

# Making 2,180 pages more useful: the Diyari dictionary of Rev. J. G. Reuther

Peter K. Austin

ORCID ID: 0000-0002-3180-0524

## Abstract

The Diyari language, spoken in northern South Australia, is unusual for an Australian Aboriginal language in having an extensive range of written sources dating from the mid-19<sup>th</sup> to the mid-20<sup>th</sup> century, many created by Lutheran missionaries. Outstanding among the missionary materials is a 14 volume manuscript in German by the Rev. J.G. Reuther, four volumes of which comprise a Diyari to German dictionary. Reuther (1981) is a translation into English of Reuther's manuscript; the dictionary part consists of 2,180 printed pages, but is only available as a microfiche. There is a high level of interest in this dictionary from the descendants of the people Reuther studied with, especially members of the Dieri Aboriginal Corporation (DAC), who are keen to revitalise the language and develop practical teaching materials. In their current forms, both Reuther's original and Scherer's translation present many difficulties for use, both by community members and researchers. This paper reports on current research and development by David Nathan and myself to add value to and increase the utility of the Reuther-Scherer dictionary, and to provide access to its contents in various ways.

## 1. Introduction<sup>1</sup>

The Diyari language,<sup>2</sup> spoken in northern South Australia, is relatively unusual among Australian Aboriginal languages in having an extensive range of written sources dating from the mid-19<sup>th</sup> to the mid-20<sup>th</sup> century (Austin 2021, Chapter 6; Stockigt 2017, Chapter 8). Many of these were compiled by missionaries at Bethesda Mission at Killalpaninna (on Cooper Creek, east of Lake Eyre), which the Lutheran church operated from 1866 to 1915 (Proeve & Proeve 1952; Stevens 1994). At the mission, Diyari was the language of daily use, and was employed by the missionaries in their church services, school, farm work, and everyday activities. Literacy was taught to Diyari speakers, some of whose writings survive to the present (Austin 1986; 2014; Ferguson 1987; Gale 1997). Today, there are a few fluent speakers, and many others who have some knowledge of the language (Austin 2014).

Outstanding among the missionary sources is a 14 volume manuscript written in German by the Rev. J.G. Reuther (see Section 2). Four volumes of this, amounting to 1,238 manuscript pages, comprise a Diyari to German dictionary, containing 4,183 numbered

---

<sup>1</sup> I am grateful for assistance and feedback in compiling this paper from Jane Simpson and David Nash. Philip Jones provided careful commentary on an earlier draft that identified a number of factual and historical errors. My work on Diyari since 2013 has been carried out in collaboration with Greg Wilson and members of the Dieri Aboriginal Corporation. The research reported here is the result of collaboration over many years with David Nathan and is currently partially funded by a Leverhulme Emeritus Fellowship (2021-2023). I dedicate this paper to the memory of the late Luise Anna Hercus, who introduced me to Diyari and provided much valuable assistance with my work on the language from 1974 to 2018.

<sup>2</sup> There are several spellings of the language name that appear in this paper. Reuther used "Diari", current descendants prefer "Dieri", while I spell as Diyari, in accordance with the analysis presented in Austin (2021).

entries and many thousands of sub-entries and example sentences (see Section 3).<sup>3</sup> There is a high level of interest in the dictionary from the descendants of the people Reuther studied with, especially members of the Dieri Aboriginal Corporation (DAC), however in its current form the dictionary, and its 1981 translation into English by Philipp Scherer (see below), presents many challenges for use (see Section 3). This paper reports on current research by David Nathan and myself to increase the utility of the dictionary and to provide access to its contents in various ways, as well as linking it to other sources, such as the Diyari-English reference dictionary and multimedia resource that I am developing. This resource attempts to bring together all available materials from published and unpublished sources, including my fieldwork carried out in 1974-1977, and collaboration with the DAC from 2013 until the present.<sup>4</sup>

## 2. J. G. Reuther and his dictionary

Rev. Johann Georg Reuther (1861-1914) was a German-speaking Lutheran missionary who was in charge of Bethesda mission at Lake Killalpaninna for eighteen years (1888 to 1906).<sup>5</sup> In addition to his missionary work and running the farm and pastoral property, Reuther carried out ethnographic research on Diyari language and culture, partly in the face of opposition from the Lutheran synod, who insisted that he should concentrate on Christian missionary activities.<sup>6</sup> Reuther amassed a substantial collection of physical objects that he obtained from the Diyari and neighbouring groups who resided at Killalpaninna through a series of exchanges and commissions (Figure 1). This included the controversial ‘toa’ objects that are discussed in Jones & Sutton (1986), and Jones (2007, 2012).



*Figure 1. Picture of Reuther and his wife Pauline in his study at Bethesda. Source: Lutheran Archives, P027/41/05316.*

---

<sup>3</sup> The early entries in Volume I contain some comparative materials on neighbouring languages, mostly cognates and a few example sentences. It may be that Reuther had planned to extend this to later volumes but stopped for some reason.

<sup>4</sup> See <https://dieriyawarra.wordpress.com/about/> (accessed 2021-04-15)

<sup>5</sup> For more details of Reuther's biography and the history of the mission see <http://missionaries.griffith.edu.au/biography/reuther-johann-georg-rev-1861-1914#sendnote6anc> (accessed 2021-08-10)

<sup>6</sup> Reuther's language work, such as the 1897 translation of the New Testament with Carl Strehlow, was applauded, but he lost the support of the Synod after his enthusiasm took an ethnographic turn from around 1900.

