Published documents relating to the language of Aboriginal people in the Milparinka area.

Members of the Momba tribe. Frederic Bonney photograph

Compiled by the Milparinka Heritage and Tourism Association 2023

Content:

9

THE YARLI LANGUAGES

Luise A. Hercus and Peter K. Austin
Australian National University and SOAS

1. Introduction

The Malyangapa language, traditionally spoken in far north-western New South Wales, has been classified into various subgroups of Australian Aboriginal languages, including the Karnic languages of the Lake Eyre Basin. Using all the available data on this language we consider previous classifications and regard Malyangapa as part of a small subgroup of languages, the Yarli subgroup, once spoken in the far north-west corner of New South Wales and adjacent areas in South Australia and Queensland.

The words quoted from Malyangapa, Yardliyawara, Diyari and Paakantyi are from our own transcriptions. Words from other languages are spelt according to the relevant standard works, Adnyamathanha according to Schebeck (2000a), Wangkumara and Yandruwantha according to Breen (forthcoming and n.d.). The only changes that have been made are in the notation of the rhotics.[1]

1.1 Languages

The three languages in the proposed Yarli subgroup are:

Malyangapa — recorded by Stephen Wurm in 1957 with Hannah Quayle, born near Yancannia in about 1875, and with George Dutton and Alf Barlow. Peter Austin has made a detailed study of these data (Austin 1986). Luise Hercus did some recording with George Dutton in the mid-1960s on Malyangapa; Jeremy Beckett had previously worked with him on social and cultural traditions. Luise Hercus also recorded Laurie Quayle, son of Hannah

[1] Notation of rhotics: *r* = alveolar tap, *rr* = alveolar trill, *R* = retroflex glide. Abbreviations used are: ACC, accusative; ALL, allative; CAUS, causal; ERG, ergative; FUT, future tense; IMPER, imperative; INCH, inchoative; LOC, locative; NOM, nominative; PAST, past tense; PRES, present tense; PURP, purposive.

Quayle, checking some of the earlier materials. He passed away in 1976, and with his death the language became extinct.

Wadikali — known from a 72 word vocabulary in Tindale's 1934 Diamantina notebook, taken down from Ned Palpilina 'Blanche Ned', who was said to be the last Wadikali. His country was Yandama Creek, but his mother had left there just before he was born and he had spent much of his life at Blanchwater in Pirlatapa country (Hercus 1987, Hercus and Koch 1996) There are descendants of Wadikali people, but the language has evidently not been spoken since the 1930s.

The name Wadikali might suggest that the language belongs to the Paakantyi subgroup as there are similarly formed names for Paakantyi people with a term *kali,* which is said to be an archaic word meaning "people". Hence we have the names Wilyakali, Thangkakali, Bula-ali and Pantyikali "the Creek people" whose language was called Wanyiwalku. The use of the term *kali* is however by no means confined to Paakantyi: Malyangapa people called the Adnyamathanha, i.e. "the Stone People" of the Flinders Ranges, by the term *Yarnda-ali,* which is simply a translation and also means "the Stone People". As the linguistic evidence in §4.1. below will show, Wadikali is not a Paakantyi language: it is clearly part of the closely-knit subgroup of Yarli languages.

Yardliyawara — based on limited material from two speakers, Barney Coffin, recorded by Bernhard Schebeck and Luise Hercus, and Fred Johnson, with whom Bernhard Schebeck did some recording of vocabulary and short sentences (Schebeck 1987). Bernhard Schebeck has very kindly made his data available to us. More work on the Coffin recordings is in progress.

Three vocabularies in Curr (1886-87) belong to the area:

1. No. 69 Evelyn Creek, by A. Dewhurst, Esq, Curr (1886-87 II:156):

Dewhurst and Crozier, as quoted (II:152), informed Curr that this area belonged to the 'Pono' people, but owing to the influx of other people there was "a great mixture of dialects". Curr goes on to speculate that the "Pono Blacks belonged to the Cooper's Creek (i.e. Wangkumara) rather than to the Darling Tribes". This may well be so, as it seems likely that 'Pono' is a transcription of *purnu*, which in Wangkumara means "country". The vocabulary written down by Dewhurst contains some admixture of Wangkumara but is mainly Malyangapa; e.g. "fire" is *wiyi* versus *'kal:'a'* in Wadikali*, kardla* in Yardliyawara; "beard" is *ngankuru* as opposed to *nganku* in Wadikali and Yardliyawara. There does, however, also seem to be some influence of Paakantyi; e.g. *yimba* for "you", cf. Paakantyi (*ng)imba*.

2. No. 69 Evelyn Creek, by H. Crozier, Esq, Curr (1886-87 II:154)
This is probably Wadikali with some admixture of Wangkumara and Pirlatapa.
3. No. 69a Near the North-west Corner of New South Wales, by A.W. Morton, Esq, Curr (1886-87 II:160)

This seems to be mainly Wadikali, though the introduction speaks of 'Mulya napa' people living in the area. All the available evidence, and especially that of Tindale, points to the extreme northwest of New South Wales being Wadikali country.

1.2 Areas and locations

Like many other language-owning groups, the people of this region were divided up into a number of local clans. In view of the general disruption and depopulation during the course of the nineteenth century, information on this has been lost. The area was particularly vulnerable on account of the discovery of gold in the Milparinka-Tibooburra area. Police and pastoralists sometimes refer to groups of people by names that are otherwise unknown, such as 'Pono' quoted above (§1.1.). It is possible that these references are to small local groups, which were displaced in the wake of the first European settlement. There is however rough general agreement among all the sources as to the area originally occupied by speakers of Yarli languages. The most important of these sources is Beckett's published and unpublished work with George Dutton in 1957-58:

> 1. George Dutton spoke of Malyangapa people being at Salisbury, Cobham, Yantara Lakes, Mt Pool and Mt Arrowsmith. The name was written as 'Milya-uppa' by Reid in Curr II:180. Reid's 'Milya-uppa' vocabulary, from Torrowotto is however not Malyangapa at all, but straight Paakantyi. This may well be due to the displacement of people, which resulted in there being a mixed population at Torrowotto, some Paakantyi, some Malyangapa. Wurm's main consultant, Hannah Quayle, placed Malyangapa country very much as George Dutton had, as "Tibooburra, Salisbury Downs and Milparinka."
>
> 2. George Dutton spoke of "Wadikali, like Malyangapa (i.e. it is close to Malyangapa), go from Mt Pool, Mt Sturt, Yandama, Tilcha from there to Lake Frome."
>
> 3. Yardliyawara was spoken on the eastern side of the Flinders Ranges, and Adnyamathanha people referred to it as 'Wooltana talk' (Wooltana being the name of a station on the north-eastern side of the Flinders).

There has been some confusion regarding the location of Wadikali people from Tilcha to Lake Frome. This was caused by a statement in the work of R.H. Mathews (1898:242): "At Lake Boolka and Tilcha are the Endawarra and Berluppa people respectively." Mathews was basing himself on information from correspondents, including letters from the police sergeant B. Hynes from Tibooburra in 1897-98. Hynes wrote 28.5.1898: "the Tilcha Blacks are called Berluppa". In a later communication 18.8.1898, too late to be used by Mathews in his article, he wrote: "Tilcha is now only a back station of Yandama and I believe there are no blacks there at present." Hynes was passing on information he had been given by R.B. Daws, the manager of Tilcha: he was talking about the state of affairs at that particular time, not about the ancestral homeland of particular groups of people. The homeland of particular groups is precisely what George Dutton was speaking about.

'Berluppa' or 'Biraliba' are variant spellings for the Pirlatapa, who were not linguistically associated with the Yarli group but were closely akin to Diyari (Austin 1990b). There may well have been a group of them visiting Tilcha. The Pirlatapa were the immediate neighbours of the Wadikali and the Yardliyawara, as indicated by George Dutton and confirmed by all the other available evidence. They were strongly associated with the Blanchwater area, as is clear from oral evidence from South Australia (Hercus and Koch 1996); and according to George Dutton (Beckett 1958) they were at "Callabonna, Quinyambie Station and through to Lake Elder, Congie Bore and Cooney Bore."

The 'Endawarra', who are mentioned by Mathews as being at Lake Boolka, about 30 km south of Tilcha, were even further from their country: Endawarra is a spelling for Yandruwantha. We know from Tindale's evidence (1934) that the Yandruwantha had joint initiation ceremonies with Wadikali people; so this too probably refers to a temporary situation. See Tindale's 1940 map for further details.

There is a major change in Tindale's maps between 1940 and 1974 for this area. Into what was on his earlier map Yardliyawara and Wadikali country, Tindale has inserted another group, Ngurunta. This name is known also from Curr 1886-87 (II:180): "The tribes which bound the Milya-uppa are the Ngurunta on the west, the Momba on the south...."

In connection with Ngurunta Tindale (1974:216) also mentions the anonymous and very fragmentary vocabulary in Curr 1886-87 (II:173) with the vague title 'Country north-west of the Barrier Range'. There is however no

indication that this brief vocabulary belongs to Ngurunta or any of the Yarli subgroup; every single word in it is Paakantyi and Curr himself mentions this: "The following words, contributed anonymously, some of which correspond with those of the Common (Paakantyi) vocabulary, show that the tribe which uses them is of Darling descent."

