%

Wangkumara/Badjidi

27–17	1.59	(145, 35)	1.43	low, about 75%
-------	------	-----------	------	----------------

Margany/Kalali (contiguous)

26–15	1.73	(147, 39)	1.98	low, > 80%
-------	------	-----------	------	------------

¹³ The cognacy of Wangkangurru (and Arabana) *thanhi* and *nhanhi* with the other Karnic forms is supported by the trio *wayi/watji/wanhi* for ‘cook’.

Table 4 Comparison of noun and verb morphology, Wangkumara and Yandruwandha

	Wangkumara	Yandruwandha
<i>Nouns</i>		
Nominative	–ia (M), –nhani (N), –0 (D)	–0
Ergative	–ulu (M), –(a)ndru (N), –ngu (D)	–li
Accusative	–(i/a)nha, –nhanha	–0
Dative	–(a)nga	–ngadi
Locative	–langa, –ngala (D)	–yi
Ablative	–(a)ndru	–nguda
<i>Verbs</i>		
Imperative	–0	–0
Recent past	–garli	–nhana
Present	–garla	–rla
Purposive	–rra, –da	–nga
Reflexive	–ii	–yindri

Margany/Badjidi (contiguous)

37–9 4.1 (141, 35) 10.5 about 99.9%

Kalali /Badjidi

45–17 2.65 (126, 35) 9.1 about 99.8%

Kalali /Paakantyi (contiguous languages but separated by some Paakantyi dialects)

16.9–7.9 2.14 (136, 38) 1.90 low, > 80%

Badjidi/Paakantyi, (contiguous languages but separated by some Paakantyi dialects)

25–17 1.47 (130, 35) 0.95 very low

These figures are all very low and, more importantly, are consistently substantially lower for the verbs than for the nouns. It seems that the several Wangkumara dialects, Kalali and Badjidi are a group of three isolates. An alternative view is that Wangkumara and Kalali (at least) are distantly related, and that they came together after a long separation, perhaps having figures of something like 10–20 at the time, and have since been borrowing (mostly nouns) from one another. [Compare some of the figures discussed in Section (e) below.] The substantial grammatical differences between them (see Breen 2007: 194) incline me to reject the latter alternative.

Now we consider the relationship of North-west Karnic, i.e. Pitta-Pitta, Wangka-Yutjurru, Wangkangurru and associated dialects, to Central Karnic (represented here by Diyari, Ngamini, Mithaka and Yandruwandha).

Wangkangurru/Ngamini, contiguous

%N%V	%N/%V	nN, nV	chi ²	significance
41–26	1.58	(154, 36)	2.61	about 90%

Pitta-Pitta/Diyari, not contiguous

19–24	0.79	(159, 37)	0.49	very low
-------	------	-----------	------	----------

Pitta-Pitta/Mithaka, contiguous languages but separated by the Pitta-Pitta dialect Mayawarli; Mithaka poorly attested

49–29	1.69	(136, 28)	2.98	>90%
-------	------	-----------	------	------

Pitta-Pitta/Ngamini, not contiguous

22–20	1.10	(162, 41)	0.112	extremely low
-------	------	-----------	-------	---------------

Pitta-Pitta/Yandruwandha, not contiguous

19–16	1.19	(181, 43)	0.145	extremely low
-------	------	-----------	-------	---------------

These figures, especially those with the higher cognate percentages, could suggest a genetic relationship, a long separation, and a return to contiguity (as in the alternative scenario described above for Wangkumara and Kalali). The rather high verb percentages argue for their having some genetic relationship. I propose (but tentatively) that only these two branches are descended from a proto-Karnic language.

Comparing Karnic languages to neighbouring languages we have, in addition to Wangka-Yutjurru/Warluwarra and Wangkangurru/Yankunytjatjara given above [in Sections 2(b) and 2(c) respectively],

Yalarnnga (isolate)/Pitta-Pitta (Karnic), contiguous

%N%V	%N/%V	nN, nV	chi ²	significance
15–10	1.50	(164, 43)	0.64	very low

Wangka-Yutjurru/Antekerrepenh (Arandic), contiguous

11.65–7.0	1.66	(163, 43)	0.78	very low
-----------	------	-----------	------	----------

Pitta-Pitta/Antekerrepenh, separated by Wangka-Yutjurru, to which Pitta-Pitta is closely related

6.2–4.9	1.27	(162, 41)	0.099	extremely low
---------	------	-----------	-------	---------------

These figures seem to show that languages to the north and west are quite unrelated. I have no data on the Thura-Yura and Yarli languages to the south. Neighbouring languages to the east are very poorly attested. Further east we have the Mari languages, and we can make some long-distance comparisons:

Margany/Yandruwandha, not contiguous; separated only by Wangkumara

%N%V	%N/%V	nN, nV	chi ²	significance
7.9–2.3	3.43	(165, 44)	1.85	low, about 80%

Margany/Pitta-Pitta, not contiguous

4.8–6.1	0.79	(188, 41)	0.121	extremely low
---------	------	-----------	-------	---------------

These figures suggest that Karnic and Mari belong to two different families.