Reuther compiled a 2,600 page 14 volume handwritten manuscript of Diari (Diyari) language and culture, of which four volumes comprise a Diyari-German dictionary (Figure 2 is a sample showing the handwriting and style of the documents). The manuscript was purchased by the South Australian Museum (SAM) for £75 in 1915 (nine years after Reuther had left the mission in 1906, and after he died by drowning in a horse and cart accident in 1914). It is catalogued as item AA266.

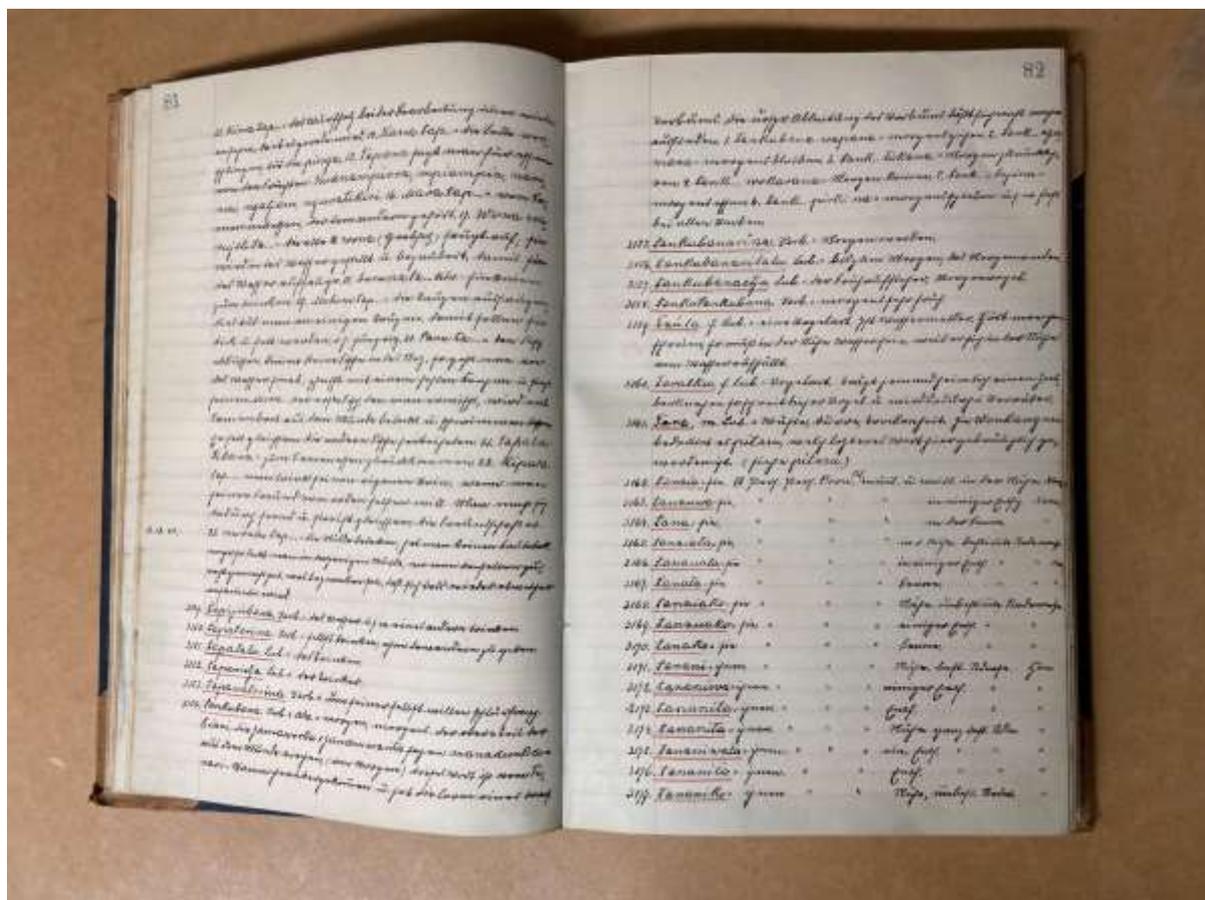


Figure 2. Sample of Reuther’s manuscript. Photo © 2021 Philip Jones, used with permission.

Various attempts were made to translate parts of the Reuther manuscript into English from the 1930s to 1960s, but the dictionary volumes remained largely untouched.

In 1974, the Australian Institute of Aboriginal Studies (AIAS, now Australian Institute of Aboriginal and Torres Strait Islander Studies, AIATSIS) provided funding for Pastor Philipp Scherer, the first archivist of the Lutheran Church of Australia, to translate the whole of Reuther’s manuscript into English. When the AIAS funding ran out, Scherer completed the work unfunded. Luise Hercus, together with her mother Theodora Schwarzschild, and in consultation with Gavan Breen and myself, translated Reuther’s Diyari grammar (Volume 5 of the manuscript) and his grammatical notes on Wangkangurru and Yandruwantha. Figure 3 is a sample (page 1885) from Scherer’s 1974 translation of the dictionary, showing part of Reuther’s volume IV page 80 (compare Figure 2).

3148.