None of the senior people in South Australia and on the NSW side in the 1960s ever mentioned the Ngurunta, and this includes Barney Coffin, who travelled frequently between the two states. Tindale's information, however, does seem to be from a person interviewed by him in the 1960s, so memory of the group as an entity had survived in this limited way. The area in question, which is mainly inhospitable sandhill country, was generally regarded by these senior people in NSW as being part of Yardliyawara. See Tindale's 1974 map for further details.

1.3 Culture

As regards social organisation the group is uniform in having a matrilineal moiety system. Yet, as elsewhere, belonging to the same linguistic subgroup does not necessarily imply social and cultural uniformity. All the three sets of people– Malyangapa, Wadikali and Yardliyawara– were circumcising and, along with their westerly neighbours, they had a form of the Wilyaru secondary initiation ritual (see Beckett 1967). Nevertheless it seems that the three groups did not perform joint ceremonies but joined in with their respective neighbours. Wadikali and Malyangapa joined in with what was called 'Milia', a circumcision ceremony and myth shared with Wangkumara/Kungardutyi people and centred on Cobham Lake in Malyangapa country. Wadikali people also shared in Yandruwantha initiation ceremonies, according to the entry mentioned above by Tindale in his Diamantina notebook (1934). Yardliyawara people joined in ceremonies with the Adnyamathanha.

There are numerous myths and song cycles traversing the whole area. Some were shared by all, along with Paakantyi people, such as the story of the Two Snakes from the Paroo who travelled all the way to the Paralana Hot Springs in Yardliyawara country (Beckett 1958). The *Kurlimuku* song cycle was also shared widely, as Barney Coffin pointed out to us, "Four nations sings the same song. Malyangapa, Wadikali and Kungardutyi and Wanyiwalku, that is four nations."

The people speaking Yarli languages clearly remained associated with one another, but each had cultural associations, involving intermarriage, with

outside groups as well. In the case of the Malyangapa it was especially with the Paakantyi group Wanyiwalku/Pantyikali; in the case of the Wadikali it was with the Karnic speaking Kungardutyi /Wangkumara and Pirlatapa; in the case of the of the Yardliyawara it was with the Pirlatapa and the Thura-Yura speaking Adnyamathanha.

1.4 Genetic unity versus diffusion

Yardliyawara and Malyangapa are so close to one another, and what we know of Wadikali is also so close, that Proto-Yarli is more or less self-evident. The differences between the languages are largely due to outside factors. There is evidence for linguistic characteristics cutting across this whole area and apparently arising from borrowing and diffusion. For example, Malyangapa and Wadikali show phonetic lengthening of single consonants at the beginning of the second syllable following the initial stressed syllable. This feature is shared with Paakantyi to the east, and with Karnic. In Yardliyawara laterals and sporadically also the nasal *n* have become prestopped in this position (thus compare Malyangapa *yarli* "person" with *yardli* in Yardliyawara), a feature shared with neighbouring Karnic languages and Adnyamathanha.

Furthermore, the Yarli languages show bound pronouns for subject and object suffixed to the verb, a feature shared with both Paakantyi and Adnyamathanha. Bound pronouns are not found in Karnic.

Finally, there are lexical items which are distributed according to these regional diffusion patterns. Two examples from the vocabulary in Appendix 9 are quoted in Table 1 to show the intricacy of these diffusion patterns.

Table 1: *Lexical items showing regional diffusion patterns*

	kangaroo	kangaroo	bird	bird	bird
Adnyamathanha	*urdlu*		*yirta*		
other Thura-Yura	*kurdlu* PNK		*thirta* KUY		
Yardliyawara	*kurdlu*		*thirta*		
Wadikali	*'kol:o*	*talda* (Morton)		*ju:li*	
Malyangapa		*tharlta*		*yurli*	
Wangkumara		*thaldra*			*maranga*
Paakantyi		*tharlta*			
Paakantyi dialect	*kurlu* Wilyakali			*yurli* duck Pantyikali	

Adnyamathanha had the closest geographical and social ties with Yardliyawara: other Thura-Yura languages like Parnkalla and Kuyani were

further away. It therefore appears that these words had spread to Yardliyawara from Adnyamathanha before the occurrence of lenition of initials in Adnyamathanha.

The first example also shows the spread of pre-stopping. The second example shows a word, *yurli* "bird", which appears to have been a joint innovation of the Yarli languages. This appears to have been subsequently lost from Yardliyawara. Some examples of morphological diffusion are discussed in §5.2. There is also a possible layer of recent borrowings from Adnyamathanha into Yardliyawara in our data, since both the speakers who survived to be recorded were also speakers of Adnyamathanha.

2. History of classification

Over the past 84 years, that is from the time of Schmidt (1919a) on, the Yarli languages have been classified into a number of linguistic subgroups:

1. with the neighbouring Karnic languages, that is with Wangkumara to the north and Pirlatapa the north-west
2. with the Thura-Yura languages to the south-west
3. with the Paakantyi or Darling River subgroup to the east.

As indicated above, in many ways this area of western New South Wales and north-eastern South Australia is marked by cultural and linguistic diffusion and shows evidence of phonological and morphological features shared across genetic subgroups. Despite this we are able to isolate characteristics of Malyangapa that it shares with those neighbouring languages with which we propose it forms a genetic subgroup, namely Wadikali and Yardliyawara. We call this the 'Yarli subgroup'.

2.1 Schmidt

The first published classification of Yarli languages using lexical data was by Schmidt (1919a), who calls the language of the area 'Evelyn Creek language', basing himself on vocabularies by Dewhurst, Crozier and by Morton in Curr (1886-87 II). Unfortunately, Dewhurst's vocabulary appears to be mixed, with some influence from Wangkumara and other Karnic languages. Schmidt had available to him only those Curr vocabularies and no morphological data; yet he was sufficiently impressed with the special features of 'the Evelyn Creek

language' to make it a special subgroup of a big group which included the Karnic and the Thura-Yura languages. (See further the map in Schmidt 1919a.)

2.2 Tindale

Tindale thought of Malyangapa and 'Wanjiwalku', a Paakantyi language, as being 'the same'. In discussing 'Wanjiwalku' (1974:200) he states: "both this group and the Maljangapa speak one language (Wanjiwalku)"– an idea that may well have stemmed from the fact that his main informant, George Dutton, was a fluent speaker of both these languages. Tindale did a lot of very intensive language work with George Dutton: he transcribed two long myths and compiled a special separate notebook on Wanyiwalku grammar. There are no linguistic data on Malyangapa in Tindale's work, and it would seem that he had no means of comparing the two languages.

There is not much similarity between any Paakantyi language and Malyangapa, as will be shown in §4.2. below.

2.3 O'Grady, Voegelin and Voegelin

To the north and north-west the Yarli languages are bordered by Karnic languages: for general discussion of the Karnic subgroup see Austin (1990a), Bowern (1998, 2001c). In O'Grady, Voegelin and Voegelin (1966:123) the 'Yalyi subgroup' consists of Karenggapa, Malyangapa and Wadikali, while Yardliyawara is placed in the Yura subgroup of south-west Pama-Nyungan.

The name Karenggapa requires some explanation. The first mention of the name Karenggapa is by J.A. Reid in Curr 1886-87 II:180: "The tribes which bound the Milya-uppa are ... those of the Paroo to the east and the Karengappa on the north." Karenggapa is mentioned by Tindale both in his 1940 work and in 1974:193 as the name of people around Mt Bygrave and the southernmost part of Bulloo Downs. It must have been a small local group, as none of the senior people recorded in the late 1950s and the 1960s had any recollection of the Karenggapa. This included elders who had memories reaching back to the last decades of the nineteenth century. The name Karenggapa has survived in the area only as the name of an old tank at the southern end of the Carryapundy Swamp. There is massive evidence from place names and statements in the mythology that the people originally living around Mt Bygrave and the southernmost part of Bulloo Downs, the area associated with the Karenggapa by Tindale, were speakers of a form of Wangkumara (Hercus 2001). The wordlist quoted by Tindale as belonging to the Karenggapa is by J.A. Reid

from Torowotto Swamp, much further to the south, and is, as stated above, entirely in Paakantyi. There is thus no evidence whatsoever to associate a 'Karenggapa' language with the Yarli subgroup.

The 1966 map *Aboriginal Languages of Australia: a preliminary classification* by O'Grady, Wurm and Hale follows this same classification and has Wadikali, Malyangapa and Karenggapa forming a subgroup.

2.4 Wurm

Wurm (1972:133) has a 'Yalyi' subgroup of the Dieric group. This subgroup consists of 'Nadikali' (presumably Wadikali) and Malyangapa. Karenggapa is no longer mentioned, but 'Yadliyawara' is still in the Yura subgroup of the southwest or Nyungic Group.

Walsh and Wurm (1982) have a Yarli subgroup (Wadikali and Malyangapa) of the Karnic languages. Yardliyawara has been reclassified into the Karna/Diyari group.

2.5 Dixon

Dixon (2001: maps on pages 72, 76, 94 and 96) also classifies 'WAd' as part of Karnic. 'WAd' is described as consisting of 'Maljangapa, Yardliyawara and Wardikali', but on those maps it appears as if Yardliyawara were not included.

Bowern (1998 and 2001c:255) has already given excellent reasons why there does not seen to be any close link between the Yarli languages and Karnic. These and other reasons will be discussed in §4.3. below.