(e) *Arandic*

I have also applied my counting to the Arandic languages [including Kaytetye, whose membership I have doubted for some time (see Koch 2004: 129); also, see Breen (1990: 156) for some early relevant figures]. Counts given above for Antekerrepenh with Wangka-Yutjurru and Pitta-Pitta suggest no relationship on the eastern side. The following counts are for within Arandic (as generally understood), involving the three northernmost members, the two southernmost and one in between. Kaytetye and Lower Arrernte are regarded as separate languages; the others are more closely related, perhaps all as dialects of a single language. The vocabularies used are from the latest revised version of Hale's comparative Arandic Wordlist. The number of comparable pairs of nouns ranged from 220 to 264 and of verbs from 38 to 46. Nominal categories are body parts (55), fauna (54), material culture (37), qualities (translated as adjectives in English) (32, of which nine cannot be applied to inanimate objects), inanimate nature and flora (each 25), people (24, mostly kin terms), time/location/direction (16). There are 48 verbs. The 14 pronominal terms are not counted.

Kaytetye/Northern Alyawarr, contiguous

%N%V	%N/%V	nN, nV	chi ²	significance
55–42	1.31	(235, 43)	2.56	moderate, nearly 90%

Kaytetye/Eastern Arrernte, not contiguous

45–34	1.32	(264, 46)	1.62	low, about 80%
-------	------	-----------	------	----------------

Kaytetye/Southern Arrernte, not contiguous

38–31	1.23	(258, 45)	0.74	very low
-------	------	-----------	------	----------

Kaytetye/Lower Arrernte, not contiguous

31–30	1.03	(239, 42)	0.044	extremely low
-------	------	-----------	-------	---------------

Northern Alyawarr/Southern Alyawarr, contiguous

79–97	0.81	(227, 38)	7.17	> 99%
-------	------	-----------	------	-------

Northern Alyawarr/Eastern Arrernte, not contiguous

58–74	0.78	(229, 44)	4.06	about 95%
-------	------	-----------	------	-----------

Northern Alyawarr/Lower Arrernte, not contiguous

46–67	0.69	(220, 39)	5.71	about 98%
-------	------	-----------	------	-----------

Southern Arrernte/Lower Arrernte, contiguous

58–79	0.73	(243, 40)	6.20	nearly 99%
-------	------	-----------	------	------------

All counts involving Kaytetye have a noun figure higher than the verb figure. All other counts have the verb figure substantially higher than the noun figure.

Hale's comparative list includes 18 bound morphemes, nine nominal suffixes and nine verbal. I have scanned these, without filling in any of the gaps (which could certainly be done, though perhaps not completely). Of the nine verbal suffixes, Kaytetye does not share any with any of the others. The others share from four to eight with one another. Of the nominal suffixes, Kaytetye shares six with the nearest of the others (Northern Alyawarr) ranging down to two with Southern Arrernte. Lower Arrernte shares, on average, about half of the suffixes with the others (apart from Kaytetye). The others share most of the suffixes with one another.

I compared Kaytetye with its non-Arandic neighbours, Warlpiri on the northwest, Warumungu on the north and Wakaya on the northeast. For Warlpiri and Warumungu I used lists matching the Hale Arandic lists;¹⁴ however, a handful of verbs were not counted because they were compound. For example, Warlpiri has no verb root meaning 'to hear'; it is expressed as *purda-nya-* in which *nya-* is the verb 'to see', or better, 'to perceive'. *Purda* does not occur other than in this compound. Comparisons involving Arandic languages are complicated by the sound changes they have undergone (see Koch 1997). The pairs of verbs counted as cognate for Warlpiri and Kaytetye are (Warlpiri form first) *nyina-/an-* 'sit', *nguna-/nw-* 'lie', *wangka-/angk-* 'speak', *parnti-/iytn-* 'smell', *ngarni* (pres.)/*ayn-* 'eat', *kulpa-/alp-* 'return', *pangi-/ang-* 'dig', *yinyi* (pres.)/*tny-* 'give', *yirra-/arr-* 'put', *marnpi-/atnp-* 'touch', *luwa-/w-* 'hit with missile', *kampa-/amp-* 'burn'. In the case of 'eat' I assume that, since deletion of the first syllable of the parent form has removed the whole root, the Kaytetye form is

