Historical linguistic geography of south-east Western Australia

David Nash


Presented to the Linguistic Issues in Native Title Claims workshop, Australian Linguistic Society (ALS) annual conference, University of WA, Saturday 2 October 1999.

Editorial contact:

John Henderson
Department of Linguistics
University of Western Australia
35 Stirling Hwy
Crawley WA 6009
ph. 08 9380 2870 fax 08 9380 1154
<John.Henderson@uwa.edu.au>

CONTENTS
1 Introduction ........................................................... 1
2 Vocabulary comparison ................................................. 2
  2.1 Previous work ..................................................... 3
  2.2 Comparison of south-east Western Australia vocabularies ....... 3
  2.3 Procedural difficulties ........................................... 9
  2.4 Significance ....................................................... 10
3 Maps ................................................................... 10
4 Pseudomaps ............................................................ 13
  4.1 Three languages? .................................................. 16
  4.2 Interpreting the results of pseudomapping ..................... 20
5 Conclusion ............................................................. 20
6 References ............................................................. 21

Map 1. '4.1 The South-East' language map ............................... 1
Map 2. Locations where language data collected (to 1980s) .......... 6
Map 3. Words for 'man' (or 'person') by location ...................... 11
Map 4. Words for 'elbow' by location .................................. 12
Map 5. Pseudomap of 26 vocabularies ................................ 14
Map 6. Pseudomap of 23 vocabularies (Map 5 less H, N, Q) ........ 16
Map 7. Sources of Map 5 with code letter marked according to 'person' word: kabun (underline), mariba (bold), or mirniny (italic) .... 18
Map 8. Locations of main vocabulary sources ........................... 19

version 22 September 2001
1 Introduction

This paper demonstrates something of the historical picture of relationships between language varieties that can be drawn from the rather meagre records that survive. In particular, we show how the technique of two-dimensional scaling can be applied to graphically represent relationships among vocabulary lists.

Our focus is on the languages of south-eastern Western Australia, the languages which are neither Western Desert nor of the Nyungar group: languages known today as Ngatju, Ngatjumaya, Marlpa, Kapun ('Gubrun'), Karlamayi, and Mirniny ('Mirning').¹ This group does not have an accepted name. O'Grady et al (1966) dubbed them the 'Mirniny Subgroup' which elevates the name of one particular subset. Here we follow Thieberger² 'languages of the South-East'; see below for more on language names.

¹ This work has been carried out in the context of consultancies since November 1998 for the Goldfields Land Council, whose support is gratefully acknowledged. The work in this paper has been partly in collaboration with Nick Thieberger, who has made helpful comments on an earlier version. I have also benefited from comments made when the paper was presented at the workshop on 2 October 1999, and from editorial comments.

² von Brandenstein used the term ‘Dundas District’ after the administrative region, in turn from Lake Dundas.

version 22 September 2001
I show that the documentary record of Aboriginal languages of the study region, from 1864 to the 1960s, has pretty much the same vocabulary in the same relative locations over that time, that is, the relative locations of languages (that is, the spatial arrangement of the various places of residence of speakers) show a continuity from the earliest records.

The historical comparative work necessarily concentrates on the distribution of vocabulary items, since few records contain expressions larger than the word. Few collections include numerous full sentences, and connected texts are recorded only in 1969-70 by von Brandenstein (1980). Texts, or at least a good collection of sentences, are needed to adequately study morphology and syntax. Without documentation of these parts of a language, comparisons with other languages can only be partial (even with languages related as dialects). Nevertheless, vocabulary comparisons of language records, even meagre ones, which have identifiable locations can provide useful indications of affiliations with languages more fully recorded in other places. Good examples are CC Hunt’s (1864) record of 14 words, and Helms (1896:325) record of Hampton Plains vocabulary (collected in 1891), which are most similar to languages of the South-East rather than any other subgroup, notably the Western Desert subgroup.

A promising avenue of inquiry is to investigate the associations detectable between place and language. A fairly straightforward kind of evidence is a direct relation between a toponym (place name) and a word (both its form and meaning) in a particular Aboriginal language. A couple of such analysable toponyms is indicative; a cumulative pattern of such links is more convincing. Investigation of this kind is postponed until we have a reasonable idea of the varieties of language across the area of study -- the kind of understanding which emerges from this paper and from which toponym studies can be a subsequent step.

Another link between language and space is investigated later in this paper. Clines of linguistic properties, or ‘linguistic distance’, can be correlated with spatial distance. If there is a good fit between the two measures, then it we can say that each language variety has a spatial niche, the district where it fits in. Then the variation of vocabulary over south-east Western Australia is studied to bring out its spatial component.

2 Vocabulary comparison

The comparison here reported is of the historical records of vocabulary of the languages of south-east Western Australia. The historical records (to 1970) are considered first. The vocabularies recorded before Aboriginal people’s residence moved to the larger towns also provides a detailed linkage of language variety with location, especially with respect to the vocabularies recorded by Daisy Bates (n.d.), which are usually assigned to specific Indigenous locations. While not the subject of this paper, the
study of old records is a good basis also for separate comparison with the
language knowledge of people still alive.