3. Lexical evidence for the subgroup

3.1 The lexical distinctiveness of Yarli

Lexical comparisons of Wadikali, Yardliyawara and Malyangapa are difficult, given the limited amount of data we have, particularly on Wadikali. Nevertheless, even a cursory survey of the available materials shows that the three are lexically very close and not particularly closely related to the nearest Karnic language, Wangkumara. A comparative vocabulary illustrating this appears in Appendix 9, along with a commentary further substantiating the evidence.

Most of the similarities between the three languages represent innovations in the Yarli languages; some, however, are joint retentions. There are a number

of lexemes that are shared by all three Yarli languages and by no other languages in the area. These include such basic words as: “be hungry”, “dog”, “go”, “good”, “little”, “moon”, “speak”, “stick (n)”. These all appear to be lexical innovations of the Yarli subgroup.

Secondly, there are lexemes shared by two Yarli languages and no others in the area, where the third Yarli language is simply undocumented or has that word replaced by a loan from a neighbouring language. Examples are the words for “euro”, “leg”, “uncle”, “see”, “no”. These also appear to be lexical innovations of the Yarli subgroup.

Some lexemes are only found in all three Yarli or only two Yarli languages plus a neighbouring language, where there was probably borrowing out of Yarli; e.g.. “bite”, “ground”. These also appear to be lexical innovations of the Yarli subgroup.

There are also lexemes inherited from Proto-Pama-Nyungan (pPN) but not found in those particular forms in neighbouring languages. These include “eye” and “lie down”. Although these are by no means innovations, they differentiate the Yarli languages from their neighbours.

Some lexemes are found in the Yarli languages only, but they have regional cognates involving major differences in form and/or meaning; e.g. the words for “arm (upper)”, “bring”, “emu”. Thus *pardu* “bring” is cognate with a verb meaning “hold” in a number of Karnic languages, including Yandruwantha *pardra*. The Yarli word *kalarti* “emu” differs from but does have some resemblance to Paakantyi *kalthi* and Kaurna *kari* “emu” and even Western Desert *karlaya* could be a distant cognate. These differences, however, are sufficiently significant to distinguish the Yarli words from those in neighbouring languages.

Finally, there are a number of basic items of vocabulary which are shared with neighbouring languages and which are inherited from proto or regional Pama-Nyungan; e.g. some body-parts, “to eat”, “give”, “food” and “possum”. Their presence in the Yarli languages shows resemblance to neighbouring subgroups, but not adherence to one or the other, because all those subgroups have them.

3.2 Lexical differences from Karnic

To the north and north-west the Yarli languages are bordered by Karnic languages: for general discussion of these see Austin (1990a), Bowern (1998 and 2001c). In the vocabulary listed in Appendix 9 we have mainly considered

Wangkumara, as being geographically and socially the nearest Karnic language. Bowern shows that the Yarli languages do not form part of Karnic. One of her arguments is based on the lexical comparison of all the Karnic languages. She gives (2001c:250) the following lexical cognate percentages, given here as Table 2, for Malyangapa in relation to a widespread number of members of the Karnic group. These numbers are sufficiently low to go towards proving her point.

Table 2: *Lexicostatistical percentages between Malyangapa and Karnic languages*

Language	Percentage
Pitta-Pitta	16
Arabana-Wangkangurru	29
Mithaka	30
Yaluyandi	21
Ngamini	21
Diyari	33
Yandruwantha	22
Wangkumara	35

4. Morphological evidence for the subgroup

4.1 Internal comparisons

4.1.1 Wadikali

Morphological comparisons within the Yarli subgroup are difficult because of the limited data from Wadikali. Tindale's materials do, however, include a couple of entries that show parallels to Yardliyawara and Malyangapa structures.

(1) The entry "flat ground" has *wankanga 'pakaita* which is almost certainly:
wanka-nga *paka-yitha*
meat-LOC go-PURP
"to go for meat"

(2) The entry "breast" has *'min:ami'teita* which could represent
minha *mitya-yitha*
what suck-PURP
"something for sucking"

(3) The entry "salt lake" has the words *pak:uta pakanu* crossed out but this almost certainly represents:
paku-tha paka-nu
lake-ALL go-PURP
"to go to the lake"

All of the bound morphemes in these phrases, *-yitha* "purposive", *-nu* "purposive", *-nga* "locative" and *-tha* "dative, allative" have identical parallel forms in Yardliyawara and Malyangapa. There can be no doubt that we are dealing with a single group of languages here.

4.1.2 Yardliyawara and Malyangapa

The recordings of Yardliyawara so far studied show no appreciable morphological differences from Malyangapa. There are only minor divergences, and two examples of these are given here.

The first example is that of special time-marking. There is no sign in Yardliyawara of the special time-marking verbal suffixes for morning and evening: this seems to be confined to Malyangapa. Those morning and evening forms were not used by the Yardliyawara speaker, and even one Malyangapa speaker was heard to use a noun "in the morning" instead of expressing time as part of the verb *wanirithu miRinga* "I leave (you people) in the morning". It seems highly likely that this specialised paradigm for time was used only in Malyangapa and probably in Wadikali, but we have no means of knowing for certain about Wadikali. One thing is clear: it was not a genetic but a regional diffusional feature. The actual forms used were not diffused, but the grammatical category was. We can deduce this from very important but as yet unpublished material by Breen from Yandruwantha (Breen forthcoming:§1.11). In this language there are verbal affixes referring to the time of day, including *-thalkana* meaning "early in the morning", based on *thalka* "upward", and *-yukara* meaning "at night", based on a verb "lie down'. The Nhirrpi dialect of Yandruwantha, recorded by Wurm and studied by Bowern (1999b:§4.4.2) shows those same features. It is from the Nappa Merrie area, very close geographically to Wangkumara. Similarly Wangkumara has a suffix *-pa* which refers to action in the morning or action upward, and a suffix *-waga*, which refers to action at night and is based on a verb meaning "sleep" (Breen n.d.).

The corresponding Malyangapa suffixes were formed from the actual word for "morning", while the origin of the "night" suffix is not so clear.

Special suffixes for action in the morning and evening have not been recorded for Karnic languages other than Yandruwantha and Wangkumara; so it looks like a purely regional phenomenon. It probably originated in Yandruwantha because this language has the most developed system of this kind. It has special forms, apart from those already quoted, for "in the morning (not quite so early)", "during the day", "within the last hour or two" and so forth. The diffusion of this feature into Malyangapa and possibly Wadikali but not Yardliyawara clearly does not impinge on the Yarli languages being regarded as a unity.

The second example concerns a verbal form not shared between the three Yarli languages. There seems to be in Yardliyawara a past causal participle *-utu,* which does not appear in the Malyangapa data, as for instance in the Yardliyawara sentence in (4).

(4) *Wanka iniki-nha wanyu-r-utu, pulkata!*
meat that-NOM bad-INCH-CAUS throw away IMPER
"Throw that meat away because it's gone bad." (Barney Coffin)

4.2 Grammatical differences from Paakantyi

There are very good reasons for recognising that, while there are some cognates and some lexical borrowings from Paakantyi, such as *kumpaka* "woman, wife", *yartu* "wind" *wanka* "meat", the grammatical systems of the two language groups are quite distinct. This is evident from the pronouns, both free and bound, as shown in Table 3. The Malyangapa singular forms are given here, with any divergences in Yardliyawara being noted.

Table 3: *Malyangapa and Paakantyi pronouns*

	Malyangapa		Paakantyi	
1sg ERG	*ngathu*	*-thu*	*ngathu*	*-thu*
1sg NOM	*nganyi*	*-nyi*	*ngapa*	*-apa*
1sg ACC	*nganyinha*	*-nyi* (Yard. also *-ayi*)	*ngayi, nganha*	*-ayi, -anha*
2sg ERG	*yintu*	*-ntu*	*ngintu*	*-ntu*
2sg NOM	*yini*	*-ni*	*ngimpa*	*-mpa*
2sg ACC	*yininha*	*-ni*	*nguma*	*-uma*

Note that the only shared pronouns are 1sg ERG *ngathu* (and in the dual the first person *ngali*); but these are also well known outside these languages and descend from a more distant ancestor. The suffix *-ayi,* an optional form in Yardliyawara, heard from Barney Coffin, was probably due to Adnyamathanha

influence. In the Yarli languages the singular bound pronouns follow an 'ergative-absolutive pattern' (syncretising NOM and ACC) while in Paakantyi the pronouns have three distinct forms. In both languages in the dual and plural the systems are 'nominative-accusative' (syncretising the ERG and NOM).

Verb morphology shows a range of differences also. Both language subgroups have a single verb conjugation and a general verb structure of Root+Tense+NOM pronoun for intransitive sentences and Root+Tense+ERG pronoun+ACC pronoun for transitive sentences

The forms and meanings of their inflectional categories are rather different, however. This is shown particularly in tense marking as indicated in Table 4.

Table 4: *Malyangapa and Paakantyi tense marking*

	Malyangapa	Paakantyi
past	*-nganta-*	*-ty-*
yesterday past	*-la-*	
last night past	*-ngantinta-* (absent from Yard)	
this morning past	*-miRinganta-* (absent from Yard)	
present	*-rnta-*	*-ø-*
future	*-yi-*	*-t-*
evening future	*-ngantiyi-*	
morning future	*-miRiyi-*	
imperative	*-ø-*	*-ø-*

4.3 Grammatical differences from Karnic

4.3.1 General

Bowern (1998:30) has listed some of the main morphological features that distinguish the Yarli languages from Karnic. The Yarli languages do not share the change of the locative case to the dative, a change that occurs in all Karnic languages but Arabana-Wangkangurru. Furthermore, the ablative in the Yarli languages is not based on the ergative. The link between the ergative and ablative is a trait shared by all Karnic languages. Also, in the Yarli languages there are different demonstrative forms from Karnic and there are no deictic increments. Finally, unlike the more easterly of the Karnic languages, those of the Yarli subgroup show no sign of gender marking in nouns or pronouns. Table 5 lists some nominal/pronominal features.