¹⁴ Compiled by Belinda Keillor.

based on the present tense suffix. This may also be the case for ‘give’, but another possibility is that it is a comparatively recent loan from Kaytetye into Warlpiri. In the case of ‘smell’ the Warlpiri verb is intransitive but the Kaytetye verb is transitive. In the case of ‘touch’ the Kaytetye verb seems to have alveolar where the Warlpiri has retroflex, but the corresponding Eastern Arrernte verb has the retroflex (*arnp-*) while in Kaytetye the verb *artnp-* means ‘run’.

Figures are as follows:

Kaytetye/Warlpiri (contiguous)

%N%V	%N/%V	nN nV	chi ²	significance
17–31	0.55	(222, 36)	3.5	moderate, about 93%

Kaytetye/Warumungu (contiguous)

9.5–27	0.63	(227, 33)	8.8	> 99.5%
--------	------	-----------	-----	---------

Kaytetye/Wakaya (contiguous) [repeated from Section 2(b)]

8.3–6.5	1.28	(144, 31)	0.052	extremely low
---------	------	-----------	-------	---------------

Obviously we can disregard Wakaya as a likely relative of Kaytetye, but the other two seem very likely.

I then did counts of two other Arandic wordlists with Warlpiri, with the following results:

(Western) Anmatyerr/Warlpiri (contiguous)

%N%V	%N/%V	nN nV	chi ²	significance
17.6–24	0.73	(199, 33)	0.67	very low

Eastern Arrernte/Warlpiri

16.0–26.5	0.60	(243, 34)	0.69	very low
-----------	------	-----------	------	----------

The Anmatyerr wordlist lacks many flora and fauna names, which probably means the noun percentage should be a little higher and the significance a little lower—one third of the 36 pairs that could be compared in these fields were cognate. (Eastern Arrernte, which is not contiguous with Warlpiri, shares 28% of 63 items with it in these two categories.)

A count for Warlpiri vs Warumungu gave:

%N%V	%N/%V	nN nV	chi ²	significance
17.9–16.2	1.10	(215, 34)	0.096	extremely low

(An earlier count with a much smaller sample, 91 nouns and 29 verbs, gave percentage figures of 16 and 21, respectively, again with low significance.)

These figures suggest to me that there was a proto-language (or a series of proto-languages) which gave rise to four (at least) branches: Warlpiri, Warumungu, Kaytetye and Arandic (minus Kaytetye). If we look at the figures given in Section 2(c) above for the Wati languages with Walmatjarri, we find evidence for two more branches of such a proto-language: Warlpiri is related to Walmatjarri in that they belong to the two branches of Ngumpin-Yapa [McConvell and Laughren (2004) supported by my count with figures 18–25] so we can tentatively add Ngumpin and Wati to the descendants of our proto-language (proto-Nyungan?).

My explanation of the Kaytetye/(other) Arandic percentages is that Kaytetye had been separated from the Arandic branch for a long time (as had the other branches) and would have been related to the other branches as they are to one another, with figures of something like 10–20. It then came back into contiguity with (other) Arandic languages and began to borrow from them, predominantly nouns, thus elevating the noun figure above the verb figure. It may well be that Kaytetye is more closely related to the (other) Arandic languages than to any of the other branches, but this remains to be proven. In any case, Harvey's observation (1997: 184) that there is likely to be a high level of diffusion into a language spoken by a small population from a contiguous language with a much larger population seems to be appropriate for the Kaytetye/Arandic situation.

4. Discussion

The category 'noun' is not monolithic. I divide nouns up into 10 semantic fields, not counting two fields, one demonstratives plus interrogatives and the other personal pronouns, which I have not included in my recent counts. These 10 fields can be grouped into three sets: body parts, flora and fauna, and remainder. The body-part field is the one which contains most of the nouns which a pair of languages have in common if they are not related and not contiguous. Flora and fauna (two fields), especially the latter, provide most of the common items in most unrelated and contiguous pairs. The remaining seven fields have, between them, hardly any common forms if the languages are unrelated, and even if they are related but rather distant. For example, the Mari languages Margany and Gugu Badhun share 44% of body-part terms but only 13% of the human classification, inanimate nature, culture and qualities fields together. The corresponding figures for the Ngarna languages Warluwarra and Yanyuwa are 18% and 5%.

