

A YOLNGU-MATHA DICTIONARY-  
PLANS AND PROPOSALS

R. David Zorc

1. INTRODUCTORY NOTES <sup>1</sup>

In May and June 1979, I had the opportunity to attend a lexicography workshop under the direction of Professor Ladislav Zgusta in Manila. There he stressed the differences among:

(i) lemma - basic information necessary to locate a word in the language, including grammatical information and irregularities in form or pronunciation, speech register, and so on;

(ii) main entry - semantic information and phrase or sentence examples;  
and

(iii) subsidiary information - cross-references and etymology.

The emphasis at all times was *basic* information; economy of style and presentation are crucial in an age of costly publication prices and users who want easy answers to their queries. Hence the lexicographer must ask himself two crucial questions:

(i) What information must I include? and

(ii) What information may I exclude?

for *each* entry! He stressed the difficulty of making dictionaries and referred to a classic quip: "It is no longer necessary to have prisons or capital punishment, all we need do is have our criminals make a dictionary - that will be punishment enough" (cf. Zgusta (1971:15)).

During 1983, I plan to begin work in earnest on a pan-Yolngu-Matha dictionary (other commitments have precluded any earlier), and I anticipate the project to run through to 1985. Since this is a dictionary representing speech varieties across a language family, I foresee a large (but hopefully not insurmountable) number of problems, which I will address here in more general terms, following some questions posed by Peter Austin (see introduction).

---

Peter Austin, ed. *Papers in Australian linguistics* No. 15:  
*Australian Aboriginal lexicography*, 31-40.  
*Pacific Linguistics*, A-66, 1983.

© R. David Zorc

## 2. USES AND USERS

Someone might well ask: "Why make a Yolngu-Matha dictionary in the first place?". The idea arose because I have heard criticisms of existing dictionaries and wordlists, mainly from the Yolngu themselves: "That word isn't Gupapuyngu, it belongs to the Dhuwa moiety, and Gupapuyngu is Yirritja!". Similar comments have been made in response to stories by various Aboriginal authors. The fact is that any speaker may use any word from some 55 Yolngu-Matha speech varieties, depending on family background and upbringing, current death taboos, or even 'style'. Hence, I have begun typing data cards with language, dialect, or moiety information, based on the provenance suggested, and on cross-checking against many language assistants. Yolngu-Matha speakers may be among the world's leading 'purists', for when they see a book or wordlist labelled 'Gupapuyngu', they expect to find words that are genuinely *only* Gupapuyngu, even if Gupapuyngu speakers have borrowed forms from other Yolngu-Matha dialects. They do not carry this purism into Makassarese or English loanwords, but are concerned over the 'ownership' of the Yolngu-Matha lexicon. Hence, the Yolngu-Matha dictionary will label each form on the basis of a number of cross-cutting native principles (moiety, broad speech group, and clan group; see 4 (ii) below).

Realistically, a Yolngu-Matha dictionary is more likely to have a limited demand than to become a 'best seller'. Nevertheless, I do not plan to address an audience of linguists (who should be able to deal with virtually any format), but rather the interested layman, including workers and teachers in Aboriginal communities who would like to learn something of the language, and the literate Northeast Arnhemlander. The last group are literacy workers, council members, and primary school teachers who need to write down and insure accuracy of spelling in their language, and who require further information on English spelling, meaning, and usage. This choice of audience already imposes the need for a good deal of information, including phonological/phonetic, grammatical, semantic, sociolinguistic and cultural, and puts a strain on the principle of economy. The Balanda (European) will need to get a good deal of information from the Introduction, where the phonetics and grammar will be summarized, before he or she can use the dictionary efficiently. The Yolngu will be seeking probable English<sup>2</sup> counterparts to words in his or her language, including labels as to the rudeness, politeness, or eruditeness of English levels and the collapsing of grammatical differences he or she may well be used to. For example, *guḷa* is a noun meaning 'faeces, excreta' (erudite), 'stool, bowel movement' (polite), 'shit' (rude), while *biḷk-thun* is a verb meaning 'defecate' (erudite), 'move bowels' (polite), 'shit' (rude), 'pooh' (baby-talk). The Yolngu often use only the rude forms, because they are unaware of the alternatives, while the European attempts to make a verb of *guḷa* or a noun of *biḷk-thun*, and hence must remember the grammatical information as to the uses of a noun and all its suffixes as opposed to a verb of Group 1 and all of its tense/aspect forms and derivations.