IV,80	tapana\$ (v) = 'to drink'
	1) <u>ngapa tapana</u> = 'to drink water'
	2) <u>paua tapana</u> = 'to slurp [or sip] seed-pulp'
	3) <u>muntja tapana</u> = 'to suck on a patient', i.e. to suck out the rubbish <sup>2</sup> at the seat [or source] of the trouble. The <u>kunki</u> does this.
	4) <u>mitali ngapa tapana</u> = 'for the ground to absorb the water'
	5) <u>gildi tapana</u> = 'to drink the fat'
	6) <u>ngama tapana</u> = 'to suck [at] the breast'
	7) <u>paja kapi tapana</u> = 'to suck out birds' eggs'
	8) <u>durintji tapana</u> = 'to suck the marrow out of a bone'

\fn1. Reuther: "wenn bei jemanden zum einen das andere kommt," – whatever that may mean.

\fn2. Reuther: "Unrat." P.A.S.

*Figure 3. Scherer's translation of Reuther's dictionary, page 1885.*

As Ganter (2019) notes:

Once Scherer's opus of translation was completed, negotiations between AIAS and SAM stalled as the Museum asserted its legal ownership of the original manuscript, and the publication of the translation was suppressed as nobody felt in a position to fund a publication.

In 1981 a microfiche of the whole manuscript was published by AIAS (see Figure 4) – we refer to this as the 'Reuther-Scherer dictionary' in what follows. The microfiche is difficult to use, not only because specialist equipment is needed to read the document, but also because it appears as white text on a black background (as in Figure 5).



Figure 4. AIAS microfiche publication of Scherer’s translation. Photo © 2021 David Nash



Figure 5. Sample pages of dictionary on microfiche. Photo © 2021 David Nash.

### 3. Significance and challenges with the dictionary

In the following sections we discuss the structure and significance of the Reuther- Scherer dictionary, and some difficulties with its form and content. We then outline developments in digitisation of the work and subsequent research on the digital files.

#### 3.1 Significance

Reuther’s dictionary contains 4,179 numbered entries, with over 16,000 sub-entries, mostly compounds or phrases that exemplify particular meanings or uses of the entry word (see

Figures 2, 3 for examples). There are over 1,700 notes that provide additional information about entries, such as relationships to Diyari mythology or ethnographic information about traditional practices, as in the instance of *munlja tapana* 'to suck on a patient' seen in Figure 3. It is a remarkable record of material culture, ceremony, trade, mythology, and associations between them and the landscape. Much of this mythological and traditional knowledge is not available in materials arising from subsequent research because it was lost following the closure of the mission and the disruption of the Diyari community and its transmission of culture. In addition, the dictionary is an extremely valuable source for idioms or other ways of speaking which reflect Diyari cosmology or categorisation. An example of this is the many idioms based on body-part terms which appear scattered throughout the examples in the dictionary.

In the translation, there are 1,692 footnotes added by Scherer, mostly comments on unusual German expressions employed by Reuther, but also some notes on the content; unfortunately, due to his lack of knowledge of Diyari, a number of these footnotes are misunderstandings or misinterpretations of Reuther's material.

### 3.2 Challenges

Hercus (2017: 118) notes:

It is difficult to comprehend how Reuther achieved this work. He had no typewriter, as they were only just coming into usage; no filing cards, just sheets of paper; no encouragement apart from the work of Flierl [his fellow missionary – PKA] and the collaboration with his co-worker Carl Strehlow; and there was hostility from the synod to contend with.

Reuther did not work alone, however, since, as Philip Jones (p.c. 2021-08-09) points out: "Reuther certainly had a 'research assistant' – his wife, Pauline, nee Stolz (daughter of a key individual in the Lutheran hierarchy, which protected Reuther against the sort of critique levelled at his rival, Otto Siebert)".

The existing version of the dictionary is organised alphabetically and has no corrections, so it is possible it is a clean copy of working notes, however none of these seem to have survived.

It seems clear that Reuther had little lexicographic training or knowledge. Structurally, the Reuther-Scherer dictionary is rather poorly organised and consequently difficult to use. According to Stockigt (2017: 332), in 1937 Norman Tindale visited at his home in Germany Otto Siebert, who had been appointed from March 1894 to 1902 as 'Busch Missioner', with responsibility for ministering to the Aboriginal camps around Killalpaninna and Kopperamanna. Siebert "described Reuther as 'lame at languages'" (Tindale 1937). Ganter (2019) adds that Siebert also commented that Reuther's "work was confused and disjointed". There is evidence to support this suggestion:

- (a) the dictionary shows a degree of repetition, probably because Reuther wrote entries without checking earlier exemplification, or having simply forgotten what he wrote previously. Thus, alongside Volume IV entry 3148. *tapana* (v) = 'to drink' which has

23 examples (see Figure 3 above, Figure 6 below), we also find in Volume I entry<sup>7</sup> 229. *dapana* (v.t.) = 'to drink; to suck, to suck up; to kiss; to chew; to eat (of grounded seed); to belch or burp; to wet or moisten; to pour, swallow' with 31 examples, only 12 of which overlap with the examples in entry 3148 (and even then the translations or additional explications and cultural notes with the overlapping examples do not always exactly coincide);