Average percentage figures for counts in six semantic fields, plus verbs, are given in Table 5, which can be thought of as an updated and expanded version of Table 7.1 in Breen (1990). The pairs of languages are grouped in five categories:

- A. contiguous dialects or closely-related languages (five pairs);
- B. closely-related languages separated only by a dialect or dialects of one (two pairs);

- C. contiguous unrelated languages (six pairs);
- D. distant related languages (three pairs);
- E. distant unrelated languages (four pairs). ‘Distant’ means ‘separated by at least one other language’.

Pairs in the various categories are:

- A. Bidyara/Gunggari, Gunya/Margany, Pitta-Pitta/Wangka-Yutjurru, Northern Alyawarr/Southern Alyawarr, Ngamini/Diyari;
- B. Warluwarra/Bularnu, Northern Alyawarr/Lower Arrernte;
- C. Warluwarra/Wangka-Yutjurru, Wanyi/Wakaya, Garrwa/Yanyuwa, Margany/Wangkumara, Wangka-Yutjurru/Antekerrepenh, Warluwarra/Kalkutungu;
- D. Bularnu/Yanyuwa, Warluwarra/Wakaya, Margany/Gugu Badhun;
- E. Bidyara/Kalkutungu, Wanyi/Yanyuwa, Pitta-Pitta/Antekerrepenh, Margany/Paa-kantyi.

Figures in the line below column headings give the range of numbers of items in that category. In a couple of cases all numbers except one are in a narrow range, and one (involving the Hale wordlist) is much higher; this number is given separately.

The ‘Common Australian’-type words, including the pronouns, may come from the unobservable level (see Breen 1997: 87–91), a substratum, preceding the origins of the language families I have proposed or hinted at. These words are particularly prominent in the body part and verb categories, and they, along with perhaps a handful of regional flora or fauna terms, may be almost the only words that pairs of unrelated languages share. For example, Margany (Mari) and Pitta-Pitta (Karnic) share:

kupakupa ‘old man’ (regional; at least one of the languages has also another word, so it counts as a half);
nganka ‘beard’, *mara* ‘hand’, *ngama* ‘breast’ (half), *kuna* ‘faeces’, *thina* ‘foot’, all ‘Common Australian’;
paya ‘bird’ (half), *kulpar(r)i* ‘emu’ (half), *kuti*~*kuturu* ‘swan’, *wakarla* and similar ‘crow’, *kila*+ ‘galah’ (half), mostly regional, but the word for crow is ‘Common Australian’ and probably onomatopoeic and ‘galah’ (often *kilakila* or *kilantji*) may be more than regional;

Table 5 Average cognate counts for nouns in semantic fields, and verbs

	Human classification (13–15; 24)	Body parts (31–47)	Fauna (19–45)	Inanimate nature (14–20)	Material culture (10–14; 32)	Qualities (12–16)	Verbs (38–47)
Group A	76	77	74	55.5	75	64	78
Group B	53	50	33	37	62	28	65
Group C	8.5	14.4	22	6.3	15.3	5.3	7.0
Group D	5.3	25	12.2	13.3	15.2	22	36
Group E	4.4	13.1	11	2.5	8.7	4.25	5.4

kutja/katja 'rotten' (half), possibly a chance resemblance;
patja ~ *patha* 'bite' (widespread), *madi* ~ *mada* 'get' (half) (perhaps 'Common Australian'), *paka* 'dig' (widespread, sometimes *paku*);
 and a number of pronouns.

In follow-up work, an improved division of words into semantic fields should be considered. For example, we should investigate the distribution of words for items like 'tail', 'feather', 'egg', 'tree', 'leaf', 'bark', 'root', which should perhaps be treated as body parts, leaving only species names to be in the fauna and flora fields.

I do not have figures to support my impression (Breen 1990: 154) that Dixon's (1980: 254) claim that there is no 'basic vocabulary' which is less susceptible to change than less 'basic' vocabulary seems to be correct. The results of any exploratory research it might have been based on [which would presumably have preceded the work described in Breen (1971a)] are lost. The claim is supported by the figures given in Section 2(a) for Margany and Gunya: almost identical figures were obtained for a count based on the 250-word lists and another count based on the whole of the remaining available vocabulary. However, these are two contiguous and closely-related dialects. Similar figures for Kalkutungu and Yalarnnga [Section 3(b)] are not so convincing: the verb figures are almost the same for both counts but the noun figure for the 'remainder' count is somewhat higher. This could be taken to suggest that the rate of borrowing for these, presumably less basic, words is higher—and it is in most of the semantic fields—but there is another important factor, and this is that the proportion of body-part words in the 'remainder' list is much lower, and this is almost the category with the lowest proportion of 'sames' in the 250-word list count. (Most body-part words are in the 250-word list.)