Table 5: *Comparison of Yarli languages and Proto Karnic*

	Yarli	Proto-Karnic
Nominal Ergative	*-ngu*	**-ngu/-lu*
Nominal Dative	*-tha*	**-ku*
Pronominal Dative	*-tha*	
Locative	*-nga*	**-la/ *-nga*
Ablative	*-tyali*	**-ngu*
3sg pronoun	*nhu-*	**nhan* (fem), **nhu* (masc)

Those forms that are shared between the Yarli languages and Proto-Karnic, namely the locative *-nga*, the ergative *-ngu* and the 3rd singular pronoun base *nhu-* are by no means an indication of a close relationship between the Yarli languages and Karnic: they are much more widespread and go back to pPN.

Those forms in Table 5 that are very different from Proto-Karnic, however, are significant features for the recognition of the Yarli languages as a subgroup.

The ablative *-tyali* as such is an innovation of the Yarli languages and there seems to be nothing similar in any of the language subgroups in the vicinity, Karnic, Thura-Yura or Paakantyi. From a historic perspective the first syllable of the suffix *-tyali* goes back to pPN as an ablative and "having" marker.

4.3.2 The suffix -tha

The dative/allative *-tha* is bi-valent: it is also a verbal suffix in the extended form *-yitha* (*-yi* FUT + *tha)*, as in the form *pakayitha* "in order to go" quoted above from Wadikali. Neither as a purposive nor as a dative-allative suffix can *-tha* be reconstructed for Proto-Karnic or for Proto-Thura-Yura.

There is a nominal suffix *-tha* in one distant Karnic language, namely Wangka-yutyuru, once spoken in parts of the eastern Simpson Desert and along the Mulligan Channel. It is used there as a genitive-possessive, and may well go back to the same pPN ablative and "having" marker *-tya* that was noted above for *-tyali*.

There is however a suffix *-tya* ~ *-itya*, which cannot be reconstructed for Proto-Thura-Yura, but is well attested in just one Thura-Yura language, Kaurna from the Adelaide plains (Jane Simpson pers. comm. on evidence from Teichelmann and Schürmann 1840). *-itya* has the following main functions in Kaurna:

(5) a. On nouns it is a purposive:
parngutta wild potato
parnguttitya for wild potatoes

b. On verbs it is a purposive:
punggondi to heap up (present form)
punggetitya for heaping up

c. Tt is added to the ergative form of pronouns to form an allative:
nindo you (ergative form)
nindaitya to you

It is possible that the Kaurna suffix *-itya* had a pronominal origin within Thura-Yura (J. Simpson pers. comm.). There is also a good possibility that *-itya* (a palatalised version of the extended form that is only found with verbs in the Yarli languages) was borrowed from the Yarli languages into Kaurna. Kaurna is not adjacent to Malyangapa: the Thura-Yura language Ngadjuri is in between. Unfortunately the presently available data on Ngadjuri have practically no morphological content, and it is impossible to say whether there were *-itya* forms in Ngadjuri. The proposition that *-itya* was borrowed form the Yarli languages into Kaurna via Ngadjuri therefore remains just that, a proposition. There is also a possibility that in a more distant way Yarli *-tha* is related to the directional *-tharV* of Thura-Yura. One thing remains certain: *-tha, -itha* as such is an innovation shared by all three Yarli languages.

4.3.3 The inchoative

The inchoative forms within the subgroup illustrate how by innovation as well as by the retention of different Pama-Nyungan features the Yarli languages cannot be classed with their neighbours. This is shown in Table 6.

Table 6: *Forms of the inchoative*

	YAR, MAL	YAN	WAN	DIY	ADN
to become	*-ngunti*	*-na*	*-minda*	*-ri*	*-ri*
to become (good or bad)	*-r-*	*-na*	*-minda*	*-ri*	*-ri*

The inchoative suffix *-r-* is used in both Malyangapa and Yardliyawara as a verbaliser, but only, as far as our recordings go, with the adjectives “good” and

"bad", as in the sentence quoted above (4) from Yardliyawara (repeated below) and as in (6), which is from Malyangapa.

(4) *Wanka iniki-nha wanyu-r-utu, pulkata!*
meat that -NOM bad-INCH-CAUS throw away IMPER
"Throw that meat away because it's gone bad." (Barney Coffin)

(6) *Wanyu-r-arnta-nyi ngurna-yi -nyi*
bad- INCH-PRES-1sg intr lie down-FUT-1sg intr

palyu mingku-ra-yi-nyi.
soon good-INCH-FUT-1sg intr
"I'm beginning to feel no good, I'll lie down and I'll come good by and by." (George Dutton)

With other adjectives a verbalising inchoative suffix *-ngunti* is used, as in *mantha-ngunti* "cool down", lit. "get cold".

The suffix *-ngunti* appears to be an innovation in the Yarli languages. The *-r-* verbaliser, however, is widely known in several forms (both with a retroflex *R* and with an alveolar tapped *r*), and the relationship between the various forms is not clear. It can be reconstructed for Thura-Yura (Simpson and Hercus this volume, chapter 8) and it is found in parts of Karnic (e.g. Diyari). It is certainly not a feature that would imply any close association of the Yarli languages with either Karnic or Thura-Yura.

5. Summary and conclusion

There is phonological evidence such as pre-stopping (applying to Yardliyawara only), lexical evidence such as a number of 'regional' words, and grammatical evidence such as the use of specialised 'time of day' markers in Malyangapa which all point towards a measure of linguistic diffusion cutting across genetic relationships. But the overwhelming testimony of much unique joint lexical and grammatical innovation in the Yarli languages provides a solid and deeper link between them. There can be little doubt that they form a small separate subgroup of Pama-Nyungan. Whether Schmidt was right and whether there was once a higher grouping of Karnic, Yarli and Thura-Yura remains as yet uncertain.

Developing Interactive Knowledgebases for Australian Aboriginal Languages — Malyangapa

Peter K. Austin

Department of Linguistics and Applied Linguistics

University of Melbourne. Parkville. Vic 3010 Australia

16 May 2002

1 Background[1]

This paper is a description of an interactive database model which I have developed to describe and analyse data on Australian Aboriginal languages. The approach is exemplified with materials from Malyangapa, an extinct language once spoken in western New South Wales. I have done this for a number of reasons. Firstly, I wish to test the concept of an interactive database for language documentation by applying it to a language for which the corpus is relatively self-contained but rich enough that a valuable documentation project results. Secondly, I wish to test the flexibility of current computer software to implement such a system. Thirdly, I wish to challenge the dominant paradigm in the academic study of Australian Aboriginal languages that has emphasised the writing of descriptive grammars, and to argue that an equally important goal of research should be the production of fully articulated and richly described corpora, especially for languages that are not longer spoken (which means the vast majority of Australian languages), which can provide an enduring legacy for future generations of researchers and language community members. Current developments in computer software, data models, and emerging cross-platform standards of representation such as XML, mean that this task is easily within the reach of all practising linguists.

The Malyangapa interactive knowledgebase is implemented in the SIL Shoebox 5.0 computer program as a series of ASCII text files that contain indexes and pointers between them, with rich linkage between the different data and metadata information. The system is designed to

---

[1] This paper was originally prepared for the Workshop on Australian Aboriginal Languages organised by Rachel Nordlinger, University of Melbourne, in March 2002. I am grateful to the late Stephen Wurm for access to his fieldnotes and recordings, and to Luise Hercus for sharing with me her detailed knowledge of the indigenous languages of eastern Australia. Needless to say, neither scholar is responsible for the details of my analysis of Malyangapa. Thanks also to SIL for making their powerful Shoebox program available to the linguistic world, and to David Nash and David Nathan for helpful comments on an earlier draft. This paper was written while I was on sabbatical leave from the University of Melbourne, which I thank for giving me the opportunity for an extended period of research leave.

enable users to move between different types of data and to fully explore the relationships between words and sentences, capturing as much information as possible about the Malyangapa materials. References to the original fieldnotes enable all instances to be located and checked. The general design for this interactive knowledgebase is one that has been successfully applied to other Australian Aboriginal language materials, some of which are much more complex and extensive (Austin 2001). It is my hope that it can serve as an example of what is possible using current computer software in terms of providing richly annotated documentation of a closed corpus of materials on a language that is no longer spoken. A planned future phase of this work will see export of the data files in a more widely used format such as extensible markup language (XML) for distribution among interested scholars and community members.

The Malyangapa language was traditionally spoken in far western New South Wales. Materials on the language were recorded by Stephen Wurm in 1957 with Hannah Quayle, born near Yancannia in about 1875, and Alf Barlow. This material consists of 48 pages of fieldnotes (24 double-sided sheets) plus a brief tape-recording, amounting to 386 sentences and containing a vocabulary of 358 items. Luise Hercus also did some recording with George Dutton in the mid-sixties on Malyangapa; Jeremy Beckett had previously worked with him on social and cultural traditions. Luise Hercus also recorded Laurie Quayle, son of Hannah Quayle, checking some of the earlier materials. He passed away in 1976, and with his death the language became extinct. The Hercus and Beckett data has not yet been incorporated into the current data files.