As stated earlier, vocabulary is the basis of comparison of the language
records since few old collections include expressions larger than the
word. Our initial comparison of the historical records of vocabulary has
focussed on 'basic vocabulary', from a standard list of meanings used in
comparing the 'core' vocabulary of languages. The particular list of
meanings used here is that used for 40-odd languages in the Sourcebook of
Central Australian Languages (Menning & Nash 1981) (hereafter, SCAL), and
includes the O’Grady (1960) adaptation3 for Australia of the Swadesh 100-
item and 200-item lists used worldwide. Concentration on 'basic
vocabulary' goes with a common (though not unchallenged) assumption that
such vocabulary is more stably inherited than vocabulary relating to more
culturally-specific meanings. In addition, there is the practical
consideration that historical records tend to record vocabulary for
'basic' meanings, so it is more available for comparison.

2.1 Previous work

The first quantification of language relationships in south-east Western
Australia came in a 1960s project spearheaded by G.N. O’Grady. This
lexicostatistically based classification was set out in a special issue of
Anthropological Linguistics and summarised on a wall map (O’Grady et al
1966a,b)4; its methodology was described by O’Grady & Klokeid (1969). In
the south-east of Western Australia, the map shows three languages grouped
in dialectal relationship, that is in pairs with over 70% of shared basic
vocabulary. From west to east in south-east Western Australia, they are
named ‘Kalarko’, ‘Ngadjunma’, ‘Mirning’. These three together with
Kalamai formed a subgroup, that is at least one of the three dialects
shares between 50% and 70% of basic vocabulary with Kalamai. The
collection of the four is named ‘Mirniny Subgroup’.

The O’Grady et al (1966a,b) classification provided the structure for A
revised linguistic survey of Australia (Oates & Oates 1970):

XXIX Pama-Nyungan family
   A Southwest group
      f Mirniny Subgroup
         55.1a Mirning A22 [AIAS-A9]
         55.1b Ngadjunma A25 [AIAS-A3]
         55.1c Kalarko A11 [AIAS-A2]
         55.2 Kalamai A12 [AIAS-A4]

2.2 Comparison of south-east Western Australia vocabularies

We have taken vocabulary lists representative of the variety of all known
sources and compared them to show similarities and differences. The lists
range over time and space, from the 1880s to 1970 and from various
locations within the south-east Western Australia region or from people

---

4 The accompanying article in Anthropological Linguistics, and O'Grady & Klokeid (1969:299-301), state the percentage of shared basic vocabulary
ranges: dialect over 70%, language 50%-70%, subgroup 25%-50%, group 15%-25%, and family under 15%.

version 22 September 2001
who were from that region. As mentioned above, these lists represent language records from the study area, other than of Nyoongar or WD languages.

We undertake this comparison for several reasons:
• to investigate the continuity of the languages spoken in the region,
• to interpolate where a wordlist probably came from if there is no geographic reference on the list, and
• to classify the lists, especially lists for which we only have the name of the speaker and no name for their language.

All the sources used in the study reported here are published or on open access in public archives in Australia. Those with more than ten items of the test list of ‘basic’ vocabulary are listed in Table 1. The sources in this table are located on Map 2 and Map 8. Table 1 also lists in the first column an identification code for those sources, all with more than 30 test-list items, for identification use in Table 2 and the analysis of §2.2 below.
Table 1. Source wordlists in comparative table, in chronological order, with size and location