In my recent research, for Warlpiri and Warumungu I originally used vocabularies I had compiled many years ago for literacy-teaching purposes, containing roughly half as many nouns as the lists I ultimately used, which were compiled to correspond to Hale's comparative Arandic lists. (The numbers of verbs were not very different.) Five counts involving pairs of rather distantly related languages gave the following noun cognate figures:

13% of 126 compared to 17% of 222;
 6% of 110 compared to 9.5% of 227;
 14% of 114 compared to 17.6% of 199;
 14% of 118 compared to 16% of 243;
 16% of 91 compared to 18% of 215.

These figures do suggest that having a larger list leads to higher cognate figures for distantly related languages. The distribution of the nouns among semantic fields was very similar in the two lists; a couple of differences suggest no cause for the differences in the cognate counts. However, none of the lists were compiled with basicness in mind.

This raises the question of what should be the composition of a vocabulary sample for comparative purposes. Lehmann (1966: 17) says 'words for lower numerals,

kinship, domestic animals, everyday activities' and 'the lower numerals, pronouns, items referring to parts of the body, to natural objects—animals, plants, heavenly bodies, and so on' (1966: 108). Campbell (1998: 126) says 'body parts, close kinship terms, low numbers, common geographical terms'. The latter two lists, if we add 'everyday activities', correspond more closely to Swadesh's pioneering 100-word and 200-word lists [as they appear in Lehmann (1966: 112–113); Campbell (1998: 201–202) gives only the 100-word list]: the only domestic animal in either list is the dog, and there are no kinship terms in the shorter list and just four in the longer. For Australia, O'Grady *et al.* (1966: 23) say that 'Curr effectively predetermined the scope of any widely applicable list in 1886 by publishing a 120 item wordlist obtained from several hundred communities, often of now extinct languages'. However, their list [as published by O'Grady and Klokeid (1969)] has half as many again body-part words in their 100-word list as Curr has in his 120—32 compared to 20 (and also, as one would hope, many more verbs).

Surely a basic list for Australian languages, to be culturally appropriate, should have a strong complement of kinship terms (as Hale's list has). However, some widespread kinship terms do tend to have different referents in different areas. To make a list apply to the whole continent it must not have flora or fauna specific names (except 'dog'). However, these must be available on a regional basis. The whole concept of basic vocabulary needs to be researched; perhaps better than a comparison of noun cognate percentages versus verb cognate percentages would be a basic versus non-basic cognate percentages, the non-basic maybe predominantly flora and fauna specific names (and so varying from region to region).

The question of whether the method of comparison described here could be applied outside Australia has hardly been studied. See Breen and Blake (2007: 71, note 3) for comparable data obtained by Blake for a few European languages, representing two branches of Indo-European. Overall percentages and percentages for verbs plus adjectives were compared and the differences looked significant. I have done counts for three Malayo-Polynesian (Austronesian) languages for which I have some information, and the results do not look promising. A reviewer quotes Beckwith (2004) as claiming 'that verbs are borrowed just as easily as nouns in agglutinative languages, on the argument that (1) Uralic, Turkic, Mongolic, etc. share a certain amount of basic vocabulary, including pronouns and verbs, (2) the languages are unrelated, and therefore (3) pronouns and verbs are readily borrowed'. Premise (2) would not be accepted by all linguists, and a method like this one might provide ammunition for one side or the other in the debate.

References

- Alpher B & D Nash 1999 'Lexical replacement and cognate equilibrium in Australia' *Australian Journal of Linguistics* 19: 5–56.
 Austin P 1981 *A Grammar of Diyari, South Australia* (Cambridge Studies in Linguistics, 32) Cambridge: University Press.