## 3. THE CORPUS

The data will be drawn from stories and wordlists produced by students of the School of Australian Linguistics (S.A.L.), as well as the works of several other scholars, including Lowe 1976 and Christie 1979 on Gupapuyngu, Davis (in press) on Gupapuyngu flora and fauna, Morphy 1983 on Djapu, Schebeck (n.d.) on Rirratjingu, Wood 1978 on Gälpu, Heath 1980 on Ritharrngu, Wilkinson 1981 on Djambarrpuyngu and Ross and Walker (in press) on Gumatj. A computer and secretarial assistance would be invaluable, but, unfortunately, I will have access to neither, at least to the scale required. At present, I plan to include all words gathered, including those known to be under taboo, but such forms will be marked with a prefixed dagger, for example, †*guya* 'fish' [cf: *ngarirri*]. I am open to suggestion on this issue, but language assistants have agreed with this procedure and use it themselves, since one is allowed to write, but not say, tabooed names or words. However, sacred and/or secret ceremonial terms will not be included (at the request of Yolngu assistants) because the dictionary is likely to be seen by women and children.

## 4. THE LEMMA

The following information is considered critical:

(1) The Yolngu-Matha word (or morph, in the case of productive suffixes) in the established orthography. Alphabetical order will also follow current conventions: *a, ä, b, d, ḡ, dh, dj, e, g, i, k, l, ḷ, m, n, ŋ, nh, ny, ng, o, p, r, rr, t, ṭ, th, tj, u, w, y*<sup>3</sup>. Glottal stop ' is ignored, but forms with it follow those without, for example, *bala* 'away' precedes *bala* 'house' and *djawar-yun* 'be tired' precedes *djawar'-yun* 'pierce, spear, inject'. Digraphs are treated as if they were single letters, and follow the completion of a non-digraph. Thus, all *ḡ-* forms occur before *dh-*, all *dh-* forms before *dj-*, and so on. Any departure from this approach reaps unpleasant consequences. A Gupapuyngu Dictionary prepared by Michael Christie ran into heavy criticism because he decided to ignore all digraphs and followed English alphabetical order strictly, with, for example, *dj-* forms after *di-/ḡi-* entries. In a letter to Yolngu-Matha linguists dated 20th March 1982, I suggested I might disregard differences in vowel length and collapse entries of *ä-a*, *e-i*, and *o-u*, and received several strong (and one vituperative) reactions against this proposal. Yolngu-Matha orthography and alphabetical order has become a sacred cow. If a clear phonetic statement is included in the introduction, I do not see the need for phonetic information in the lemma (such as Schebeck has included in his examples (see page 53)), although there are a few cases of exceptions that would need exemplification, for example, the clan name *marika*, with accent on the penult rather than the first syllable and the interrogative particle *muka*, with accent on the ultima, and so on. Besides the mass total of ink and space saved, the dangers of indicating a given pronunciation may be interpreted as prescriptive where many dialectal or even idiolectal alternates exist: *gumatj* = [ *gúmatʷ* ] or [ *gúmætʷ* ]. This is particularly the case in a pan-Yolngu-Matha dictionary, but may be less so in a strictly Dhangu'mi (Rirratjingu) dictionary.

I might add that I strongly oppose semantic orderings, because of the difficulty of information retrieval or location. In wordlists, this procedure is warranted, particularly if an alphabetised appendix is included, but in a dictionary many semantic groupings are blurred or downright arguable.