<table>
<thead>
<tr>
<th>code</th>
<th>year</th>
<th>source</th>
<th>test list</th>
<th>Lat S (deg min)</th>
<th>Long E (deg min)</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>&lt;1881</td>
<td>Curr 34 Eyre</td>
<td>72</td>
<td>32°08'</td>
<td>126°18'</td>
<td>Eyre's Sand Patch</td>
</tr>
<tr>
<td>B</td>
<td>1880</td>
<td>Curr 35 Eucla</td>
<td>81</td>
<td>31°40'</td>
<td>128°52'</td>
<td>Eucla</td>
</tr>
<tr>
<td>C</td>
<td>1891</td>
<td>Helms</td>
<td>63</td>
<td>33°?</td>
<td>124°?</td>
<td>Symons Hill Rockhole, Fraser Ra.; Hampton Plains</td>
</tr>
<tr>
<td>D</td>
<td>1898</td>
<td>Garnier Coolgardie</td>
<td>50</td>
<td>30°58'</td>
<td>121°10'</td>
<td>Coolgardie</td>
</tr>
<tr>
<td>E</td>
<td>1907-14</td>
<td>Bates 2A, 1a 39/2-33 Gauera</td>
<td>105</td>
<td>33°30'</td>
<td>123°30'</td>
<td>Brooks' Place,</td>
</tr>
<tr>
<td>F</td>
<td>1907-14</td>
<td>Bates 2A, 1b 39/34-47</td>
<td>58</td>
<td>33°37'</td>
<td>123°53'</td>
<td>70 Mile, Israeli Bay</td>
</tr>
<tr>
<td>G</td>
<td>1907-14</td>
<td>Bates 2A, 3 39/55-60</td>
<td>33</td>
<td>32°15'</td>
<td>126°09'</td>
<td>Wongala Wardaiwarn</td>
</tr>
<tr>
<td>H</td>
<td>1907-14</td>
<td>Bates 2A, 4 39/62-70</td>
<td>22</td>
<td>31°40'</td>
<td>128°52'</td>
<td>Eucla district</td>
</tr>
<tr>
<td>I</td>
<td>1907-14</td>
<td>Bates 2A, 5 part 1 9/78-113</td>
<td>106</td>
<td>31°40'</td>
<td>128°52'</td>
<td>Jinyila</td>
</tr>
<tr>
<td>J</td>
<td>1907-14</td>
<td>Bates 2A, 6 part 1 39/122-132</td>
<td>40</td>
<td>31°55'</td>
<td>126°04'</td>
<td>Nunginija, Yayoudle, Cocklebiddy</td>
</tr>
<tr>
<td>K</td>
<td>1907-14</td>
<td>Bates 2A, 8 39/191-212</td>
<td>106</td>
<td>33°03'</td>
<td>123°23'</td>
<td>Deralinya</td>
</tr>
<tr>
<td>L</td>
<td>1907-14</td>
<td>Bates 2B, 2 40/30-40</td>
<td>40</td>
<td>33°35'</td>
<td>120°03'</td>
<td>Woolbanup Hill</td>
</tr>
<tr>
<td>M</td>
<td>1907-14</td>
<td>Bates 2C, 1 47/2-26</td>
<td>104</td>
<td>32°10'</td>
<td>121°35'</td>
<td>Dundas district</td>
</tr>
<tr>
<td>N</td>
<td>1907-14</td>
<td>Bates 2C, 2 47/29-32</td>
<td>25</td>
<td>31°50'</td>
<td>121°40'</td>
<td>Widgiemoooltha, Norseman</td>
</tr>
<tr>
<td>O</td>
<td>1907-14</td>
<td>Bates 2C, 3 47/35-56</td>
<td>82</td>
<td>30°58'</td>
<td>121°10'</td>
<td>Coolgardie</td>
</tr>
<tr>
<td>P</td>
<td>1907-14</td>
<td>Bates 2C, 4 47/59-84</td>
<td>89</td>
<td>30°45'</td>
<td>119°12'</td>
<td>Karratjibbin</td>
</tr>
<tr>
<td>Q</td>
<td>1907-14</td>
<td>Harvey in Bates 2C, 5 47/114-119</td>
<td>15</td>
<td>32°10'</td>
<td>121°35'</td>
<td>Norseman</td>
</tr>
<tr>
<td>R</td>
<td>1939</td>
<td>Tindale T86 Ngadjunma (Ngadju purara)</td>
<td>40</td>
<td>33°36'</td>
<td>123°53'</td>
<td>Israeli Bay</td>
</tr>
<tr>
<td>S</td>
<td>1939</td>
<td>Tindale T89 Kala:mai (includes 4 of genealogy Sheet 116 Kabu:(d)n)</td>
<td>45</td>
<td>31°14'</td>
<td>119°19'</td>
<td>Southern Cross</td>
</tr>
<tr>
<td>T</td>
<td>1939</td>
<td>Tindale T83 Mining</td>
<td>37</td>
<td>31°40'</td>
<td>128°52'</td>
<td>Eucla</td>
</tr>
<tr>
<td>U</td>
<td>1959</td>
<td>GN O'Grady S source</td>
<td>78</td>
<td>31°51'</td>
<td>132°42'</td>
<td>born nr Koorooringibbie</td>
</tr>
<tr>
<td>V</td>
<td>1958</td>
<td>GN O'Grady J source</td>
<td>60</td>
<td>31°40'</td>
<td>128°52'</td>
<td>Eucla /Carnarvon</td>
</tr>
<tr>
<td>W</td>
<td>1960</td>
<td>GN O'Grady P source</td>
<td>100</td>
<td>32°15'</td>
<td>126°03'</td>
<td>Twilight Cove /NE Norseman</td>
</tr>
<tr>
<td>X</td>
<td>1960</td>
<td>Hale &amp; O'Grady</td>
<td>63</td>
<td>30°55'</td>
<td>118°12'</td>
<td>Mukinbudin</td>
</tr>
<tr>
<td>Y</td>
<td>1970</td>
<td>CGvB RG</td>
<td>32</td>
<td>32°25'</td>
<td>123°53'</td>
<td>Balladonia / Norseman</td>
</tr>
<tr>
<td>Z</td>
<td>1970</td>
<td>CGvB PF</td>
<td>39</td>
<td>32°10'</td>
<td>121°35'</td>
<td>Norseman</td>
</tr>
</tbody>
</table>

Where the location is vague, I have used the coordinates of the place central to the district (for instance Norseman for ‘Norseman district’); similarly where the source mixes several places I have provided coordinates for a point approximately in the middle of the implied area.
Map 2. Locations where language data collected (to 1980s) and year of collection.
Comparison was restricted to equivalents of the 168-term test list used in SCAL, ignoring other meanings that may be represented in particular sources. To a limited extent near synonymous glosses are merged, for instance, words for ‘man’ may be taken from those forms glossed as ‘Blackfellow’ in an old record. Ideally the whole of each list should enter any comparison, but the restriction to the SCAL test list uses forms for meanings which are indeed most widespread in the available data and makes the task more manageable.

Vocabularies were entered in a comparative table with a column for each source and a row for each meaning. The forms were compared row by row. Forms judged to have a common root were assigned to the same correspondence set. (An alternative lexicostatistic method is to require full stem correspondence.) Note that correspondences are counted only if ‘homosemantic’ (having the same meaning).