- Austin P 1990a 'Classification of Lake Eyre languages' *La Trobe Working Papers in Linguistics* 3: 171–201.
- Austin P 1990b 'The last words of Pirlatapa' in P Austin, *et al.* (eds) *Language and History: essays in honour of Luise A. Hercus* Canberra: Pacific Linguistics. pp. 29–48.
- Austin P 1991 'The Karangura language' *Records of the South Australian Museum* 25(2): 129–137.
- Austin P 1994 'Diyari' in N Thieberger & W McGregor (eds) *Macquarie Aboriginal Words: a dictionary of words from Australian Aboriginal and Torres Strait Islander languages* North Ryde: Macquarie Library. pp. 125–143.
- Barrett B 2005 *Historical reconstruction of the Maric languages of central Queensland* Unpublished Master of Linguistics thesis, Australian National University, Canberra.
- Beale A 1974 *A grammar of the Biri dialect of North Queensland* Unpublished MS, Australian National University, Canberra.
- Beckwith CI 2004 *Koguryo, the Language of Japan's Continental Relatives: an introduction to the historical-comparative study of the Japanese-Koguryoic languages, with a preliminary description of archaic Northeastern middle Chinese, volume 21* Brill's Japanese Studies Library Leiden: Brill.
- Black P 1997 'Lexicostatistics and Australian languages: problems and prospects' in D Tryon & M Walsh (eds) *Boundary Rider: essays in honour of Geoffrey O'Grady* Canberra: Pacific Linguistics. pp. 51–69.
- Black P 2006 'Equilibrium theory applied to Top End Australian languages' in K Allan (ed.) *Selected papers from the 2005 Conference of the Australian Linguistic Society* Available at: <http://www.als.asn.au/proceedings/als2005/black-equilibrium.pdf>.
- Blake BJ 1979 *A Kalkatungu Grammar* (PL B-57) Canberra: Pacific Linguistics.
- Blake BJ 1987 *Australian Aboriginal grammar* London: Croom Helm.
- Blake BJ 1988 'Redefining Pama-Nyungan: towards the prehistory of Australian languages' in N Evans & S Johnson (eds) *Aboriginal Linguistics 1* Armidale: University of New England. pp. 1–90.
- Blake BJ 1990 'Languages of the Queensland/Northern Territory border: updating the classification' in P Austin, RMW Dixon, T Dutton & I White (eds) *Language and History: essays in honour of Luise A. Hercus* (PL C-116) Canberra: Pacific Linguistics.
- Blake BJ & G Breen 1971 *The Pitta-Pitta Dialects* (Linguistic Communications 4) Melbourne: Monash University.
- Blake BJ & G Breen 1990 'Guwa' in G Breen (ed.) *Salvage Studies of Western Queensland Aboriginal Languages* Canberra: Pacific Linguistics. pp. 108–144.
- Bowern C 1998 *The case of Proto-Karnic: morphological change and reconstruction in the nominal and pronominal system of Proto-Karnic (Lake Eyre Basin)* Unpublished BA (Hons) thesis, Australian National University, Canberra.
- Bowern C 2006 'Another look at Australia as a linguistic area' in Y Matros, A McMahon & N Vincent (eds) *Linguistic Areas: convergence in historical and typological perspective* Basingstoke, England: Palgrave MacMillan. pp. 244–266.
- Breen G 1971a *Aboriginal Languages of Western Queensland* (Linguistic Communications 5) Melbourne: Monash University. pp. 1–88.
- Breen G 1971b *A description of the Waluwara language* Unpublished MA thesis, Monash University, Victoria.
- Breen G 1973 *Bidyara and Gungabula: grammar and vocabulary* (Linguistic Communications 8) Melbourne: Monash University.
- Breen G 1981a 'Margany and Gunya' in RMW Dixon & BJ Blake (eds) *Handbook of Australian Languages Volume 2* Canberra: ANU Press. pp. 274–393.
- Breen G 1981b *The Mayi Languages of the Queensland Gulf Country* Canberra: AIAS.
- Breen G 1990 *Salvage Studies of Western Queensland Aboriginal Languages* (PL B-105) Canberra: Pacific Linguistics.