Malyangapa is relatively closely related to the neighbouring Wadikali and Yardliyawarra languages forming with them the Yardli group (Austin and Hercus 2002). There is limited data on these other two languages, and it is planned in future to include this material for comparative purposes.

2 Research on Australian Aboriginal languages

The academic study of Australian Aboriginal languages has been dominated over the past thirty odd years by a research paradigm that saw the primary goal of linguistic research as the writing of descriptive grammars in a more-or-less standard structuralist format covering phonology, morphology and syntax, with some brief mention of sociolinguistics and peculiarities of language use such as special speech styles for taboo contexts[2]. The grammars that emerged from this tradition were typically single-volume compact studies of 300-400 pages, or in the case of truly moribund languages, a 50-100 page sketch following the format of the *Handbook of Australian Languages* (Dixon and Blake 1978-2000). Almost completely lacking have been comprehensive dictionaries and text collections (as pointed out in Austin 1991), and there has been no concern for corpus-based approaches to linguistic research.

[2] This model has been more generally espoused for *all* languages by Dixon, who writes: "if every person who called themselves a linguist settled down to provide a full description of a single previously undescribed language, then he or she would justify the title" (Dixon 1994:229, cf. Dixon, 1997:135ff "What every linguist should do"). What Dixon means by 'full description' is a single volume descriptive grammar.

There are a number of problems that arise as a direct result of this research and publication paradigm:

- the resulting grammars are necessarily limited both in terms of their depth and breadth of coverage. There is usually insufficient detail to enable testing of the author's claims, and no concern with *tokens* of language forms or categories that would enable checking of distributions, frequencies, or other token patterns in a corpus. This point has been forcefully made by Heath (1984:5): "The extensive exposition of textual citations and statistics in many chapters of this volume may strike some readers as reflecting a personal fetish of mine. While this may be true, it is a fetish I would defend. ... it gives a more patient (or more skeptical) reader a feeling for the raw data which underlie the analysis and the opportunity to "cross-examine" the author by going directly to the data. It also encourages readers with highly specialised interests, or with a different theory of language, to discover new patterns which I overlooked or did not have space to discuss". He goes on to say: "my concern with documentation reflects my own sad experiences as a reader of other linguists' grammars, which almost never provided me with the information I wanted to undertake my own (re-) analysis of the language in question. It also reflects my experience that most published grammars are based on material obtained in unreliable direct-elicitation (sentence-translation) sessions, and/or utterances which were produced by the linguist with or without "confirmation" from a native informant"[3];
- there is a lack of appreciation among linguists operating in this paradigm that grammar writing is a tertiary level of language documentation — the primary documentation is the audio, video and other recorded media together with the original fieldnotes and transcriptions made by the researcher in collaboration with native speakers[4]. This was pointed out almost thirty years ago by Goddard (1973:86): "most descriptive linguists probably feel that their finished grammars have a greater validity, in some sense, than their raw fieldnotes. But **the field notes are the primary documents, the nearest thing to the actual speech events there is**, and they should always ultimately be deposited in a suitable library or public archive, together with explanatory information on dates of fieldwork, relevant characteristics of informants, changing transcriptional conventions, and indexes. Only if this practice become more general can the present situation be improved, in which numerous cases of possible informant errors, artifacts of elicitation methods, misprints, and miscopyings remain

[3] An example of this problem is the phenomenon found in a large number of eastern Australian languages and described in Austin 1997 whereby a single verb affix has either applicative or causative effect depending on the semantics of the root to which it is attached. Most of the grammars of the relevant languages (which typically fail to describe the split) contain one or two examples of the phenomenon but do not examine it in detail — only by combing through dictionaries (where they exist) and by cross-linguistic comparison is it possible to uncover the semantic range of this covert category. No corpora are available for examination.

[4] The secondary level is the stage between fieldnotes and grammar writing when fieldnotes are reworked and retranscribed, example sentences are selected, analysed and glossed, paradigms are assembled, and the linguist 'works out' the structure of the language.

forever undetected or in doubt because of the impossibility of checking them against the primary documents" [emphasis added, PKA];

- as a result of the death of most of the indigenous languages of Australia, including their almost total extinction in the southern part of the country, research on these languages has entering a new phase where corpora are necessarily static[5] and analysis of the languages will need to draw on traditional philological approaches to extinct languages. Important in such work will be detailed annotations of primary documents, and careful comparisons of all available sources, including pre-modern materials collected by amateurs or less trained observers (see Austin and Tindale 1986 for one example of this; Blake's publications on the languages of Victoria over the past 10 years are also another instance). A shift of emphasis from grammar writing to corpus documentation will need to take place as a result. Goddard (1973:86) argued the case for this orientation for Americanist linguists a generation ago: "the linguist who has a philological approach looks not only to the past but also to the future; **he must be concerned with minimizing the problems which the documents he produces will cause his successors. This means making explicit in the fullest practicable way all the information about a form or a corpus that a future investigator might seek.** It is impractical, of course, to give full particulars for every form ever cited in print. But it is possible to do more along these lines than most Americanists have been accustomed to in the past." [emphasis added, PKA]. Goddard (1973:88) argues further that there is a need to focus: "on the fact that there are and will be only a finite number of documents recording the native languages of North America. **It is necessary to make the fullest and most careful use of what there is, and to exercise the greatest diligence in preserving this corpus for the future in the most useful possible form.**" [emphasis added PKA]. Exactly the same arguments apply to current research in Australia.

For these various reasons, the time is ripe for a shift in direction in Australian language documentation back to examination and analysis of primary documentation, including provision of fully detailed data and metadata descriptions for the existing corpora. Fortunately, computer software of various types is now becoming available to make this task easier than it has ever been in the past. This paper is intended as an example of how it is possible to implement corpus documentation in a flexible database format.

3 Software

For this project I have chosen to use the Shoebox 5.0 program developed by the Summer Institute of Linguistics (www.sil.org) which is a general tool for information management, oriented towards linguistic data. Shoebox maintains a series of ASCII files in a standard file

[5] Some southern languages, such a Kaurna or Kamilaroi, have expanding corpora as a result of language revival projects and associated language engineering. This material is however qualitatively different from data collected from actual speakers or rememberers of the languages.

format where each record begins with a data type code (in the form \x) followed by the data and a carriage return. Typically there is a (bilingual) lexicon listing all morphemes and a database of glossed sentences listing sentence forms, morpheme-by-morpheme glosses and free translations. Shoebox provides a number of functions such as semi-automatic filling-in of information ('interlinearisation' or 'glossing'), generating a wordlist (a list of all items in a chosen data field with an associated count and index list of occurrences), generating a concordance (a listing of chosen data items with their preceding and following context), and automatic numbering of a data set. Shoebox also serves as a viewer of lexicon, texts, wordlist and concordance, allowing the user to have multiple windows open on screen. It has a 'jumping' function that instantiates hypertext links by search and pattern match (rather than anchors and pointers as in HTML), linking data in different files and allowing the user to move between material in linked fields in related data files.

Shoebox provides export into XML format but this is weakly developed and does not capture the hierarchical structure of glossed texts which is only indirectly represented in Shoebox by virtue of vertical alignment on screen (encoded as spaces within the text data files).

4 Data structures

The core of the Malyangapa knowledgebase is a set of files of two types: data files and metadata files. The data files capture knowledge about lexical information and sentence structure. The metadata files deal with metalinguistic terminology (a controlled vocabulary encoded as abbreviations), and information about the people involved in the project (as speakers and recorders). Since there is only one set of notes from Wurm's fieldwork and all the material is elicited, there is no metadata recorded on such topics as source genre, date and time of recording, transcriptional practices, or other background information[6].

The data and metadata files are linked by exploration pathways ("jump paths"), hypertext links between named fields within one file and fields in a related file. Clicking on an item in a jump path field updates the related file window with the linked data (thus clicking on a morpheme in a glossed sentence updates the lexicon window with that morpheme's record, or clicking on a sentence number in the wordlist file displays the related sentence in the glossed text window). The exploration pathways are shown in Diagram 1.

Diagram 1 goes here

It is also possible to encode these links as anchors and pointers (as with HTML) by transforming the Shoebox files into marked up data files for viewing by other software such as web browsers. A sample of what this might look like can be seen in the Appendix. I have not yet written the necessary routines for converting the Shoebox files to such a format.

[6] Such metadata is stored for other languages I have been working on (see Austin 2001).