Most lists lacked a majority of the 168 SCAL items (while some items had several words, that is several synonymous forms for a particular meaning). The number of SCAL items in each list (with at least one word) is given in the third column of Table 1. Since we allowed multiple synonyms, affinities between lists score a little higher than if the classic forced-choice-of-one is used.

The resulting percentages are in Table 2, using identification codes given in Table 1.
### Table 2. Percentages of corresponding vocabulary between pairs of 26 south-east Western Australia sources.

| code | Curr Eyre | Curr Eucla s | Helm 2A,1 | 2A,2 | 2A,3 | 2A,4 | 2A,5 | 2A,6 | 2A,8 | 2B,2 | C,1 | C,2 | C,3 | C,4 | C,5 | T86 | T89 | T83 | O’G | O’G | O’G | Hale vB | vB | S | J | P | RG | PF |
|------|-----------|--------------|-----------|------|------|------|------|------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A    | 65        | B             | 44        | 32   |      |      |      |      |      |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| B    | 27        | 27            | 52        |      |      |      |      |      |      |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| C    | 49        | 40            | 39        |      |      |      |      |      |      |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| D    | 55        | 50            | 75        | 59   | 75   | 60   |      |      |      |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| E    | 75        | 76            | 50        | 19   | 65   | 46   |      |      |      |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| F    | 34        | 29            | 55        | 46   | 67   | 57   | 44   | 50   | 44   | 52   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| G    | 53        | 41            | 68        | 40   | 64   | 68   | 44   | 50   | 47   | 54   | 69   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| H    | 31        | 25            | 61        | 51   | 42   | 46   | 34   | 44   | 31   | 49   | 55   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| I    | 41        | 33            | 89        | 57   | 50   | 61   | 42   | 67   | 43   | 55   | 82   | 64   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| J    | 33        | 29            | 52        | 40   | 29   | 36   | 34   | 36   | 35   | 30   | 50   | 57   | 50   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| K    | 34        | 32            | 56        | 44   | 37   | 43   | 46   | 36   | 39   | 31   | 34   | 55   | 64   | 50   | 79   |     |     |     |     |     |     |     |     |     |     |     |     |
| L    | 33        | 29            | 52        | 40   | 29   | 36   | 34   | 36   | 35   | 30   | 50   | 57   | 50   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| M    | 48        | 46            | 56        | 32   | 37   | 50   | 36   | 45   | 46   | 33   | 46   | 36   | 55   | 50   | 70   | 86   | 80   | 49   |     |     |     |     |     |     |     |
| N    | 38        | 31            | 62        | 35   | 60   | 67   | 25   | 64   | 37   | 29   | 66   | 60   | 50   | 61   | 42   | 54   | 88   |     |     |     |     |     |     |     |
| O    | 48        | 46            | 56        | 32   | 37   | 50   | 36   | 45   | 46   | 33   | 46   | 36   | 55   | 50   | 70   | 86   | 80   | 49   |     |     |     |     |     |     |     |
| P    | 38        | 31            | 62        | 35   | 60   | 67   | 25   | 64   | 37   | 29   | 66   | 60   | 50   | 61   | 42   | 54   | 88   |     |     |     |     |     |     |     |
| Q    | 53        | 41            | 68        | 40   | 64   | 68   | 44   | 50   | 47   | 54   | 69   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| R    | 31        | 25            | 61        | 51   | 42   | 46   | 34   | 44   | 31   | 49   | 55   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| S    | 48        | 46            | 56        | 32   | 37   | 50   | 36   | 45   | 46   | 33   | 46   | 36   | 55   | 50   | 70   | 86   | 80   | 49   |     |     |     |     |     |     |     |
| T    | 33        | 29            | 52        | 40   | 29   | 36   | 34   | 36   | 35   | 30   | 50   | 57   | 50   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| U    | 48        | 46            | 56        | 32   | 37   | 50   | 36   | 45   | 46   | 33   | 46   | 36   | 55   | 50   | 70   | 86   | 80   | 49   |     |     |     |     |     |     |     |
| V    | 38        | 31            | 62        | 35   | 60   | 67   | 25   | 64   | 37   | 29   | 66   | 60   | 50   | 61   | 42   | 54   | 88   |     |     |     |     |     |     |     |
| W    | 48        | 46            | 56        | 32   | 37   | 50   | 36   | 45   | 46   | 33   | 46   | 36   | 55   | 50   | 70   | 86   | 80   | 49   |     |     |     |     |     |     |     |
| X    | 33        | 29            | 52        | 40   | 29   | 36   | 34   | 36   | 35   | 30   | 50   | 57   | 50   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Y    | 48        | 46            | 56        | 32   | 37   | 50   | 36   | 45   | 46   | 33   | 46   | 36   | 55   | 50   | 70   | 86   | 80   | 49   |     |     |     |     |     |     |     |
| Z    | 33        | 29            | 52        | 40   | 29   | 36   | 34   | 36   | 35   | 30   | 50   | 57   | 50   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| #    | 72        | 81            | 64        | 46   | 105  | 58   | 33   | 22   | 106  | 40   | 106  | 40   | 104  | 25   | 82   | 89   | 15   | 40   | 45   | 37   | 78   | 60   | 100  | 63   | 32   |

# = number of items in this source fitting the SCAL test list

version 22 September 2001
The percentages in Table 2 range from 13% (BZ) to 100% (HQ), and come from fractions with a large range of denominators, all a lot less than the potential 168. For instance, sources A and G have 72 and 33 test list items respectively, but only 16 of these overlap, that is are word pairs with the same meaning. Of these overlapping 16 items, 12 are judged to correspond, giving a correspondence percentage of 12/16=75%, as can be read in Table 2 at the intersection of column A and row G.