- Breen G 1997 'Taps, stops and trills' in D Tryon & M Walsh (eds) *Boundary Rider: essays in honour of Geoffrey O'Grady* Canberra: Pacific Linguistics. pp. 71–93.
- Breen G 2003 'Wanyi and Garrwa comparative data' in N Evans (ed.) *The Non-Pama-Nyungan Languages of Northern Australia: comparative studies of the continent's most linguistically complex region* Canberra: Pacific Linguistics. pp. 425–462.
- Breen G 2004a 'Evolution of the verb conjugations in the Ngarna languages' in C Bowern & H Koch (eds) *Australian Languages: classification and the comparative method* (Current Issues in Linguistic Theory 248) Amsterdam: John Benjamins. pp. 223–240.
- Breen G 2004b Innamincka Talk: a grammar of the Innamincka dialect of Yandruwandha with notes on other dialects (Pacific Linguistics 558) Canberra: Pacific Linguistics.
- Breen G 2004c *Innamincka Words: Yandruwandha dictionary and stories* (Pacific Linguistics 559) Canberra: Pacific Linguistics.
- Breen G 2007 'Reassessing Karnic' *Australian Journal of Linguistics* 27(2): 173–199.
- Breen G 2009 'The Biri dialects and their neighbours' *Transactions of the Royal Society of South Australia* 133(2): 219–256.
- Breen G 2011 'Reassessing Karnic: a reply to Bowern (2009)' *Australian Journal of Linguistics* 31(1): 137–142.
- Breen G & B Blake 2007 *The Grammar of Yalarnnga: a language of western Queensland* (Pacific Linguistics 584) Canberra: Pacific Linguistics.
- Campbell DM 1998 '[Wordlist in] Murunuda' in J Mathew (ed.) *Eaglehawk and Crow* Melbourne. pp. 205–272.
- Capell A 1956 *A New Approach to Australian Linguistics* (Oceania Linguistic Monograph 1) University of Sydney.
- Curr EM (1886–1887) *The Australian Race: its origin, languages, customs, place of landing in Australia, and the routes by which it spread itself over that continent* (4 volumes) Melbourne: John Ferres, Government Printer.
- Dixon RMW 1980 *The Languages of Australia* Cambridge: Cambridge University Press.
- Dixon RMW 1981 'Wargamay' in RMW Dixon & BJ Blake (eds) *Handbook of Australian Languages Volume 2* Canberra: ANU Press. pp. 1–144.
- Dixon RMW 1997 *The Rise and Fall of Languages* Cambridge: Cambridge University Press.
- Dixon RMW 2002 *Australian Languages: their nature and development* (Cambridge Language Surveys) Cambridge: Cambridge University Press.
- Green R 2003 'Proto Maningrida with proto Arnhem: evidence from verb inflectional suffixes' in N Evans (ed.) *The Non-Pama-Nyungan Languages of Northern Australia: comparative studies of the continent's most linguistically complex region* Canberra: Pacific Linguistics. pp. 369–421.
- Hale K 1974 *Comparative Arandic Wordlist* (version prepared by H Koch, with further editing and additions by G Breen).
- Harvey M 1991 'The temporal interpretation of linguistic diversity in the Top End' in *Archaeology and Linguistics: Aboriginal Australia in global perspective. Conference papers, drafts, volume 2*. pp. 75–92.
- Harvey M 1997 'The temporal interpretation of linguistic diversity in the Top End' in P McConvell & N Evans (eds) *Archaeology and Linguistics: Aboriginal Australia in global perspective* Melbourne: Oxford University Press. pp. 179–185.
- Harvey M 2009 'The genetic status of Garrwan' *Australian Journal of Linguistics* 29(2): 195–244.
- Hercus L 1993 *Paakantyi Dictionary* Canberra: the author.
- Hercus L 1994 *A Grammar of the Arabana-Wangkangurru Language, Lake Eyre Basin, South Australia* (PL C-128) Canberra: Pacific Linguistics.
- Johnson S 1990 'Social parameters of linguistic change in an unstratified aboriginal society' in P Baldi (ed.) *Linguistic Change and Reconstruction Methodology* Berlin: Mouton de Gruyter. pp. 419–433.
- Kirton J & Y Nagai 1984 *Yanyuwa-English Dictionary Files*.