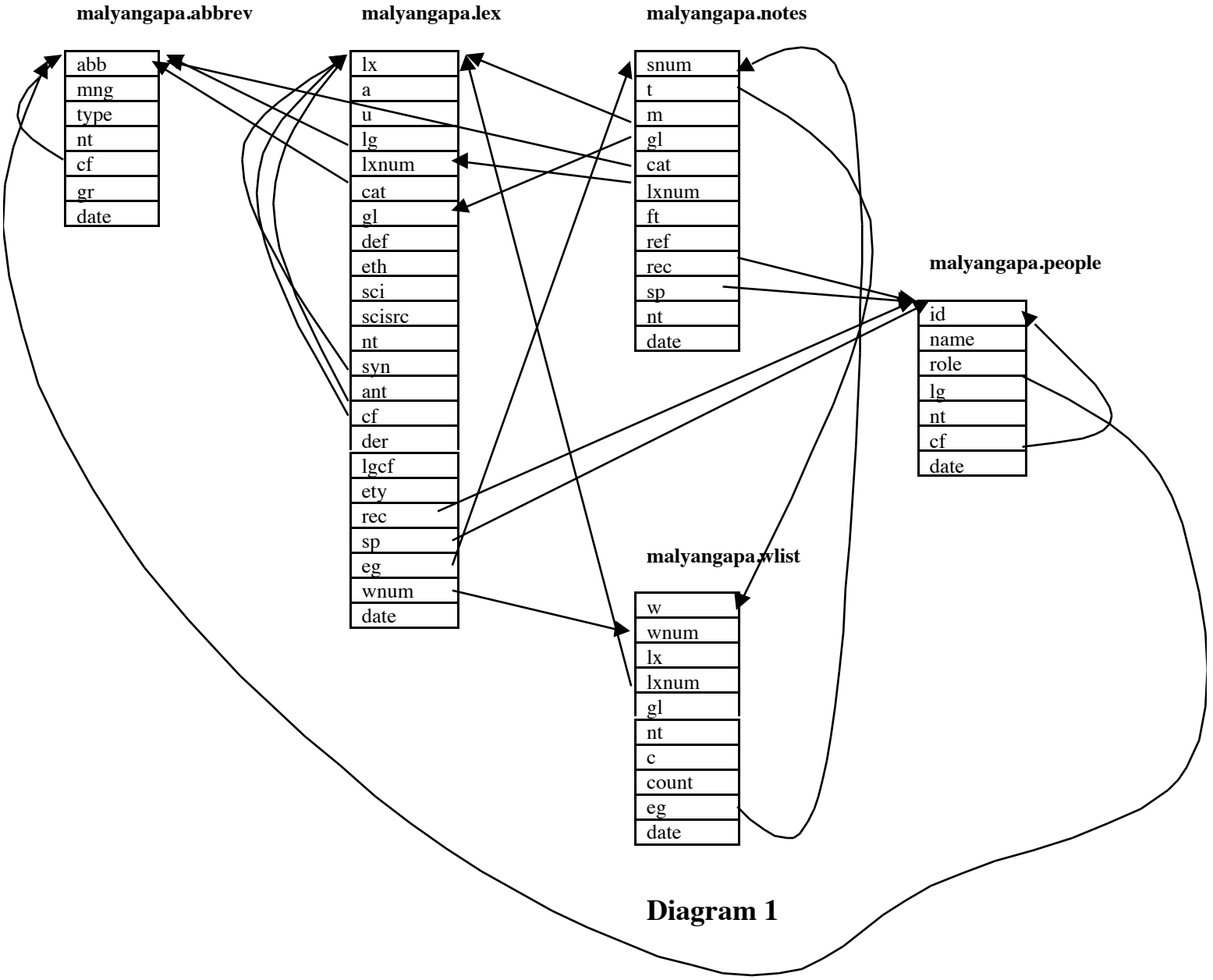


Diagram 1

Note: Arrows indicate hypertext links

4.1 Data files

There are two types of language data files: those dealing with lexical information and those dealing with sentence analysis. The lexical material is stored in two files:

- a lexicon that gives lemmas in Malyangapa with their glosses and definitions in English
- a wordlist that lists all occurrences of word forms in the sentences, together with their lemmas, the number of occurrences, and references to the sentences within which the items are found. This wordlist was generated using Shoebox's wordlist and glossing functions

The lexicon has the following structure:

malyangapa.lex

lx	the Malyangapa lemma spelled in a practical orthography
a	alternative forms of the lemma, used for morphophonemic alternations[7]
u	underlying forms of morpheme combinations where the surface form is not simple concatenation of lemmas, or the 'shortest match' principle adopted by Shoebox gives the wrong morphological parse[8]
lg	language of the lemma [currently Malyangapa but to be expanded to the other Yarli languages]
lxnum	unique numerical identifier for lexical entries
cat	morpho-syntactic category
gl	gloss, usually a single English word, or a sequence of English words separated by periods [used in sentence glossing and finderlist generation]
def	definition of the lemma in English
eth	ethnographic information about the referent
sci	scientific name for plants and animals

[7] Shoebox uses a simple concatenative morphology model. If the lemma has allomorphs these must be listed in an 'a' field, which Shoebox is set up to look at before using the 'lx' form.

[8] We give two examples of this from the Malyangapa lexicon: firstly, the surface form pulanha at the end of a verb is ambiguous between the suffix lemma -pulanha 'third person dual transitive object pronoun' (X *verb* them two] and the sequence of suffixes -pula 'third person dual transitive subject' plus -nha 'third person singular transitive object' [They two *verb* him/her/it]. Under the lemma -pula the field a contains -pulanha and the field u contains –pula -nha. A second example occurs under yuRinga 'be deaf'; here a lists yuRingarntayi and u lists yuRinga -rnta –yi ['be.deaf -pres -emph'] because there are two other morphemes yuRi 'ear' and -ngarnta 'past' which Shoebox would otherwise associate with yuRingarnta using its 'longest match' principle.

scisrc	source for scientific identification
nt	any other notes on form, meaning or usage
syn	form of any synonymous term
ant	form of any antonymous term
cf	form of any other related term[9]
der	list of any derived forms of the lemma
lgcf	cross-reference for lemmas in other related languages [not currently implemented]
ety	etymological information about the lemma including reconstructed proto-forms [not currently implemented]
rec	initials of person who recorded the entry [cross-reference to **people** file]
sp	initials of speaker who provided the entry [cross-reference to **people** file]
eg	reference number for example sentence in which the lemma occurs [cross-reference to **notes** file]
wnum	reference number for word forms containing the lemma [cross-reference to **wlist** file]. This field gives access to all the word forms (and hence all the sentence examples) actually occurring in the corpus for any individual lemma[10]
date	date stamp for entry [generated by Shoebox]

[9] Further sense relations could be distinguished by setting up additional fields (eg. metonymy, hyponymy) if these can be determined from the data. For Malyangapa the material is insufficient to enable this degree of semantic specification.

[10] The wnum reference to the wlist file gives the user indirect access to all occurrences of a lemma in the notes through the lists of wordform sentence numbers generated by Shoebox. The eg field in the lexicon can contain a list of just those occurrences selected by the linguist analyst to illustrate the lemma.

An example of a lexical entry is:

```
\lx    kurntu
\a
\u
\lxnum 086
\lg    Ml
\cat   n
\gl    many
\gl    much
\def   much, many, plenty
\eth
\sci
\scisrc
\nt
\syn   marru
\ant
\cf
\der
\lgcf
\ety
\rec   SW
\sp    HQ
\eg    036
\wnum  113; 114; 115; 116
\date  12/Mar/2002
```

The Malyangapa wordlist is a full list of all wordforms occurring in the sentence examples, together with a count of the number of tokens of each wordform and a list of the example sentence reference numbers in which the wordform occurs. The data in this file is generated using Shoebox's wordlist function run over the example sentences file (**notes**, described below), and then numbered with a unique wordform identifier (using Shoebox's number function). Analysis and lemmas of the wordforms was generated by using Shoebox's glossing (parsing) facility against the lexicon. The relevant wordform identifier numbers were then written back into the wnum field for each lemma in the lexicon. The value of the wordlist is that it gives all occurring forms of lemmas, together with their token frequencies, and hypertext links back to the sentences from which they are extracted.

malyangapa.wlist

w	wordforms occurring in the sentence materials
wnum	unique numeral identifier for wordforms [generated by Shoebox number function]
lx	lemmas for wordform [generated by Shoebox parsing with cross-reference to **lex** file]

lxnum	lemma identification numbers [generated cross-reference to **lex** file]
gl	glosses for lemmas [generated cross-reference to **lex** file]
nt	notes
c	number of wordform tokens in sentence data [generated by Shoebox wordlist function]
count	count of word form occurrences in decimal format [generated by Shoebox wordlist function, can be sort field for frequency analysis]
eg	reference number for example sentence in which the wordform occurs [generated by Shoebox, cross-reference to **notes** file]
date	date stamp for entry [generated by Shoebox]

An example of a wordlist entry is:

```
\w     ngapanga
\wnum  226
\lx    ngapa -nga
\lxnum 076   -075
\gl    water -loc

\c 3
\count 000003
\eg 071; 207; 306
\date 17/Feb/2002
```

Malyangapa sentence data is stored in a single file that contains surface sentence forms[11], their aligned morpheme-by-morpheme glosses (generated using Shoebox's glossing (parsing) function against the lexicon), free translations of each sentence, identification of speaker and recorder, and any additional notes (often comments on problems of analysis or the surface forms).