From inspection of Table 2 one can see that most of the vocabularies appear to be related in a way that might be described as a ‘dialect web’. Voegelin et al (1963) introduced the term ‘family-like language’; ‘dialect chain’ has also been used. The terms apply to the south-east Western Australia situation of Table 2 in that most varieties (assuming for the moment that each source records a single language variety) can be linked to any other by way of intermediate links and with no link having less than 70% of vocabulary in common. So for instance sources O and U share only 19%, but consider the chain joining them of O-P 79%, P-S 86%, S-Q 80%, Q-T 86%, T-U 88%.

2.3 Procedural difficulties

In carrying out the comparisons behind Table 2, there were procedural difficulties of well-known kinds.

First, there are the usual difficulties in dealing with vocabulary recorded in old sources. The lists vary considerably in their reliability due to the variable expertise of the person writing the language. Some have internally inconsistent spelling conventions which makes it difficult to determine what the writer was trying to record. Further, spellings for a given word can vary as exemplified in the forms on Map 4. These and similar hazards, and methods for dealing with them, are discussed in Thieberger (1995), a suitable manual.

More particular to this project were these areas where decisions and assumptions were necessary:

(a) Multiple synonyms were recorded for some meanings. For example Source B, the Curr Eucla list, has jering ‘sandhill’, wanda ‘beach sandhills’, and both are entered in the comparative table for the meaning ‘sandhill’. Hence Source B is scored as identical to both Source U, which has jirriny alone, and with Source V, which has wantarri alone.

(b) In some instances we need to decide whether forms are independent, wary of covert double counting. The meanings ‘to drink’ and ‘to eat’ are often not expressed independently, for instance, sharing a stem meaning ‘ingest’, distinguished sometimes by compounding with the word for ‘water, liquid’, or the word for ‘food’.

(c) In some instances a form may be considered to be double-headed and so a decision is required as to how to score its correspondences. For instance words for ‘to speak’ may combine a root ‘language ~ mouth’ with a verb stem ‘to make noise’. We have here scored correspondence by reference to the (inflecting?) verb root, and not to the cognate object with which it is compounded.

(d) The ‘homosemantic’ constraint, usually adopted in lexicostatistics, studiously ignores corresponding forms with different meanings, even where they are probable cognates and the meaning shift is plausible. (A more detailed comparison would take account of presumed semantic changes.)
(e) In a few instances a word is identifiable as a recent borrowing. An example is in source H (Bates 2A,4) “Mai ammadharra, boorrborr” ‘hungry’, where the first of the terms is clearly from the Western Desert language, including a distinctive Western Desert language suffix -tjarra.

### 2.4 Significance

When interpreting the varying percentages in Table 2, we have to be keep in mind a sampling error. The test list of 168 items is not large to start with; the sources are missing a proportion of test list data, and any pair of sources overlap in considerably fewer than the potential 168.

Statistical significance has been studied in the literature on lexicostatistics (for example Embleton 1986 and references there). To draw inferences about subgrouping it is preferable to use a list of something like 200 meanings. The comparative vocabulary data available to us is simply not this extensive. Black 1976:84n3 provides a method for “a rapid estimation of the percentage-point differences required for two percentages to be significantly different”. It may be inferred from Black’s table that, for a 100-item list, a 9 percentage point difference is needed for a difference to be significant at the 20% confidence level (and 14 percentage point difference at the 5% confidence level) (for percentages in the range 5%-95%). In the analysis at hand, a 100-item list is about the largest size we have for a pairwise comparison.

Ideally a more detailed philological comparison should be carried out, with careful attention to all analysable detail. As a practical compromise, it is productive to compare the less numerous sources in the light of findings from having compared the more numerous sources, remembering that the oldest records are commonly among the less numerous ones. Usually the short records show affinity with several of the geographically closest later or longer records.

Another aspect of using small vocabulary lists to compare languages is the possible effect on the calculations of conscious ‘marker words’ (including the various terms for localised groups and languages). An informant might express group allegiance in their self-conscious answer to an inquiry as to the word for ‘water’, or ‘Blackfellow’ (a common term in Bates’ lists) for instance, and give an emblematic reply, whereas the word provided for, say, ‘elbow’ (cf. Map 4) is not a conscious marker of group identity (not reported in Aboriginal Australia, at least).

### 3 Maps
Map 3. Words for ‘man’ (or ‘person’) by location
Map 4. Words for 'elbow' by location.
4 Pseudomaps

Since the data of the languages of the study area are almost entirely vocabulary, thereby limiting us to lexical evidence for language classification, we have little choice but to pursue lexicostatistic similarity measures. Such similarity measures among the recorded language varieties (as in Table 2) are spatially interpretable. Indeed a two-dimensional synoptic layout of the various sources presents them in a way in which clusterings of degrees of similarity may be readily perceived like distances on a map.