- (b) there are multiword examples of the use of particular forms included under one entry which do not appear (nor are they cross-referenced) under the headword for another form in the example, e.g. *dantjumana tapana* = 'to drink sparingly' appears in example 64 under 209. *dantjumana* = 'carefully', but not under 3184. *tapana* = 'to drink'. Occasionally, some senses of polysemous lemmas are only seen in examples found elsewhere and not included with the lemma itself;
- (c) the numbered sub-entries exemplifying uses of an entry are often randomly presented and are not ordered in terms of semantic sense relationships or selectional restrictions, especially for polysemous items such as *thapa-* (see section 4). Sub-entry combinations with literal interpretations are mixed together with those that are idiomatic, e.g. *ngapa dapana* = 'to drink water' (cf. *ngapa* 'water') is followed immediately by *kana dapana* = [lit:] 'to drink a person', i.e. in the event of sickness to suck it out of a person's body (cf. *kana* 'human being'). Some idioms are clearly tagged by including [lit:] with the literal translation, but others are not, e.g. *kirra dapana* = 'to kiss a kirra 'boomerang', - which means as much as 'to spit on one's hands';
- (d) many of the numbered entries are derived forms of roots, and in standard lexicographic practice they would probably be represented as sub-entries of the root lemma, e.g. 3149. *tapijirbana* 'to drink water (etc.) belonging to someone else' which can be analysed as *thapa-yirpa-rna* 'drink-benefactive-participle', derived from the root *thapa-* 'to drink'.

In terms of the form of entries in the dictionary, the orthography used by Reuther (and other missionaries and writers) generally under-differentiates consonants and over-differentiates vowels. The missionary orthography uses five vowel symbols <a, i, u, e, o>, however Diyari has just three phonemic vowels (*a, i, u*), each of which has a range of allophones (see Austin 2021, Chapter 2); Reuther's <e> is typically *a* in the neighbourhood of laminal consonants, while his <o> can be either *a* or *u*, depending on context. For consonants, he has a voicing contrast (e.g. <d> in *dapana* versus <t> in *tapana*, cited above, or <b> versus <p>, and <k> versus <g>, e.g. *gildi* 'fat' in Figure 3 vs. *kana* 'person'), however voicing is only distinctive word-medially for apico-domal stops (*rt* versus *rd*). Diyari distinguishes lamino-dental, apico-alveolar, and apico-domal points of articulation for stops, nasals and laterals, however Reuther uses <t> or <d> for all of *th, t, rt, rd*, and <n> for *nh, n*, and *rn*,<sup>8</sup> along with <l> for *lh, l* and *rl* (thus his *tapana* is *thaparna*). There is also a three-way rhotic contrast between apical flap *r*, trill *rr*, and post-alveolar glide *r̥*, all of which merge as orthographic <r> (or occasionally <rr>) in the missionary spelling.

<sup>7</sup> Diyari does not distinguish voiceless [t] from voiced [d] so these should be a single item; for problems with the missionary spelling system for Diyari see below.

<sup>8</sup> Stockigt (2017) shows how the grammatical analysis of the missionaries is undermined by their failure to distinguish between the various nasals.

There are also grammatical issues with the entries: for part-of-speech Reuther clearly distinguishes transitive and intransitive verbs (fundamental for case-marking of arguments), but over-differentiates adjectives from substantives (nouns), for which there is no language-internal evidence (Austin 2021). He also marks “masculine” or “feminine” gender for nouns, influenced by his own native German and by other missionaries’ analyses (Stockigt 2017), however in Diyari gender is only expressed in third person singular pronouns entirely on a semantic basis of the referent as feminine versus non-feminine (Austin 2021, Section 3.3.2). Nominal roots do not express gender as such. Reuther also does not distinguish the word classes of adverb, particle, and interjection.

For all these reasons, the dictionary is not easy for users to access and understand, especially those who are attempting to learn the language. In the following sections we discuss how the work can be made more useful by digitalising it and adding an explicit representation of structural content and typographic information.

#### **4. Digitisation and value adding**

The Reuther-Scherer dictionary contains very interesting and valuable information about lexical semantics, collocations, idiomatic expressions, mythological knowledge, and ethnographic knowledge and practices, much of which is lacking in more recent research on Diyari. However, as we have seen, it remains difficult to use in its current form because of issues with the orthography, grammatical information, and scattered and inconsistent nature of the content. It is also very difficult to search in both its paper version (which only exists in the AIATSIS library) and the microfiche publication.

In 1989 David Nash and Jane Simpson, working at AIATSIS on the *National Lexicography Project*, scanned the 2,180 pages of Volumes I to IV of the Scherer typescript using a Kurzweil *Discover 7320 Model 30* scanner and optical character reader. Over 25 sessions between 1st March 1989 and 24th April 1989 they created 44 plain text files. Figure 6 shows a sample scan of page 1885.

3148. #tapana\$ (v) = `to drink.'

1) #nqapa tapana -\$ `to drink water.'

2) #paua tapana\$ = `to slurp [or sip] seed-pulp.'

3) #muntja tapana\$ = `to suck on a patient,' i.e. to suck out the rubbish<sup>2</sup> at the seat [or source] of the trouble. The #kunki\$ does this,

4) #mitali ngapa tapana\$ = `for the ground to absorb the water.'

5) #gildi tapana -\$ `to drink the fat.'

6) #nqama tapana\$ = `to suck [at] the breast.'

7) #pa]a kapi tapana\$ = `to suck out birds' eggs.'

8) #durint]i tapana\$ = `to suck the marrow out of a bone,

1. Reuther: "wenn bei jemanden zum einen das andere kommt,"  
- whatever that may mean,

2. Reuther: "Unrat." P.A.Z.

*Figure 6. Sample scanned OCR output of Reuther-Scherer (1981).*

The accuracy of character recognition is estimated by Nash to have ranged from about 87% to about 96%. Simpson proof-read the scanned files and corrected obvious errors, however many character mis-readings in the Diyari remained (e.g. ] or J for j, nq for ng, ~ for uninterpreted characters), along with random representations of white space. Underlining in the original text came out as #...\$ surrounding the underlined characters, though occasionally with incorrect scope. Some = signs preceding glosses appear as -\$ , as in nqapa tapana -\$ in Figure 6. The resulting files added to the value of the Reuther-Scherer dictionary, but were still short of optimal.