Unfortunately, Shoebox does not support media but a nice complement to this file would be scanned images of Wurm's original fieldnotes keyed to the page reference identifier, and links to digitised sound files (Wurm made tape recordings of part of his material). I hope that this can be implemented in the future.

malyangapa.notes

snum	unique numerical identifier for each sentence [generated by Shoebox]
t	surface form of sentence in a practical orthography

[11] The *t* (for 'text') field is transcribed in a practical phonemic orthography that closely resembles but is not identical to Wurm's transcription. In work on other languages I have a separate field for the original transcription, thus recording it separately from my analysis of the surface sentence form.

m	aligned morphemic representation of wordforms [generated by Shoebox glossing (parsing) into lemmas using the lexicon]
gl	aligned English gloss of each morpheme [generated by Shoebox glossing]
cat	aligned syntactic category of each morpheme [generated by Shoebox glossing]
lxnum	aligned lemma number of each morpheme [generated by Shoebox glossing]
ft	English free translation of sentence
ref	reference to page number and sentence number in fieldnotes
rec	initials of person who recorded the sentence [cross-reference to **people** file]
sp	initials of speaker who provided the sentence [cross-reference to **people** file]
nt	notes on sentence
date	date stamp for sentence [generated by Shoebox]

An example of the data in the sentence file is:

```
\snum  386
\t     kata     wanthayi     yiniki
\m     kata     wantha -yi   yiniki
\gl    cockatoo where  -emph that
\cat   n        n      -suff dem
\lxnum 362      071    -034  035
\ft    Where is that cockatoo?
\ref   SW2/1Bs05
\rec   SW
\sp    AB
\nt    Wurm's gloss "Look at that red and white cockatoo"
\date  18/Jul/2001
```

4.2 Metadata files

The metadata files contain non-linguistic background information about the data in the lexicon and sentence collections. For Malyangapa this consists of a list of all the abbreviations and material relating to the recorders and speakers who contributed data.

malyangapa.abbrev

abb	unique abbreviation
mng	meaning of the abbreviation, usually a short description in lay terms of the functions of syntactic category labels

type	type of abbreviation, eg. category, language name
nt	notes about the abbreviation
cf	cross-reference to other related abbreviations
gr	cross-references to grammar of Malyangapa [not yet implemented]
date	date stamp for item [generated by Shoebox]

An example of an entry in this file is:

```
\abb  vtr
\mng  transitive verb
\type sub-category
\nt   transitive verbs are a sub-category of verbs; they take a
      transitive subject argument in ergative case and a
      transitive object argument in accusative case.
\cf   vi, vdi
\gr
\date 12/Mar/2002
```

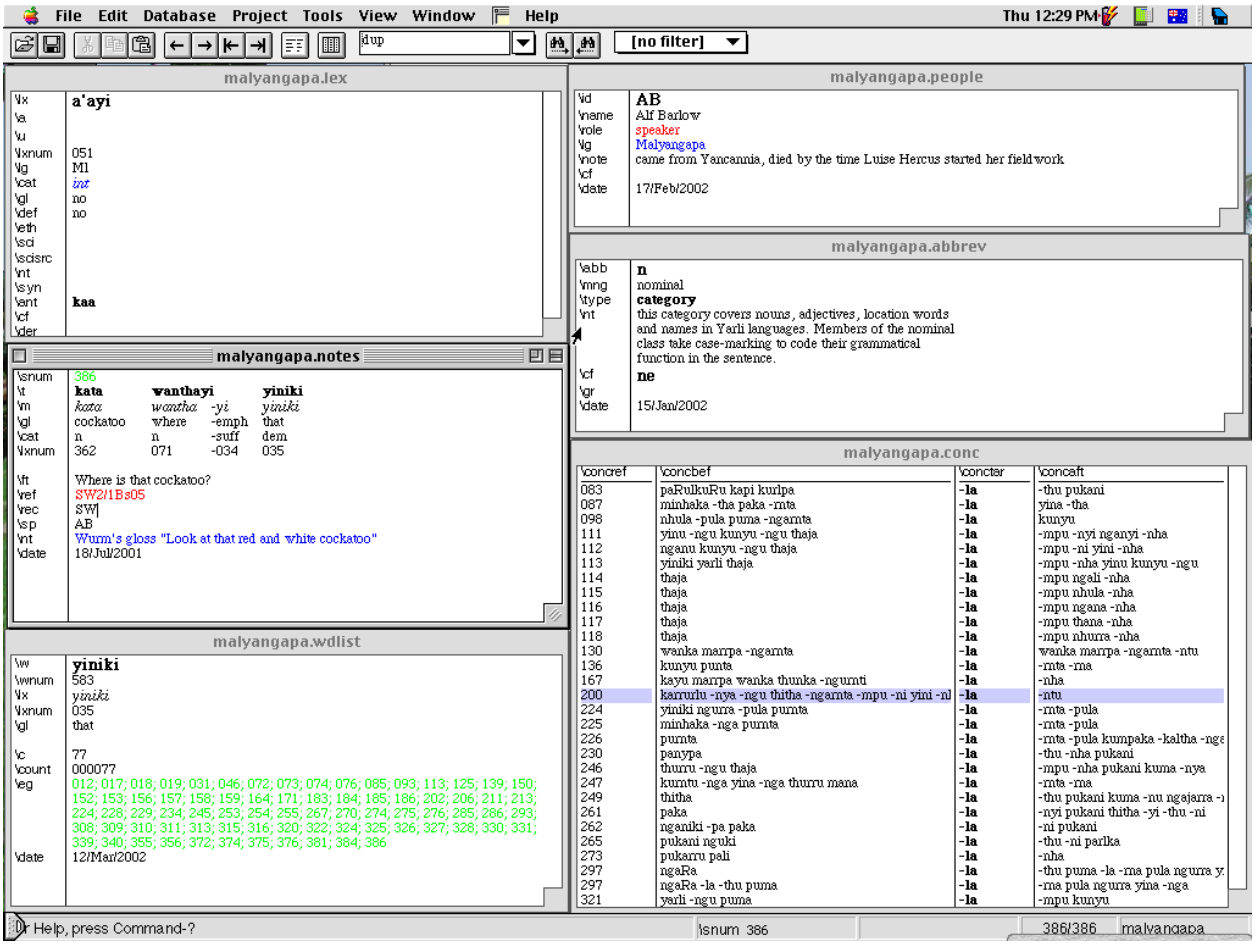
malyangapa.people

id	unique abbreviation for each person [generally initials of firstname and lastname]
name	person's name
role	role of person in language documentation
lg	language spoken (if Aboriginal language speaker)
nt	notes on individuals
cf	cross-reference to other individuals in this file
date	date stamp for item [generated by Shoebox]

A sample entry from this file is:

```
\id   LQ
\name Laurie Quayle
\role speaker
\lg   Malyangapa
\note consultant for Luise Hercus, son of HQ
\cf   HQ
\date 17/Feb/2002
```

Here is a screen shot showing a sample of the files as displayed within Shoebox — the windows display lexicon (top left), notes (middle left), wordlist (bottom left), metadata (top right and middle) and a concordance (bottom right). User interaction is by right-click (or option-click on the Macintosh) within the linked data fields (shown as arrows in Diagram 1).



5 Conclusions

The study of Australian Aboriginal languages is entering a new phase that requires a focus on detailed corpus documentation and analysis. Powerful software tools exist that enable such documentation to be accomplished, and this paper has presented an example of a data model applied to materials on one extinct language that shows the value of such an approach. I have also demonstrated how the Shoebox program can serve as a hypertext viewer of a quite complex set of data and metadata that enables users to fully explore an analysed and annotated corpus. In future research I hope to export this data into a format that allows viewing and interaction by general purpose tools such as web browsers.

6 References

Austin, Peter K. 1991 ‘Australian Aboriginal language studies’, In Michael Clyne (ed.) *Linguistics in Australia: trends in research*, 55-74. Canberra: Australian Academy of Humanities and Australian Academy of Social Sciences.

Austin, Peter K. 2001 Developing an interactive knowledgebase for the Mantharta languages, Western Australia. Presented at Department of Linguistics and Applied Linguistics, University of Melbourne, March 2001, and University of Utrecht, February 2002.

Austin, Peter K and Norman B. Tindale 1986 ‘Emu and brolga, a Kamilaroi myth’. *Aboriginal History*, 9:8–21.

Austin, Peter K. and Luise A. Hercus 2002 ‘The Yarli languages’. Paper presented at International Conference on Historical Linguistics, Melbourne August 2001, to appear in the proceedings.