This approach falls under the statistical method known as multi-dimensional scaling (MDS). Applied in psychology and many other fields, its use in linguistic applications have been few. The MDS approach was demonstrated in detail for analysing a group of fairly closely related languages by Black 1976, and Dyen et al 1992:70–76 applied MDS to Indoeuropean data. Dobson & Black 1979 is the sole published application of MDS to languages in Australia (specifically those of the Cairns rainforest). Woods et al (1986:262–5) is a textbook presentation of MDS for linguists, and Borg & Groenen (1997) is a recent statistics textbook on the subject.

The MDS graphic output, known as configurations, or pseudomaps, can in the case of spatially tagged data be compared with the ‘true’ geographic map of the locational attributes. The comparison is most straightforward when two-dimensional configurations are derived. For languages, or particular language varieties, the known local affiliation of each variety is the obvious two-dimensional spatial attribute. The approach is to apply MDS to some linguistic similarity measure to see what internal relationships are revealed, and whether one of these is the spatial arrangement of the local varieties.

The MDS procedure cannot provide an external orientation for the pseudomaps – the configurations are equivalent under rotation or mirror-image flipping. However when there is some kind of geographic meaning available for each data point, then the configuration can be rotated and flipped so as to make the MDS configuration comparable with the geographic data. The source vocabularies of south-east Western Australia are of this nature, that is all can be geographically located, more or less – some to a particular small named locality, some rather to a district.

The similarity measure employed in this paper is the correspondence percentages of Table 2. We have allowed multiple synonyms for one meaning, partly because the data comes to us this way (especially from Bates) and there is no reason for choosing one synonym above the others. A consequence is that the pairwise percentages between vocabularies may not satisfy the 'triangle inequality', which theoretically can undermine the derivation of pseudomaps. ‘In practice, however, such cases rarely result in any great deviations from triangle inequality, and such deviations could in any case be lessened or removed by some suitable nonlinear transformation of percentage into distance.’ (Black 1976:85n4)

In any case, the potential problem is avoided in that we have used ordinal (non-metric) scaling rather than metric scaling.
To produce two-dimensional scaling pseudomaps, the programs KYST and ALSCAL were chosen because they are the culmination of much refinement of multi-dimensional scaling programs in the 1970s. Both are discussed in recent textbooks. KYST was used by the largest (published?) lexicostatistical project (Dyen et al 1992:72); and the sleeker ALSCAL 84 is in routine use.\footnote{KYST is available from http://www.netlib.org/mds/kyst2a.dos/kyst2a.exe.gz, and ALSCAL 84 Alternating Least Squares Scaling (version 84.1) from ftp://192.41.30.119/pub/visuaftp/ALSCAL/}

Running ALSCAL 84 on the data of Table 2 of all 26 vocabularies produces the diagram or pseudomap of Map 5.\footnote{In the MDS calculations the data points have to jostle each other, as it were, to give an overall compromise configuration which minimises the discrepancy between geometrical distances and the similarity measure of cognate percentages. The degree of mutual squash or squeeze can be summarised in a measure known as ‘stress’. A particular method of calculating stress defines a measure known as $\text{stress}_1$ (Borg & Groenen 1997:33f). For the configuration of Map 5, $\text{stress}_1$ is 0.229, which is on the high side, but not unexpected given the assumptions of the model and error in the data. A ‘scree plot’ (Borg & Groenen 1997:37-38) has $\text{stress}_1=0.129$ in 4 dimensions, $\text{stress}_1=0.153$, in 3 dimensions, and $\text{stress}_1=0.382$ in 1 dimension.}

Map 5. Pseudomap of 26 vocabularies.

Sources are placed on the pseudomap so that the distances between any one source and all the others represent the overall relationship of similarity between that source and all others. As Dyen et al 1992:73 put it, “there is a strong inverse relationship between lexicostatistical percentage and distance. If the percentage is large the distance is small; and if the percentage is small the distance is large.” The pseudomap does not aim
to show the geographical locations of the sources. The ‘dialect web’
nature of the relationships deducible from Table 2 are readily seen in Map
5: every source can be linked to any other by way of short intermediate
links, though the extremes are some distance apart.

Note that the reduction of the potential 25 dimensions to two dimensions
means that in interpreting the pseudo-maps pairwise distance on the
pseudo-map is not uniformly proportional to linguistic similarity, but the
“best fit” aims for a good correlation between pairwise distance on the
pseudo-map and linguistic similarity. For example, in Map 5 source D is
closer to M than it is to T, while the distance (on the pseudomap) between
D and M is roughly the same as between T and V. Compare the percentages
in Table 2: D-M 51% is greater than D-T 44%, but T-V 95% is even closer
than D-M. The reason that the distance D-M is much the same on the
pseudomap as the distance T-V is that the relationship with all the other
sources is also brought to bear in the formation of the configuration of
the pseudomap: despite being almost the same (having 95% in common with
each other), T and V relate somewhat differently to the other sources, for
instance T-E 65% versus V-E 58%.

Extremes of mismatch in correspondence between percentages of Table 2 and
pseudomap distances in Map 5 are exemplified by:
• C-T (54%) which are separated by about seven times the distance
separating D-M (51%, much the same percentage);
• K-T (57%) which are separated by about eight times the distance
separating K-Z (57%, the same percentage).
The rationale for these mismatches comes from seeing them as extremes in
the general distribution of the matches between the 325 (=26*25/2) inter-
source percentages and inter-source pseudo-distances. For instance, K-T
needs to be more separate than K-Z to accommodate the way that K is
generally less similar to other sources with higher percentages in common
with T (A, B, G, I, Q, U, V, W, all >80%) and more similar to other
sources with higher percentages in common with Z (F, Q, Y, all > 70%).