In 1991 I was given a copy of the proof-read scanned text files and partially edited them using Microsoft Word (utilising its ability to search and replace on formatting), attempting to:

- (a) correct character mis-readings, based on identifying obvious errors together with my knowledge of Diyari;
- (b) replace spaces before indented material with a single <tab> character;
- (c) remove blank space at the bottom of pages;

- (d) encode Scherer's page numbers in {...} and Reuther's volume and page numbers in <...>;
- (e) add \fn. before all footnotes.

This resulted in edited files with entries as in Figure 7.

```

┌
{p1884}<IV,80>
3143. #tankatankana$ (v) = 'to add repeatedly'

3144. #tankakalalu$ (n) = '[the act of] adding repeatedly'

3145. #tankanietja$ (n) = 'someone on whom one thing is heaped upon another'

3146. #tankaterina$ (v) = 'to add oneself'

3147. #tankanaterina$ (v) = 'to add, append something for one's own sake'

3148. #tapana$ (v) = 'to drink'
    1) #ngapa tapana$ = 'to drink water'
    2) #paua tapana$ = 'to slurp [or sip] seed-pulp'
    3) #muntja tapana$ = 'to suck on a patient', i.e. to suck out the rubbish2 at the seat [or
source] of the trouble. The #kunki$ does this,
    4) #mitali ngapa tapana$ = 'for the ground to absorb the water'
    5) #gildi tapana$ = 'to drink the fat'
    6) #ngama tapana$ = 'to suck [at] the breast'
    7) #paja kapi tapana$ = 'to suck out birds' eggs'
    8) #durintji tapana$ = 'to suck the marrow out of a bone,

\fn1. Reuther: "wenn bei jemanden zum einen das andere kommt," - whatever that may mean,
\fn2. Reuther: "Unrat." P.A.S.

{p1885}<IV,80>
    9) #kumari tapana$ = 'to drink blood; to suck at a wound'
    When he is first being made a man, the #wiljaru$ opens his mouth widely and drinks
the blood [offered to him],

```

Figure 7. Edited Word file

While the resulting files were now more consistent and easier to search, they remained poorly structured and difficult to use, especially for Diyari people and language learners. A better representation was needed, so in 2014-2015 with funding support from the Dieri Aboriginal Corporation, David Nathan processed the files that I had created in order to produce an XML-marked-up rendition of the document, with tags encoding information content. This enabled David to further clean up the files to remove inconsistencies (e.g. the part-of-speech 'transitive verb' was variously encoded as vt, vt., v.t, v.t.); some of this could be done by ordered sequences of search and replace using regular expressions, but it also involved many thousands of manual edits to distinguish structural units for which there were some (but incomplete) cues in the content or the existing flagging or markup. Using typographical flags introduced in the scanning (see above), such as #...\$ to indicate the beginning and end of Diyari language text, (...) to indicate part-of-speech, and '...' to identify glosses, David

parsed the Word files to insert XML tags such as identification of the entries, lemmas (entry form plus part of speech), glosses, numbered subentries, and Diyari material within glosses or notes, along with page reference locations to the Scherer and Reuther originals. An example of the resulting well-formed XML is given in Figure 8, derived from the excerpts shown above.

```
<lemma><di>tapana</di> <pos val="v" /></lemma> <gloss>to drink</gloss>
<tabp><tabnum num="1" /> <di>ngapa tapana</di> <gloss>to drink water</gloss></tabp>
<tabp><tabnum num="2" /> <di>paua tapana</di> <gloss>to slurp [or sip] seed-
pulp</gloss></tabp>
<tabp><tabnum num="3" /> <di>muntja tapana</di> <gloss>to suck on a patient, i.e. to suck
out the rubbish at the seat [or source] of the trouble. The <di>kunki</di> does
this</gloss></tabp>
<tabp><tabnum num="4" /> <di>mitali ngapa tapana</di> <gloss>for the ground to absorb
the water</gloss></tabp>
<tabp><tabnum num="5" /> <di>gildi tapana</di> <gloss>to drink the fat</gloss></tabp>
<tabp><tabnum num="6" /> <di>ngama tapana</di> <gloss>to suck [at] the
breast</gloss></tabp>
<tabp><tabnum num="7" /> <di>paja kapi tapana</di> <gloss>to suck out birds&quot;
eggs</gloss></tabp>
<tabp><tabnum num="8" /> <di>durintji tapana</di> <gloss>to suck the marrow out of a
bone,</gloss></tabp>
<fn num="1">Reuther: &quot;wenn bei jemanden zum einen das andere kommt,&quot; -
whatever that may mean,</fn>
<fn num="2">Reuther: &quot;Unrat.&quot; P.A.S.</fn>
<location><page num="p1885" /> <reuther vol="IV" page="3" /></location>
<tabp><tabnum num="9" /> <di>kumari tapana</di> <gloss>to drink blood; to suck at a
wound</gloss></tabp>
```

Figure 8. XML representation.

