TRAN ScriBING MATERIALS FROM
THE LivINg ARCHIVE OF ABORIGINAL LANGUAGES

General notes on transcription:

- Transcribe exactly what you see in the image, including capital letters and punctuation
- Use parentheses () when they are part of the original document, and square brackets [] to insert any notes about the transcription
- Transcribe exactly the line breaks as the author has written them
- Do not leave blank lines in the text
- Do not indent new paragraphs
- Do not transcribe page numbers
- Transcribe any hyphenations at the end of a line break as they appear in the image
- If there is an English translation, type it as it appears on the screen. If there appears to be an error in the text, include [sic] after the word in question.
- If unsure of the word put a [?] in the transcribe box, if unsure of many words use [?] of [] for each word, if unsure of letters in a word put a [?] for the letter. For example: lett[?]r.
- If a word is completely illegible, transcribe [illegible]
- If a word or sentence has been crossed out or erased put these words between < and > for example: <crossed out>.
- If there are annotations on the image, transcribe these and mark as [annotation]
- If there is any species or botanical information in the text, transcribe the text and mark the relevant word/s as [species] or [botanical]

Specific notes on transcribing Indigenous Australian languages:

- Aboriginal languages have surprisingly similar sound systems considering the huge number of languages covering a huge area. However there are some differences in spelling across the different languages.
- Most Aboriginal languages have the following letters:
  
  a b d e g l j k l m n o p r t u w y

- Some languages also have these combinations and special characters:
  
  dh lh nh th rr ng nj ny dj rd rl rn rt
  ḏ ḙ ṇ ṇ ṛ ṗ ṇ ṇ ṛ ṗ
  á é í ó ú ü
You’ll notice that some English letters are missing (like c, f, q, s, v, x, z), though at least one language uses h as a consonant (separately to its use in combinations such as th or dh).
Some of the letters have the same sound as they do in English; others do not.
The combinations of letters may be quite different to English, some words are very long, some are hyphenated.

Transcribing special characters
Most of the texts do not include any special characters, and can be done using a standard ASCII keyboard.
If you are transcribing special characters, it is important to use Unicode, which ensures that a character appears correctly on any system. If a non-Unicode character is used, the character may appear as a ? or a strange symbol when displayed on certain systems or programs.
For example there is a difference between putting an underline beneath a letter – e.g. ḷ - and creating a Unicode character for an underscore ḷ – e.g. ḷ - they may look similar, but computers process them differently.
If there are only a few special characters to include, you can use the online Unicode keyboard available at http://ipa.typeit.org/full/, and type directly into the text box, selecting the appropriate character as needed, then copy the entire text into the DigiVol window.
If you expect to transcribe longer texts or several short texts which have special characters (particularly the Yolngu languages of northeast Arnhem Land), you may prefer to download a special Unicode keyboard. The instructions for installation and use are available at http://ausil.org.au/node/3719

For any enquiries, please contact livingarchive@cdu.edu.au

Background to the project:
The Living Archive of Aboriginal Languages is a digital archive of endangered literature in Australian Indigenous languages of the Northern Territory. It was developed out of a concern for the fate of the many thousands of resources created in remote Indigenous schools during the era of bilingual education in the Northern Territory. The fluctuating fortunes of the bilingual education policy meant that these materials were vulnerable, so with funding from the Australian Research Council, a team of academics created the archive and began populating it with digital versions of these materials. Since its origins in 2012, it has expanded to other communities in the NT which didn’t have bilingual programs, and now includes over 3500 items in 50 Indigenous languages, all available under a Creative Commons license from the website at http://laal.cdu.edu.au/. The materials are available as PDFs and as text files, the latter produced using optical character recognition software, though this does not always process the textual content correctly, often leaving errors in the texts. Thanks to the DigiVol volunteers these texts can be edited and uploaded to the Archive for people to read and enjoy.