The relative configuration of vocabulary similarity matches reasonably
well with the known geographic source of each record: compare Map 5 with
Map 8 or Map 2 (via the codes in Table 1 or Table 3). The striking
mismatches with geographic locations are these:
1. The most striking mismatch concerns source L. L is embedded in the
configuration in Map 5 but is actually from the Ravensthorpe district,
west of Esperance and well away from all other source locations. L
being very much a Marlba language, we cannot explain why Bates
associates it with Ravensthorpe (well within the Nyoongar area).
2. Source R was collected by Tindale in Norseman and said to be from
Israelite Bay, which is on the coast south-east of Norseman; yet Map 5
places R even further away from sources along the Bight coast (A, B, F,

To try to reduce the margin of error somewhat, ALSCAL 84 was then run on
the data, extracted from Table 2, of those 23 vocabularies with more than
thirty SCAL items, producing the diagram of Map 6.\(^8\) This entails

\(^8\) Stress\(_1\) is 0.182. A ‘scree plot’ has stress\(_1\)=0.101 in 4 dimensions
stress\(_2\)=0.128 in 3 dimensions, rising to stress\(_1\)=0.362 in 1 dimension.
This pattern points to two being the appropriate dimensionality, and to
the appropriateness of excluding sources H, N and Q.
discarding sources H, N and Q, which, coincidentally, include two of the three sources which we know are a mix of language varieties (see the ‘ind?’ column in Table 3).

Map 6. Pseudomap of 23 vocabularies (Map 5 less H, N, Q)

Map 6 still resembles a dialect web, with three clusters readily discernible, in the top-left, lower-left, and mid-right of Map 6 (normalised to correspond to the north-west, south-west and east of the study area).

4.1 Three languages?

The compelling question asks what is going on in the clustering apparent in Map 6, what are its determining factors? A three-way distinction across the study area is the label for a language variety according to its word for ‘man’ or ‘person’ (usually the same word has both meanings).

The sources are listed again in Table 3 with their word for ‘man’ or ‘person’ in the right-hand column. If the term is capitalised, then it has been supplied as a name of the language or people. The left-hand columns in Table 3 are the identification as in Table 1, including the letter key to the two pseudomaps Map 5 and Map 6. The middle column indicates (where we know from the provenance) whether the source is a single variety, or probably a compilation from several individuals and a mix of language varieties.

9 The complexities of language naming in Australia mean that commonly used names should not necessarily be taken to reflect linguistic similarity; (see McConvell’s paper on ethnonyms, this vol.)
Table 3. Source vocabularies and 'person' word

<table>
<thead>
<tr>
<th>code</th>
<th>source</th>
<th>ind?</th>
<th>'person'</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Curr 34</td>
<td>?</td>
<td>meening</td>
</tr>
<tr>
<td>B</td>
<td>Curr 35</td>
<td>?</td>
<td>majilba, meening</td>
</tr>
<tr>
<td>C</td>
<td>Helms</td>
<td>N</td>
<td>minnin, bulga, malba</td>
</tr>
<tr>
<td>D</td>
<td>Garnier</td>
<td>Y</td>
<td>kaiano</td>
</tr>
<tr>
<td>E</td>
<td>2A,1a</td>
<td>Y</td>
<td>Mining</td>
</tr>
<tr>
<td>F</td>
<td>2A,1b</td>
<td>Y</td>
<td>Mulba</td>
</tr>
<tr>
<td>G</td>
<td>2A,3</td>
<td>Y</td>
<td>Mining</td>
</tr>
<tr>
<td>H</td>
<td>2A,4</td>
<td>N</td>
<td>Mulba, mining</td>
</tr>
<tr>
<td>I</td>
<td>2A,5(1)</td>
<td>Y</td>
<td>Mining</td>
</tr>
<tr>
<td>J</td>
<td>2A,6(1)</td>
<td>Y</td>
<td>Mining</td>
</tr>
<tr>
<td>K</td>
<td>2A,8</td>
<td>Y</td>
<td>Mulba, Baaduk people</td>
</tr>
<tr>
<td>L</td>
<td>2B,2</td>
<td>Y</td>
<td>Mulba</td>
</tr>
<tr>
<td>M</td>
<td>C,1</td>
<td>Y</td>
<td>Kabboon</td>
</tr>
<tr>
<td>N</td>
<td>C,2</td>
<td>Y</td>
<td>Marrbal</td>
</tr>
<tr>
<td>O</td>
<td>C,3</td>
<td>Y</td>
<td>Kabboorn</td>
</tr>
<tr>
<td>P</td>
<td>C,4</td>
<td>Y</td>
<td>Kabboorn, kabboon</td>
</tr>
<tr>
<td>Q</td>
<td>C,5 Harvey</td>
<td>N</td>
<td>Malbert tribe, bulgoo, fahdook</td>
</tr>
<tr>
<td>R</td>
<td>T86</td>
<td>Y</td>
<td>marlba, malba:</td>
</tr>
<tr>
<td>S</td>
<td>T89</td>
<td>Y</td>
<td>kabun -&gt; kabulu</td>
</tr>
<tr>
<td>T</td>
<td>T83</td>
<td>Y</td>
<td>mining</td>
</tr>
<tr>
<td>U</td>
<td>GNOG S</td>
<td>Y</td>
<td>mirminy</td>
</tr>
<tr>
<td>V</td>
<td>GNOG J</td>
<td>Y</td>
<td>mirminy</td>
</tr>
<tr>
<td>W</td>
<td>GNOG P</td>
<td>Y</td>
<td>mirminy</td>
</tr>
<tr>
<td>X</td>
<td>Hale</td>
<td>Y</td>
<td>kapun</td>
</tr>
<tr>
<td>Y</td>
<td>CGvB RG</td>
<td>Y</td>
<td>marlba</td>
</tr>
<tr>
<td>Z</td>
<td>CGvB PF</td>
<td>Y</td>
<td>marlba, nganda(g)a (FR)</td>
</tr>
<tr>
<td></td>
<td>GNOG GAL</td>
<td></td>
<td>marlba [ms. buljgu malba]</td>
</tr>
</tbody>
</table>