- Koch H 1997 'Pama-Nyungan reflexes in the Arandic languages' in D Tryon & M Walsh (eds) *Boundary Rider: essays in honour of Geoffrey O'Grady* Canberra: Pacific Linguistics. pp. 271–302.
- Koch H 2004 'The Arandic subgroup of Australian languages' in C Bowern & H Koch (eds) *Australian Languages: classification and the comparative method* (Current Issues in Linguistic Theory 249) Amsterdam: John Benjamins. pp. 127–150.
- Lehmann WP 1966 *Historical linguistics: an introduction* New York: Holt, Rinehart and Winston.
- Mathews RH 1905 'Ethnological notes on the Aboriginal tribes of Queensland' *Journal of the Royal Geographical Society of Queensland* 20: 49–75.
- McConvell P & M Laughren 2004 'The Ngumpin-Yapa subgroup' in C Bowern & H Koch (eds) *Australian Languages: classification and the comparative method* (Current Issues in Linguistic Theory 249) Amsterdam: John Benjamins. pp. 151–178.
- Menning K (compiler) & D Nash (ed.) 1981 *Sourcebook for Central Australian Languages* (Pilot Edition) Alice Springs: Institute for Aboriginal Development.
- O'Grady GN & TJ Klokeid 1969 'Australian linguistic classification: a plea for coordination of effort' *Oceana* 39(4): 298–311.
- O'Grady GN, CF Voegelin & FM Voegelin 1966 'Languages of the world: Indo-Pacific fascicle six' *Anthropological Linguistics* 8(2).
- Simpson J (compiler) 2008 Wru-vocab-master-distrib-aseda.txt (2008 version).
- Sutton P 1973 *Gugu-Badhun and its neighbours: a linguistic salvage study* Unpublished MA thesis, Macquarie University, Sydney.
- Terrill A 1998 *Biri* (Languages of the World 258) Munich: Lincom Europa.
- Terrill A 2002 *Dharumbal: the language of Rockhampton, Australia* (Pacific Linguistics 525) Canberra: Pacific Linguistics.
- Wilson S & T Murray 1887 'Peak Downs District, Logan Downs Station' in EM Curr (ed.) *The Australian Race, Volume 3* Melbourne. pp. 54–69.

Appendix: Sources of Data

(The numbers are for locating the languages on the map.)

- | | | |
|----|----------------------------------|---|
| 1 | Alyawarr (Northern and Southern) | see Arandic |
| 2 | Anmatyerr | see Arandic |
| 3 | Antekerrepnh | Breen, unpublished |
| | Arandic languages | Hale comparative Arandic wordlists, ed. Breen |
| 4 | Badjidi | Breen, unpublished; Mathews (1905) |
| 5 | Bidyara | Breen, unpublished; Breen (1973) |
| 6 | Biri | Beale (1974) |
| 7 | Bularnu | Breen, unpublished |
| 8 | Dharumbal | Terrill (2002) |
| 9 | Diyari | Austin (1981, 1990, 1994) |
| 10 | Eastern Arrernte | see Arandic |
| 11 | Garrwa | Breen (2003) |
| 12 | Gugu Badhun | Sutton (1973) |
| 13 | Gunggari | Breen, unpublished |
| 14 | Gunya | Breen (1981a) |
| 15 | Guwa | Blake and Breen (1990) |
| 16 | Kalali | Breen, unpublished |
| 17 | Kalkutungu | Blake (1979) |
| 18 | Kaytetye | see Arandic |
| 19 | Kukatja | Menning and Nash (1981) |
| 20 | Kungkari | Breen (1990) |

- | | | |
|----|-------------------|---|
| 21 | Lower Arrernte | see Arandic |
| 22 | Margany | Breen (1981a) |
| 23 | Mayi-Thakurti | Breen (1981b); Gerhardts Laves field notes (c.1929) |
| 24 | Mithaka | Breen, unpublished; Campbell (1998) |
| 25 | Ngamini | Breen, unpublished |
| 26 | Ngawun | Breen (1981b) |
| 27 | Paakanytji | Hercus (1993) |
| 28 | Pirriya | Breen (1990) |
| 29 | Pitjantjatjara | Menning and Nash (1981) |
| 30 | Pitta-Pitta | Breen, unpublished; Blake (1979) |
| 31 | Southern Arrernte | see Arandic |
| 32 | Wakaya | Breen, unpublished |
| 33 | Wamatjarri | Menning and Nash (1981) |
| 34 | Wanyi | Breen (2003) |
| 35 | Wargamay | Dixon (1981) |
| 36 | Wangkangurru | Breen, unpublished; Hercus (1994) |
| 37 | Wangka-Yutjuru | Breen, unpublished |
| 38 | Wangkumara | Breen, unpublished |
| 39 | Warlpiri | Menning and Nash (1981); IAD (n.d.); Hale (1974) |
| 40 | Warluwarra | Breen, unpublished |
| 41 | Warumungu | Simpson (2008) |
| 42 | Yalarnnga | Breen and Blake (2007) |
| 43 | Yaluyandi | Breen, field notes; Austin (1990a) |
| 44 | Yambina | Wilson and Murray (1887) |
| 45 | Yandruwandha | Breen (2004c) |
| 46 | Yankunytjatjara | Menning and Nash (1981) |
| 47 | Yanyuwa | Kirton and Nagai (1984), dictionary files |
| 48 | Yawarrawarrka | Breen (2004c) |