David also wrote a DTD (document type definition) which specifies the structure of the resulting XML, and describes and governs the syntax of the tagged categories. As Nathan (2016) notes:

migrating the lexical data to XML format result[s] in a resource from which various outputs can be derived, such as databases, websites, teaching and learning materials, interactive multimedia apps, archivable materials, and a solid basis for any further data work. XML is an ISO standard; it is guaranteed archive-quality; it enables both automated and manual diagnosis and manipulation of data; from it a range of derivative products can be made; while at the same time it is viewable in any text editor or browser and is human-readable.

It is possible to derive various kinds of documents that display the tagged information in different ways. Once suitably marked up, the XML document can be linked to using an Extensible Stylesheet Language for Transformations (XSLT) document to select and restructure structural elements, and a Cascading Style Sheets (CSS) document to define the display characteristics when viewed in an XML processor or web browser. For example, an

edition can be created showing lemmas and glosses only (a simple vocabulary list), or using different layouts, type faces, and colours to display the data for viewing or proofing. Figure 9 is a sample for the *tapana* entries discussed above (generated by Nathan in 2016).



Figure 9. Information structure display of the Reuther-Scherer dictionary

An XML-based approach has been adopted by several researchers working on Aboriginal language materials in Australia. Jansz (1998) is an analysis of the Warlpiri dictionary developed by Ken Hale and others (Laughren & Nash 1983), which Jansz and Manning encoded in XML (see also Jansz et al. 2001). They also developed software called *Kirrkirr* that graphically displays sense relationships within the dictionary. Corris et al. (2004) explore the utility of *Kirrkirr* for Warlpiri speakers. Henderson (2008) discusses how he encoded the 1930s fieldnotes of Gerhard Laves in XML; these were rather unstructured, unlike the semi-structured Reuther-Scherer dictionary. Thieberger (2016) uses XML representations to encode vocabulary lists collected in the early 20<sup>th</sup> century by Daisy Bates, with rather simpler content (just form-meaning pairs) compared to the richness of the entries we are dealing with for Diyari. Musgrave & Thieberger (2012) describe the significance of an XML-based approach in their work on Heath's dictionary of Nunggubuyu (Heath 1982) – we discuss in Section 5 why their method is not directly extensible to the Reuther-Scherer dictionary.

## 5. Further value adding and future prospects

Work on the XML files created by David Nathan continues as part of a research project funded by a *Leverhulme Emeritus Fellowship* for 2021-2023. This has involved adding further value by enriching the structure and adding additional tags:

(a) most of the material tagged as `<tabp> ... </tabp>` can be analysed as two different types of data:<sup>9</sup>

- sub-entries within an entry – all the relevant tags have been re-encoded to `<subentry> ... </subentry>` using the fact that the `<tabp>` for subentries are numbered within each `<entry>` block;
- contextual notes of various types, e.g ethnographic or mythological background, etymologies, comparative remarks – there are 1,732 of these and they have to be manually tagged for their note type, e.g. `<note type="ethno">...</note>`.

(b) the `<gloss> ...</gloss>` data strings sometimes contain subtypes of information, usually indicated by certain characters, optionally preceded by punctuation such as e.g. (for examples) or i.e. (for free translations of examples). I have identified material which comprises example sentences and assigned `<eg> ...</eg>` tags to them, and then within these added `<literal_gloss> ... </literal_gloss>` for literal glosses and `<free_gloss> ... </free_gloss>` for free translations of examples. Thus, the Reuther-Scherer entry:

2) #kalku durpana\$ = 'to twist or twirl reeds,' e.g. #kalku jerra durpanau billi dakala\$ = reeds there twist bag to weave, i.e. 'twist the reeds in order to weave a string-bag.'

becomes:

```
<subentry>
  <subentrynum num="2"/>
  <di>kalku durpana</di>
  <gloss>to twist or twirl reeds</gloss>
  <ex>
    <di>kalku jerra durpanau billi dakala</di>
    <literal_gloss>reeds there twist bag to weave</literal_gloss>
    <free_gloss>twist the reeds in order to weave a string-bag</free_gloss>
  </ex>
</subentry>
```

Unfortunately, this is difficult to do algorithmically and may require all 16,028 examples to be manually checked.

(c) some lexical entry glosses have semantic specifications, but these are non-uniformly represented (sometimes in brackets, sometimes as i.e., sometimes as ‘of ...’) – we have added `<semspec> ... </semspec>` for this information within the relevant `<gloss> ...</gloss>` tags (due to inconsistencies in the source file these have to be assigned manually, where applicable). An example is:

---

<sup>9</sup> David Nathan (p.c., 2021-08-15) points out that this was just an “empty tag” to capture and preserve indenting in the document it was derived from. In this case, layout on the page in the Reuther-Scherer dictionary is underspecified in terms of the document semantics, which are expected to be “filled in” by the reader. It needs to be explicitly spelled out in the XML representation.

80. #dunkana\$ (v.) = 'to come out, to go out (of a wurley, or through a hole, etc.), used of humans and animals; to come forth, to emerge (from a distance or from a place of hiding); to rise; to come forth (from seed)'

which becomes:

```
<entry lbl="80">
  <lemma>
    <di>dunkana</di>
    <pos val="v"/>
  </lemma>
  <gloss>to come out, to go out
    <semspec>of a wurley, or through a hole, etc.</semspec>
    <semspec>used of humans and animals</semspec>
  </gloss>
  <gloss>to come forth, to emerge
    <semspec>from a distance or from a place of hiding</semspec>
  </gloss>
  <gloss>to rise, to come forth
    <semspec>from seed</semspec>
  </gloss>
</entry>
```

As noted above, the entries, subentries, and examples in Reuther's original dictionary and Scherer's translation are somewhat randomly presented, and occasionally repeated or scattered throughout the text. For this reason, while an XML-based representation helps to improve access to and structure of the text, the poor lexicographic structure means that well organised outputs, especially for language learners, cannot be derived directly from the Reuther-Scherer document itself (cf. Musgrave & Thieberger 2012 for Heath's Nunggubuyu dictionary). Rather, what is needed is a kind of standoff index of the document that reorganises the information in the original, while also giving the modern spellings and analysis of the Diyari materials, drawing on subsequent research. An example mock up of an entry in a learner's reference dictionary that does this for *thapa-rna* (Reuther's *dapana* and *tapana*) is given in Figure 10.