(ii) Coded information as to the dialect or provenance of a form. The Yolngu have three cross-cutting guidelines:

- (a) Moiety = *Dhuwa* or *yirritja*, or both, that is, in use by any/all speakers;
- (b) Language type = *dhuwala'mirri*, *dhuwal'mirr*, *dhay'yimirr*, *dhangu'mi*, *djangu'mi* (based on the proximate deictic for 'this');
- (c) Dialect = *Gupapuyngu*, *Gumatj*, *Dhalwangu*, *Djapu*, *Marrangu*, *Marrakulu*, *Liyagawumirr*, *Djambarrpuyngu*, *Rirratjingu*, *Gälpu*, *Ngaymil*, *Wangurri*, *Warramiri* (based on membership in one or more clans, for example, *Gupapuyngu* includes *Gaykamangu*, *Gumbula*, and *Marrkula* clans);

To these may be added two linguistically-derived criteria:

- (d) Subgroup = Northern Yolngu (including *Dhangu'mi* and *Djangu'mi*) which have *nhän* 'he/she'; or Southern Yolngu (including *Dhuwala'mirr'*, *Dhuwal'mirr*, *Dhay'yimirr*, *Ritharrngu*) which have *ngayi* 'he/she';
- (e) Vowel dropping (including *Dhuwal'mirr*, *Dhay'yimirr*, *Dhangu'mi*, *Djangu'mi*);

Thus, for the five synonyms for 'big', there are: *ngutu* (may be used by any moiety or social group), *bathala* (*Yirritja*), *yindi* (*Dhuwa*), *qumurr* (vowel-dropping) and *qumurru* (*Yirritja*, non vowel-dropping); for the intensive marker 'very much so', there are: *wirrk* (*Gumatj*, *Djapu*), *marimi* (*Dhangu/Djangu'mi*), *mirithirr* (*Dhuwal'mirr*, *Dhay'yimirr*) and *mirithirri* (*Dhuwala'mirri*). Ultimately, abbreviations will be developed for all of these categories.

This sort of information may not be available for every form in the dictionary, but will be sought. Where such information cannot be specified accurately, it will best be left out, signifying 'don't know', rather than pan-Yolngu-Matha.

(iii) Coded information as to the grammatical function of a form, including the following categories:

- (a) pronouns (personal) and demonstrative pronouns (deictics);
- (b) noun, those taking root case suffixes, such as *-dhu* (ergative), *-lili* or *-li* (allative), *-nguru* (ablative), *-kurru* (progressive), *-ngura* or *-nga* (locative);
- (c) name, those roots taking suffixes such as *-nha* or *-ny* (accusative), *-wala* or *-wul* (locative/allative), *-gung(u)* (originative);
- (d) locational, taking only a limited number of case suffixes;
- (e) verb, specified as transitive or intransitive, and with an indication of group membership. The following is my provisional classification of March 1982, with revisions based on feedback from several linguists working on Yolngu-Matha:<sup>4</sup>

- Group 1 [-un, -urru, -urruna, -una(ra)] (most productive group)
- Group 2 [-ma, -ngu, -ngala, -nha(ra)] (causative, factitive, production)
- Group 2a [-kama, -kungu, -kangala, -kanha(ra)] (6 forms noted)
- Group 2b [buma, bungu, bumara, bunha(ra) 'hit']
- Group 2c [ngäma, ngäku, ngäkula, ngänha(ra) 'hear']
- Group 2d [-thama, -thulu, -thangala, -thanha(ra)] (2 forms noted)
- Group 3 [-mirri, -mirri, -mina, -minya] (Reflexive, Reciprocal)
- Group 3a [-i+rri, -∅, -na, -nya(ra)] (includes -thi+rri inchoative)
- Group 3b [-i, +∅, -na, -nya(ra)] (7 forms noted)
- Group 4 [-a, -i, -na, -nha(ra)] (historically old stems, 16 forms noted)
- Group 5 [-an/-aṅ, -ulu, -ara, -ana] (largish group, depending on dialect)
- Group 5a [-an/-aṅ, -urru, -ara, -ana] (dialect variants of Group 5)
- Group 5b [-tjan, -yaku, -tjarra, -tjana] (3 forms noted)
- Group 5c [-tjan, -tjurru, -tjarra, -tjana] (1 form noted)
- Group 6 [-∅, -∅, -∅, -∅] (loanwords that do not change inflection)
- (f) preverb particle, indicating tense or mood, for example, *dhu*, *yurru*, *ngarru*, signalling future, *ga*, *yukurra*, *yaka*, progressive, and *nguli*, *baying*, habitual;
- (g) verb replacement particle, substituting for verbs in discourse, for example, *rur* 'stand' (= *dhärra*), *mit* 'cut' (= *gulk-thun*) and *ḡulk* 'run' (= *wanḡi*);
- (h) adjectival, semantically justifiable as a separate class in that they have true antonyms and grammatically may occur with uninflected intensives (*wirrki*, *marimi*, *mirithirri*);
- (i) discourse particles, inserting mood or giving subtle shifts to statements, for example: *muka*, *ngatja*, interrogative tag; *-nha*, *-na*, sequence, and *wäy*, optative, plea;