Dixon, R.M.W. 1994 *Ergativity*. Cambridge: Cambridge University Press

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

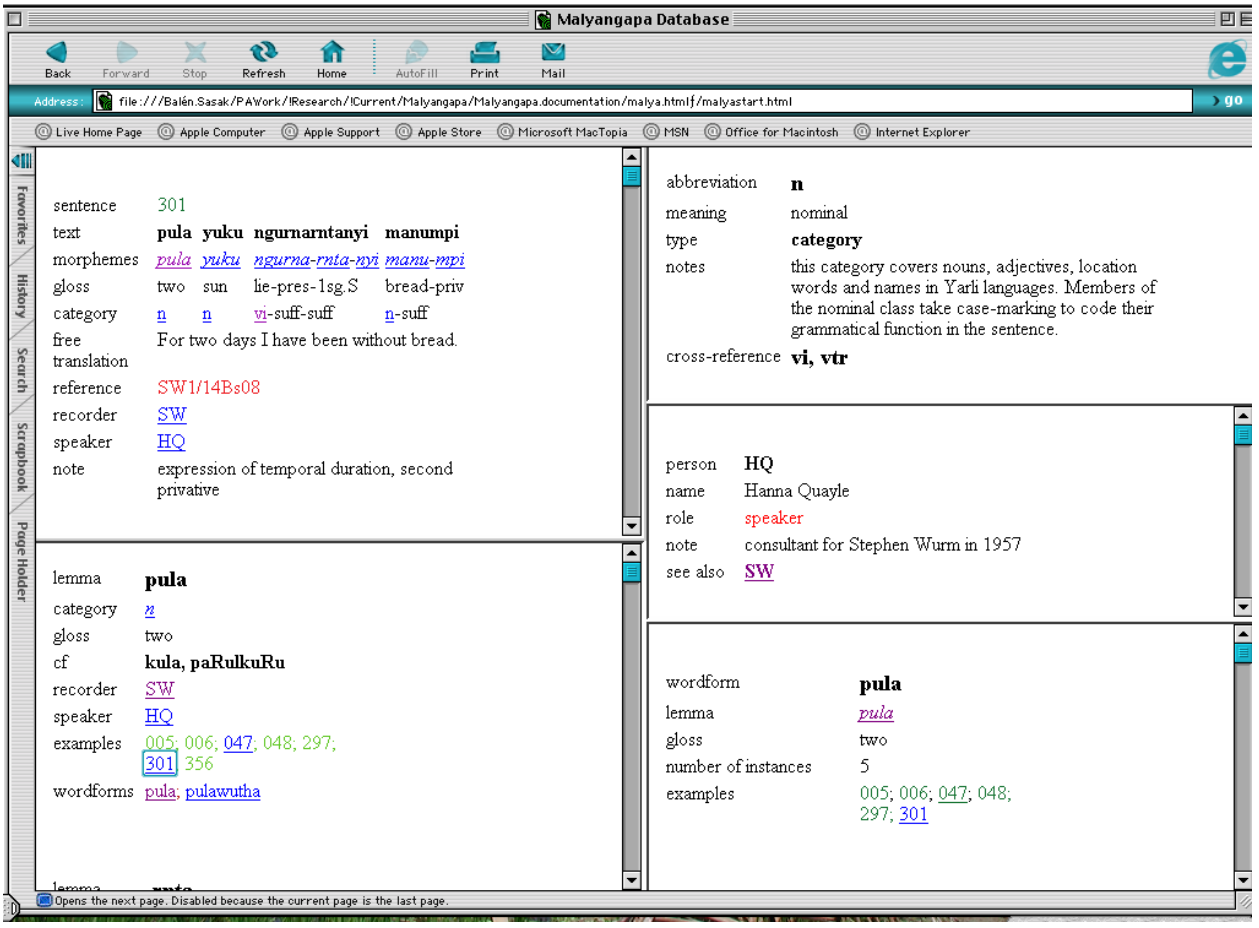
Dixon, R.M.W. and Barry J. Blake 1978-2000 *The Handbook of Australian Languages*. Vol 1-5. Canberra: ANU Press and Melbourne: Oxford University Press.

Goddard, Ives. 1973. ‘Philological approaches to the study of North American Indian languages: documents and documentation’. In Thomas A. Sebeok, ed. *Current Trends in Linguistics, Vol 10*. The Hague: Mouton.

Heath, Jeffrey 1984 *A Nunggubuyu grammar*. Canberra: Aboriginal Studies Press.

7 Appendix — Web browser files

This is a mockup view of the Shoebox database as HTML tables with hypertexts links as viewed by Internet Explorer.



No. 69.—EVELYN CREEK.

By H. Crozier, Esq.

Kangaroo	chukeroo.
Opossum	murlu.
Tame dog	koonai.
Wild dog	thirita.
Emu	kurlithe.
Black duck	pundrewunga.
Wood duck	
Pelican	
Laughing jackass	
Native companion	mulumpari.
White cockatoo	kugalurinya.
Crow	kowulka.
Swan	kuteruk
Egg	kupi.
Track of a foot	tina.
Fish	kooia.
Lobster	
Crayfish	muracuru.
Mosquito	koonti.
Fly	ulberu.
Snake	thuru.
The Blacks	yarlee.
A Blackfellow	kurna.
A Black woman	kumbuga.
Nose	minthe.
Hand	murra.
2 Blacks	kurna munderu.
3 Blacks	kurma baralkeru.
One	koola.
Two	munderu.
Three	baralkeru.
Four	munderu-munderu.
Father	kooma.
Mother	ngumma.
Sister–Elder	kanuberti.
,, Younger	karee.
Brother–Elder	kaku
,, Younger	neeyi.
A young man	karuwurle.
An old man	karu-karu.
An old woman	koorilpu.
A baby	yalibuthe.
A White man	birre-birre.
Children	
Head	kukaminta.
Eye	mirlke.
Ear	kutchera.

No. 69.—EVELYN CREEK—*continued.*

Mouth	murna.
Teeth	murna-thunthera
Hair of the head	kukawincha.
Beard	ngunka
Thunder	thuna.
Grass	kuntha.
Tongue	thurli.
Stomach	mundera.
Breasts	ngumma.
Thigh	ngura.
Foot	tina.
Bone	moko.
Blood	kurte.
Skin	ngunya.
Fat	murne.
Bowels	kurnangundere.
Excrement	koodna.
War-spear	birra.
Reed-spear	
Throwing-stick	
Shield	puragu.
Tomahawk	nali.
Canoe	
Sun	uku.
Moon	pirtall.
Star	purle.
Light	dudthera.
Dark	tinka.
Cold	muntha.
Heat	nurtekurla.
Day	dudthera.
Night	tinka.
Fire	kurla.
Water	ngapa.
Smoke	moyu.
Ground	nurte.
Wind	yurke.
Rain	koolpe.
God	
Ghosts	
Boomerang	kira.
Hill	pumperu.
Wood	kurla.
Stone	kurnu.
Camp	nginchera.
Yes	ngagu.
No	punni.
I	nginyi.
You	yine.
Bark	ngonyia-ngonyia.
Good	minko.
Bad	winu.
Sweet	minko.
Food	munu.
Hungry	purangu.
Thirsty	wirltunga.
Eat	thulinu.
Sleep	ngurwanu.
Drink	thapernu.
Walk	purlkanu.
See	thitthanu.
Sit	pula.
Yesterday	kulginyie.
To-day	kerreri.
To-morrow	koonigoonirri.
Where are the Blacks?	noweraku?
I don't know	wertarie.
Plenty	perriri.
Big	pirna.
Little	wakarraka.
Dead	
By-and-by	muta.
Come on	kuba.
Milk	thunka.
Eaglehawk	kurrera.
Wild turkey	kurlathura.
Wife	nongo.

No. 69.—EVELYN CREEK.

By A. Dewhurst, Esq.

Kangaroo	turlda.
Opossum	pilta.
Tame dog	kunnuya.
Wild dog	
Emu	kurlitchi.
Black duck	kultappi.
Wood duck	
Pelican	
Laughing jackass	
Native companion	puralko
White cockatoo	kilumburra.
Crow	kaulka.
Swan	kutteroo.
Egg	kappi.
Track of a foot	tina.
Fish	kuya
Lobster	
Crayfish	umpurra.
Mosquito	kunti.
Fly	ilburroo.
Snake	turroo.
The Blacks	yalli.
A Blackfellow	
A Black woman	kumbuka.
Nose	minta.
Hand	murra.
2 Blacks	
3 Blacks	
One	koola.
Two	boola.
Three	barlgo.
Four	krundoo.
Father	kumma.
Mother	umma.
Sister–Elder	karaloo.
,, Younger	kalawarri.
Brother–Elder	karkoo.
,, Younger	
A young man	kulta.
An old man	garroo-garroo.
An old woman	walgunnuga.
A baby	multa-bobbra.
A White man	
Children	kurndoo, waltanna.
Head	karkunta.
Eye	milparloo.
Ear	yuimerta.

No. 69.—Evelyn Creek—*continued.*

Mouth	tia.
Teeth	
Hair of the head	kakowunta.
Beard	unkurroo.
Thunder	kulpi.
Grass	muttoo.
Tongue	tarlindi.
Stomach	ayamulla.
Breasts	numma.
Thigh	yaltarra.
Foot	tinna.
Bone	mookoo.
Blood	karti.
Skin	parlatta.
Fat	mulni.
Bowels	naimoola.
Excrement	goornana.
War-spear	kal-kurroo.
Reed-spear	
Wommera	
Shield	gulgarra.
Tomahawk	karro.
Canoe	
Sun	yookoo.
Moon	pitali.
Star	purli.
Light	
Dark	
Cold	munta.
Heat	nitiulla.
Day	bookaninti.
Night	neilba.
Fire	wi.
Water	nappa.
Smoke	tooba.
Ground	nulti.
Wind	yarlto.
Rain	kulpi.
God	
Ghosts	
Boomerang	wunna.
Hill	yanda.
Wood	wi.
Stone	yunda.
Camp	ichurra.
Yes	kow.
No	
I	nunjie.
You	yimba.
Bark	bindarra.
Good	minko.
Bad	
Sweet	taukoo.
Food	talata.
Hungry	boorakinnia.
Thirsty	
Eat	talindalto.
Sleep	unangi.
Drink	
Walk	bulkarannia.
See	
Sit	ninadunnia.
Yesterday	bokanni.
To-day	kaiio.
To-morrow	pamyinga.
Where are the Blacks?	
I don't know	narooringo.
Plenty	gurndo.
Big	wilto
Little	bumpata.
Dead	palino.
By-and-by	
Come on	
Milk	tarpunda.
Eaglehawk	purti.
Wild turkey	kaldura.
Wife	