*thaparna* vtr [R229, 3148]

(a) 'to drink'

*ngapa thaparna* 'to drink water' [R229-ex1, R3148-ex1]

*kumarri thaparna* 'to drink blood', *Ethn.* When he is first being made a man, the *wiyarru* opens his mouth widely and drinks the blood offered to him by his initiators. This also used to refer to drinking the blood that has been washed off a spear that a man has been killed with; young men are required to do this, in order to become fearless. Reuther also has the meaning 'to suck a wound' [R3148-ex 9]

*kipara thaparna* 'to drink urine', *Ethn.* men drink their own urine in order to end a friendship when planning to kill a friend [R3148-ex22]

*puarrna thaparna* 'to kneel down to drink' [R3148-ex18]

*thapatha thikarna* 'to come back to drink' [R3148-ex21]

(b) 'to suck'

*ngama thaparna* 'to suck the breast' [R3148-ex6]

*paya kapi thaparna* 'to suck out bird's eggs' [R3148-ex7]

*thuringyi thaparna* 'to suck marrow out of a bone' [R3148-ex8]

*manya thaparna* 'to suck a wound', *Ethn.* the *kunki* sucks puss and matter from a wound [R3148-ex3]

*muluru thaparna* 'to suck on caterpillars', *Ethn.* this is done so they become big and fat [R3148-ex19]

(c) 'to eat, slurp up, or chew on soft or semi-liquid food or fruit'

*parwa thaparna* 'to slurp up ground seed' [R229-ex6, R3148-ex2]

*kilthi thaparna* 'to eat fat or stew' [R229-ex4, R3124-ex 5]

*danyu thaparna* 'to eat *danyu* fruit'. *Note.* for soft fruit *thaparna* 'drink' rather than *thayirna* 'eat' is used. Other fruits are *kudnampira*, *mpiampia*, *nharimayi*, *ngalyaru*, and *ngurathikiri* [R3124-ex11]

*paḥyangari thaparna* 'to chew on the gum of the *paḥyangari* tree' [R229-ex9, R3148-ex 12]

*karna thaparna* 'to eat people', *Ethn.* the members of a *piya* revenge expedition eat the raw liver of their victim. [R229-ex2, R3148-ex14]

(d) 'to lick'

*miralu thaparna* 'to lick a coolamon', *Ethn.* when a man has no tobacco, he licks the dish that he last prepared tobacco in to ensure that some may soon become available. [R-ex 23]

(e) 'to kiss, touch with the mouth'

*marna thaparna* 'to kiss on the mouth'

*parru thaparna* 'to kiss a fish', *Ethn.* If no fish land in the net, a man goes down into the water, whistles into a hollow bone, and sings his *mura* song. The first fish to be caught is then kissed while the man has bread (made of seed) in his mouth, and allowed to swim again. This is expected to entice other fish to enter the net. [R3148-ex20]

*kira thaparna* 'to kiss a boomerang', *Ethn.* this is done before throwing to ensure that the boomerang hits its target. It is also done repeatedly when a boomerang is being made to see if it is straight. [R229-ex25, R3148-ex13]

Figure 10. Structured pedagogical dictionary entry for *thapa-rna*

This kind of index provides a useful reorganisation of the Reuther-Scherer materials for the Diyari community to employ in language revitalisation and learning, and indeed forms the basis for a language learning post on the Diyari language blog.<sup>10</sup> One of our goals during the

<sup>10</sup> <https://dieriyawarra.wordpress.com/2021/03/18/kararaya-yawarra-word-of-the-week-thaparna/>, accessed 2021-05-15

current project is to create various kinds of well-structured digital materials for learners that link to the Reuther-Scherer dictionary.

## 6. Conclusions

The 2,180 page Reuther-Scherer dictionary of the Diyari language is one of the most detailed attestations of an Australian Aboriginal language, and has great potential as a resource for both learning about the language and culture, as well as serving as a source for further anthropological and linguistic research. However, it suffers from a number of structural and representational deficiencies that make it difficult to use, inconsistent, and rather poorly presented. Our approach of adding XML markup to the digitised version of the dictionary means that it can become a much more useful and accessible information source. Pedagogical material can link to the XML file, and vice versa, and other types of documents can also be generated from it, ranging from simple word lists to more complex multimedia products. Marking up with XML in the ways described here involves enrichment of the original document's intended semantics by adding consistency, explicitness, and machine readability and processability. We do this by a combination of pattern-led algorithmic insertion and manual editing and insertion, on the basis of our knowledge and interpretations as linguists. This involves drawing on all other existing sources on the language and culture (such as documentation and recordings made by Austin and Hercus in the 1970s), as well as work with current native speakers and community members. The resulting enriched file can then be used as a platform for further development of various sorts of materials for different purposes and audiences.