- (j) conjunctions, including *ga* 'and', *bala* 'then', *wo* 'or', *märr* 'so that', *bili* 'so', *maku* 'or' (with words), *nganydja* 'or' (with clauses) and *balangu* 'when, if';
- (k) interrogatives, that is, *nhä* 'what?' (corresponding to nouns), *yol*, *wara* 'who?' (corresponding to names), *wanha*, *ngala* 'which?, where?' (corresponding to locationals), *nhawuy*, *nhawi* 'what-you-may-call-it?' and *nhaltjan* 'do what?';
- (l) temporals, *yalala*, *yalnguwa* 'later on', *ngäthili* 'earlier', *yawungu*, *barpuru* 'yesterday', *gäthura* 'today', and so on;
- (m) quantifiers, such as: *wanggany* 'one', *bulal*', *märrma* 'two', *bukmak*, *warrpam* 'all';
- (n) interjections, such as *way* 'hey', *yakay* 'ouch', *go* 'come on', and so on.

Some linguists may find more classes than necessary represented here, and I am open to collapsing them, although I am trying to keep the non-linguist in mind (in terms of semantic and grammatical behaviour of the Yolngu-Matha forms). Such information needs to be abbreviated, but should not be excluded. For example, it is useful and important to know that *gungda* 'stone' is both noun (*gungdalili* 'towards the rocks') and name (*gungdawala* 'to/with Jessie M. Gaykamangu'). Similarly, one must know if a verb is transitive or intransitive and some of its basic grammatical requirements in order to use it correctly, for example, *wanga* (v-intr-4) 'talk, speak to (-gu); talk, speak about (-galangawuy); speak (a language) (-kurru).'

## 5. THE MAIN ENTRY

Semantic information should be clear and succinct. Phrase or sentence examples should be provided if they help specify the meaning. Zgusta warned against long detailed lists of possible meanings, for example, *manymak* = 'good, nice, fine, well, delicious, healthy, beautiful, pretty, neat'.

The language level should be aimed at the most general audience, for example, *manymak ngatha* 'delicious food' (rather than: 'tasty tucker'). When examples are included, one must consider if they might better belong under a different entry, for example, *manymak miyalk* 'good wife', probably is better suited for *miyalk* 'woman' than for *manymak* 'good'.

A general guide to extensions of meaning or polysemy is required within the main entry; sub-numerals are not necessary if the entry can be kept short, so that semicolons divide each major diversity, for example, *walu* 'sun; day; time(-piece)'. The last element indicates that 'time' and 'time-piece' are both appropriate meanings. Another example is *gurtha* 'fire(wood); matches, lighter'. Homonyms, on the other hand, should require either separate entries or separate numbers within an entry, for example, *bul'-yun* (1) 'play'; (2) 'swell-up'.

Yolngu-Matha is exceptionally rich in idioms, and these will be treated within the main entry, although presentation may be a problem. *ngoy* 'seat of emotion, one's inner self; guts' has over seventy idioms that could take up two full pages if indented at the start of each new sub-entry; some other body parts, for example, *buku* 'forehead', *gong* 'hand, arm, finger', have nearly as many. Elegance would appear to conflict with economy, and I am not sure which avenue to follow at present. Possibly the use of ... would allow a flow between each idiom and save space as well.

A dictionary need not (and probably should not) be an encyclopedia, but certain basic cultural information should be included, including the scientific names of flora and fauna, where possible: *muta-muta* 'red berry; eaten when ripe; juice used for treating ringworm'; *Grewia retusifolia*.

