

規劃數位族語資料庫

一、前言

台灣原住民族是以口述為文化傳承的族群，族語即為其文化傳承的重要載體。時至今日原住民族文化及社會結構已有很大的改變，加上失去教育的主體性及自主權利，往往造成在學校受的教育越多，反而離母體文化越遠的情況，這也正是其語言與文化急速流失的原因之一。有鑑於此，教育部與原住民族委員會於民國 94 年 12 月 15 日會銜發布「原住民族語言書寫符號系統」，即希望能更有系統的將族語文字化，以因應在現代教育制度下族語學習的困境，除了推動族語課程、教材的編撰，另一方面則是鼓勵族人能以自己的觀點，書寫、創作、記錄並保存族群文化的內涵。

從歷史的角度來檢視，原住民族的語言開始有文字的記錄約莫是從荷據時期的「新港文書」開始，到日據時期之後有多位日本語言學者如小川尚義、淺井惠倫、土田滋、森口恒一、月田尚美、野島本泰為台灣南島語言研究貢獻良多(李王癸, 2009)。而民國 57 年基督教聖經公會也以羅馬拼音將聖經翻譯成族語版，民國 83 年則由教育部亦委托中央研究院李王癸教授編訂「中國語文台灣南島語言的語音符號系統」。台灣原住民族語言雖因口傳方式保存不易，所幸尚有外來的傳教士、學者的記錄，使許多珍貴的語料尚得留存至今，但由於早年影音記錄及保存的技術尚不發達，可能造成部分資料因此損壞或遺失，十分可惜。

爰此，如何應用現代技術，將已留存的族語文獻資料，特別是早期由族人自行書寫、錄音、錄影的資料彙整，並以數位化方式保存，將影音的部分加以轉譯成文字，以利後續研究或教學使用，便是未來我們需要努力的課題。而本年度我們將建立一套「族語文獻調查表」用於族語文獻蒐集，之後再彙整成「數位族語資料庫語料清單」，並草擬「原住民族語言研究發展中心數位族語資料庫建置評析報告」，此報告預計會先以三年期程作為規劃並提出建議，首先會參考相關網站(如台灣原住民族數位典藏資料庫、台大台灣南島語多媒體語料庫、中央研究院台灣南島語語料庫等)提出初步規劃，也請中心顧問協助指導，於 3 月至 10 月間每月至少召開一次研商會議，亦在研究過程中進行初步的族語文獻調查(包含文字、錄音或影片等資料形式)，最後提出相關報告，以供後續族語資料庫建置參考。關於詳細的工作規劃內容將於下文說明。

二、106 年度數位族語資料庫建置工作

本計畫 106 年度的主要工作項目有三大項，一為調查及彙整族語文獻(以 1980 年以前之文字、錄音或影片為優先)，並製作「數位族語資料庫清單」，供

分期建置語料庫使用；二為召開研商會議並草擬出「原住民族語言研究發展中心數位族語資料庫建置評析報告」，內容應包含現況分析及建置目的、階段性建置任務及目標、資料庫網頁設計圖。進一步的工作內容初步規劃如下：

1. 調查及彙整族語文獻：

(1) 研究範圍及限制：

因研究經費及人力限制，現階段主要調查文獻方式是以網路搜尋及行政機關典藏文獻為主，另規劃向本中心其他工作小組（包含原住民族各族）調查是否有相關文獻資料，或請其提供可訪談對象，以利初步蒐集相關資料。

此外，將設計「族語文獻清單調查表」，除了便於後續資料的建置，更希望此計畫能透過網路推廣，以利族語文獻調查的工作可以持續進行，並能網羅更多文獻，在有限的範圍內能夠將調查成果極大化。

(2) 調查方式：

- A. 查詢各大語料庫網站並以此作為基礎，向資訊專家學者做初步的諮詢，進一步設計「族語文獻調查表」格式，俾利後續調查及建檔工作進行。
- B. 查詢網路文獻資料、行政機關之庫藏。
- C. 參與中心其他族語小組工作會議，並請各族族語老師協助提供相關文獻或資訊，再聯絡相關教會、部落、協會或族人，前往訪談並收集資料。
- D. 於網路推廣族語文獻調查工作，徵求更多族人提供相關資訊（料）。
- E. 蒐集資料類型包括文字、錄音或影片等三大項目，並以1980年以前之資料為優先。
- F. 將上述收集之文獻彙整為「數位族語資料庫語料清單」。

2. 召開研商會議並草擬資料庫建置評析報告：

(1) 會議參與對象：邀集機關、語言學者、程式設計師、耆老等代表召開數位族語資料庫研商會議。

(2) 相關流程及規劃：

- A. 參考其他現有資料庫並規劃族語資料庫雛形。
- B. 預計以三年的期程作為第一期資料庫建置規劃。
- C. 本年度自三月起到十月間，每月定期召開一次研商會議，並就會議小組成員的建議，檢視及修正資料庫建置架構、內容。

- D. 撰寫「原住民族語言研究發展中心數位族語資料庫建置評析報告」，並含現況分析及建置目的、階段性建置任務及目標、資料庫網頁設計圖。

綜合上述，本年度工作執行流程：



- 相關語料庫網站及文獻搜尋
- 設計「數位族語資料庫文獻調查表」並於網路推廣族語文獻調查工作
- 文獻調查及蒐集
- 中心其他各族工作小組及田野訪談
- 提供現階段已蒐集之「數位族語資料庫語料清單」

- 規劃族語資料庫雛形
- 以三年的期程作為第一期資料庫建置規劃
- 於三至十月每月召開研商會議，邀集機關、語言學者、程式設計師、耆老等代表參與會議，以修正資料庫建置架構、內容。
- 初步建構數位族語資料庫頁面，及示範數位化文獻一則
- 草擬「原住民族語言研究發展中心位族語資料庫建置評析報告」