The challenge ahead, of course, is to do the kind of work illustrated here in detail for one particular entry (*thapa-rna*) for the remaining thousands of entries, sub-entries, examples, and notes in the Reuther-Scherer dictionary. However, when completed, the dictionary and associated materials will be one of the richest information sources on an Australian Aboriginal language.

## References

- Austin, Peter. 1986. Diyari language postcards and Diyari literacy. *Aboriginal History* 10, 175-190.
- Austin, Peter K. 2014. And still they speak Diyari: the life history of an endangered language. *Ethnorema* 10, 1-17.
- Austin, Peter K. 2021. *A grammar of Diyari, South Australia*. Version 2.12. London: SOAS. [http://www.peterkaustin.com/docs/Austin\\_2021\\_Diyari\\_grammar.pdf](http://www.peterkaustin.com/docs/Austin_2021_Diyari_grammar.pdf) (accessed 2021-04-15)
- Corris, Miriam, Christopher Manning, Susan Poetsch & Jane Simpson. 2004. How useful and usable are dictionaries for speakers of Australian indigenous languages? *International Journal of Lexicography* 17, 33-68.
- Ferguson, Charles A. 1987. Literacy in a hunting-and-gathering society: The case of Diyari. *Journal of Anthropological Research* 43(3), 223-237.
- Gale, Mary-Anne. 1997. *Dhanum Djourra'wuy Dhawu : a history of writing in Aboriginal languages*. Underdale: Aboriginal Research Institute, University of South Australia.
- Ganter, Regina. 2019. Reuther, Johann Georg, Rev. (1861-1914). German Missionaries in Australia website, Griffith University. <http://missionaries.griffith.edu.au/biography/reuther-johann-georg-rev-1861-1914> (accessed 2021-04-15)

- Heath, Jeffrey. 1982. *Nunggubuyu dictionary*. Canberra: Australian Institute of Aboriginal Studies.
- Henderson, John. 2008. Capturing chaos: Rendering handwritten language documents. *Language Documentation & Conservation* 2(2), 212-243.
- Hercus, Luise. 2017. Looking at some details of Reuther's work. In Nicolas Peterson & Anna Kenny (eds.) *German Ethnography in Australia*, 115-135. Canberra: ANU Press.
- Jansz, Kevin. 1998. Intelligent processing, storage and visualisation of dictionary information. University of Sydney Computer Science Honours thesis.
- Jansz, Kevin, Christopher D. Manning & Nitin Indurkha. 2001. Kirrkirr: interactive visualisation and multimedia from a structured Warlpiri dictionary. *Literary and Linguistic Computing* 16(2), 135-151.
- Jones, Philip. 2007. *Ochre and Rust: Artefacts and Encounters on Australian Frontiers*. Kent Town, South Australia: Wakefield Press.
- Jones, Philip. 2011. Colonial Wissenschaft: German naturalists and museums in nineteenth century South Australia. In Peter Monteath (ed.) *Germans - Travellers, Settlers and their Descendants in South Australia*, 204-236. Adelaide: Wakefield Press.
- Jones, Philip G. 2012. Carefully sculptured words: the Lake Eyre toas in their time and place. *Journal of Museum Ethnography* 25, 35-54.
- Jones, Philip G. & Peter Sutton. 1986. *Art and Land: Aboriginal Sculptures of the Lake Eyre Region*. Adelaide: South Australian Museum in association with Wakefield Press.
- Kenny, A. 2013. *The Aranda's Pepa: An introduction to Carl Strehlow's masterpiece, Die Aranda-und Loritja-Stämme in Zentral Australien (1907-1920)*. Canberra: Australian National University Press.
- Laughren, Mary & David Nash. 1983. Warlpiri Dictionary Project: Aims, method, organization and problems of definition. In Peter Austin (ed.) *Papers in Australian Linguistics No. 15: Australian Aboriginal Lexicography*, 109-133. Canberra: Pacific Linguistics A-66.
- Musgrave, Simon & Nick Thieberger. 2012. Language description and hypertext: Nunggubuyu as a case study. In Sebastian Nordhoff (ed.) *Electronic Grammaticography*, 63-77 Hawaii: Language Documentation & Conservation Special Publication No. 4.
- Nathan, David. 2016. Report on Reuther Dieri dictionary data work: Completion of "Stage 2". Manuscript.
- Proeve, E.H. & H.F.W. Proeve. 1952. *A work of love and sacrifice : the story of the mission among the Dieri tribe at Cooper's Creek*. Tanunda: Auricht.
- Reuther, Johann G. 1981. The Diari vols 1-13. Translated by Philipp A. Scherer. Vol. 5 translated by T. Schwarzschild and L.A. Hercus. AIAS Microfiche No.2, Canberra: Australian Institute of Aboriginal Studies.
- Stevens, Christine. 1994. *White man's dreaming: Killalpaninna Mission 1866-1915*. Melbourne: Oxford University Press.
- Stockigt, Clara. 2017. Pama-Nyungan morphosyntax: lineages of early description, University of Adelaide, PhD dissertation.
- Thieberger, Nick. 2016. Daisy Bates in the digital world. In Peter K. Austin, Harold Koch & Jane Simpson (eds.) *Language, land & song: Studies in honour of Luise Hercus*, 102-114. London: EL Publishing.

Tindale, N. B. 1937. Report on trip to Germany to obtain information regarding Reuther's manuscript. Unpublished ms, South Australian Museum Anthropological Archives, Adelaide. AA 266/5/19/1.