Tribal wisdom should also be tapped (and preserved) by showing the interconnection between flora, fauna, and natural events ("when this tree bears flowers, we know that the parrot fish are abundant", "when these yams are ripe, we know the wet is close at hand", and so on), or with people and clans ("this shark is the totem of Clan X", "this tree is the totem for the *buḷany'* subsection").

## 6. SUBSIDIARY INFORMATION

After the main entry, but not necessarily indented separately, will come some, but not generally all, of the following information:

(i) synonyms, for example, *manymak* 'good' [Syn: *ngamakurr(u)*, *ngamakuli*, *laytju*]; *liya* 'head' [Syn: *ḍāmbu*, *mulkurr*, *gayawak*, *waḡangga'*].

(ii) antonyms, for example, *dhunupa* 'correct; straight; right-hand/side' [Opp: *djarrpi* 'crooked', *wiḷ'ku* 'left'], *mopuk*; 'salty (water)' [Opp: *raypiny*].

(iii) cross-references, for instance, *miny'tji* 'colour' [cf: *gurrngan*, *mol* 'black', *miku* 'red', *milkuminy*, *mulkuminy* 'blue/green', *watharr* 'white']

(iv) internal comparisons to less transparently related forms, for example, *məl* 'eye' [cf: *milkuma* 'show', *milma* 'within sight', *milkarri* 'tears', *milparrambarr* 'eyelashes']; *wanggay* 'one' [cf: *wanggayngu* 'other, different']; *baḡami* 'woman' [cf: *waḡangu* 'proper; ownership of']; *nhina* 'sit; stay' [cf: *dhut* (verb replative)]; *bur* (verb replative) 'arrive' [cf: *buna*]. For Yolngu-Matha, this would also include noun-verb pairs, for example, *gara* 'spear' [cf: *dharpuma* 'to spear']; *waltjaḡ* 'rain' [cf: *dhayun* 'to rain']; and non-productive derivations, *marrnggitj* 'medicine-man, sorcerer' [cf: *marrnggi* 'to know'].

(v) etymological information, if available, for example *bathala* 'big' [Makassarese *battala*? 'heavy, burdensome'], *bathi* 'dillybag, container' [Makassarese *patti* 'box, case, chest; coffin', cf: Malay *p ti* 'box, chest, case'], *məl* 'eye' [Proto-Pama-Nyungan \**mi:l* 'eye', cf: Warlpiri *mil+pa*].



Zgusta felt very strongly in favour of putting an etymology at the *end* of an entry (rather than after the lemma, as is done with some English dictionaries). "Is it an etymology or a warning?", he would ask. This has a further advantage here in Australia, where so few forms can be related to reconstructions - they are not glaring when they are absent. Although historical linguistics is still in its infancy here, compared to Indo-European or Austronesian, enough information is available to warrant at least some etymological information of a reliable nature. For example, Walker and Zorc's 1981 study of Makassarese and Austronesian loans in Yolngu-Matha has applications for most languages of northern Arnhemland (Burarra, Anindilyakwa, Nunggubuyu, Djinang) and could shed at least some light on the provenance of non-Australian forms. Similarly, fully assimilated English loans should be marked, as could the more definite reconstructions of Proto-Pama-Nyungan, lower-order subgroups such as Proto-Yolngu, or possibly even a Proto-Australian (?). However, the recommendation should be considered that it is better to leave out a risky etymology than produce a blatant error.

#### 7. POSTSCRIPT

I have had access to drafts of the papers in this volume by Austin, McKay, and Schebeck and find myself generally in agreement with what they say. I am encouraged to see solutions to my problems emerging from their work and recommendations. I would be happy to correspond with any scholars on points of mutual interest, and receive comments as to the directions I am taking (including the wrong turns and possible detours).

#### FOOTNOTES

<sup>1</sup> I wish to recommend Zorc 1979 as summarising a number of my thoughts on the requirements and desiderata of various types of dictionary (basic, bilingual, and etymological). Offprints are available on request, so I will not repeat or dwell upon points already dealt with in that paper.