Map 7 is the configuration of Map 5 annotated with this three-way distinction.
There is an obvious correlation between the ‘person’ word (and variety name) and relatively high percentage of vocabulary correspondences, as displayed in Map 7 (compare Map 3). The correlation is enhanced in the light of the following comments about sources marginal to the clusters:

- Source Q (compiled by the Norseman doctor) could well be a mixture, and in any case is not numerous.
- Source H has both marlba and mirniny as ‘person’ words, and is unusual for a Bates vocabulary in being assigned merely to a district (Eucla) and to no named people.
- Source C is said by its compiler (Helms) to be a mixture, and has both marlba and mirniny as ‘person’ words.
- Source E has mirniny as the ‘person’ word, but is from a location well to the west of other Mirniny locations, and closer to Marlba locations; as is clear on Map 3.

In short, in the study area the ‘person/man’ language names, or labels of language distinctions, correspond well to the three observed vocabulary clusters.
Map 8. Locations of main vocabulary sources
4.2 Interpreting the results of pseudomapping

As noted earlier, the relative configuration of vocabulary similarity matches reasonably well with the known geographic source of each record. The extent to which there is mismatch is of further interest. Mismatch could be attributed to a variety of influences. First, there could be bias in the vocabulary attributable to their manner of collection and preservation. Second, there are factors inherent to the language situation. In each individual’s interview responses there could well be a conscious ‘clustering to the norm’ of one of the named languages of the study area.

A third, and perhaps most interesting, possibility is that the vocabulary similarity measure is tapping into older patterns of relationship between the language varieties. Detectable diffusion (borrowing) of vocabulary and other linguistics traits generally proceeds from neighbour to neighbour, and the more recent a borrowing the more readily detectable it generally is; thus similarities due to borrowing and diffusion are expected to present a spatial cline. Relationships between languages dating from an earlier period, when spatial arrangements may have been different from in historical times, can be expected to leave a trace -- the genetic component of the shape of a language, where the vocabulary and other properties of daughter languages inherited from ancient ancestor languages. Note that the use of correspondence percentages, as in this paper, which include correspondences due to borrowing as well as inheritance, suits the data somewhat more to spatially-interpreted MDS (as noted by Dobson & Black 1979:58ff). Thus it is a virtue rather than a drawback that the figures of Table 2 probably include sharings due to borrowing.

Given the general persistence of particular languages short of catastrophes, that is, that a community’s language does not (indeed, hardly could) change quickly (within a generation, say), we would expect the historical record of match between location and language variety also to approximate well the situation for some generations prior to the period recorded (that is prior to 1880-1970).

5 Conclusion

The various historical vocabularies are records of a ‘dialect web’ or ‘family-like language’ with three discernible clusterings. The three clusters correspond fairly well with the word for ‘person’ in each vocabulary, words which have been in long use as language names, namely Mirniny, Marlba, and Kabun.

Further, the spatial relationships among the vocabularies are comparable across more than a century of records, and hence allow a view of locality of linguistic knowledge. The results are evidence for continuity, in that the relative configuration of vocabulary similarity matches reasonably well with the known geographic source of each record. The span of implied continuity covers the generations of the historical records, and, by reasonable extrapolation, some generations prior, approximating thus the time span relevant to demonstration of continued native title.
I see potential in extending the quantified approach of this paper (test-list percentages, pseudomapping) to more recent lexical data, bringing present-day linguistic knowledge into view. It may also prove fruitful to integrate into the comparison various grammatical elements (for instance, some nominal suffixes attested in several sources), which are known to persist as indicators of particular linguistic lineages.

6 References

A. Vocabulary sources (Table 1)


B. Other references

ALSCAL 84 Alternating Least Squares Scaling (version 84.1) ftp://192.41.30.119/pub/visuaftp/ALSCAL/

Handbook of Western Australian Aboriginal languages south of the Kimberley Region, (1993 by Pacific Linguistics, Australian National University, Web version © Nicholas Thieberger 1996).

version 22 September 2001


O'Grady, G.N. 1959. Significance of the circumcision boundary in Western Australia, BA thesis, University of Sydney.


O'Grady, G.N., S.A. Wurm, & K.L. Hale. 1966b. *Aboriginal languages of Australia (a preliminary classification)*. [map 69 x 86 cm ~ 76 x 94cm.] Scale est. 1:5,243,586 ~ 5,068,800. Victoria, B.C.: University of Victoria.


version 22 September 2001