<sup>2</sup> Although the Yolngu-Matha/English dictionary will have as one of its goals a pedagogical guide for Yolngu-Matha speakers to English usage, no English to Yolngu-Matha dictionary is planned, and would indeed be a difficult venture. *One* would be needed for *each* specific Yolngu-Matha communict, but there would be no advantage to having a long list of forms translating English glosses without duplicating the sociolinguistic information to be contained in the Yolngu-Matha sections. Should computer facilities become available to the School of Australian Linguistics, such separate English to Yolngu-Matha communict lexicons could be produced by a series of programmed commands, for example, make a list of all Gumatj, Dhuwala'mirri, and Southern Yolngu forms for an English - Gumatj lexicon, or make a list of all Djapu, Dhuwal'mirr and vowel-dropping Southern Yolngu forms for an English - Djapu lexicon, and so forth.



<sup>3</sup> In the Yolngu orthography ä represents long [a:], e represents long [i:] and o represents long [u:]. Underlining signifies retroflexion (apico-domal articulation), and digraphs with h as the second element are lamino-dental sounds. Throughout this paper the standard Yolngu symbol ŋ is spelled ng for typographic convenience.

<sup>4</sup> Copies of my paper detailing this classification (Zorc 1982b) are available on request. The four suffixes set out on each line in the following listing (except for Groups 2b and 2c, which have only one verb each) are: citation form or 'simple present', 'future' (definite or tomorrow future, positive or negative imperative), 'past' (non-specific or recent past), and 'remote past, negative of specific past' (also used in derivations for causative, reciprocal, infinitive, adjectival or anti-passive forms). The symbol + indicates a more hypothetical (less transparent) morpheme boundary than those marked with -. Elements after + are dropped in subsequent inflections. Thus, in Group 3a, +rri disappears in the remaining tense/aspect inflections, for example 'to desire' is *djal-thi+rri*, *djal-thi-∅*, *djal-thi-na*, *djal-thi-nya(ra)*. An example from Group 3b would be 'to go' *marrtj-i*, *marrtj-i*, *marrtji-na*, *marrtji-nya(ra)*, and from Group 4 'to speak' *wang-a*, *wang-i*, *wanga-na*, *wanga-nha(ra)*.

## REFERENCES

- CHRISTIE, M.  
1979 Gupapuyngu-English dictionary. MS. N.T. Department of Education.
- DAVIS, S.  
in press *Dictionary by domains - natural species - Gupapuyngu language*. N.T. University Working Papers in Theoretical and Applied Linguistics, Publication 2.
- HEATH, J.  
1980 *Basic materials in Ritharngu: grammar, texts and dictionary*. PL, B-62.
- LOWE, B.  
1976 Temporary Gupapuyngu dictionary. Michael Christie, ed. Milngimbi. Roneoed.
- MORPHY, F.  
1983 Djapu. In: R.M.W. Dixon and B.J. Blake, eds *Handbook of Australian Languages*, vol. 3:1-188. Canberra: Australian National University Press.
- ROSS, J. & A. WALKER  
in press *Gumatj wordlist: part one*. N.T. University Working Papers in Theoretical and Applied Linguistics, Publication 1.

40 R. DAVID ZORC

SCHEBECK, B.

n.d. Dhangu-Nomina and Dhangu-Verba. MS.

WALKER, A. & R.D. ZORC

1981 Austronesian loanwords in Yolngu-Matha of northeast Arnhemland.  
*Aboriginal History* 5: 109-134.

WILKINSON, M.

1981 Djambarrpuyngu Thesaurus. MS. N.T. Department of Education.

WOOD, R.

1978 Some Yuulngu phonological patterns, *Papers in Australian linguistics* No. 11. PL, A-51:53-117.

ZORC, R.D.

1979 An etymological, bilingual dictionary - esoteric or essential?  
*Philippine Journal of Linguistics* 10:63-73.

1982a Survey of Yolngu-Matha names (introductory version). School of  
Australian Linguistics. Batchelor. Typescript.

1982b Yolngu-Matha verb stem classification. School of Australian  
Linguistics. Batchelor. Typescript.