

原住民族語言研究發展中心

108 年度「數位族語資料庫維護與資料建置」執行計畫書

Akiw 研究員

一、前言

聯合國教科文組織 (UNESCO) 2009 年公布了「世界瀕危語言地圖」，其中臺灣有 5 種語言被列入極度瀕危，10 種語言被列入滅絕等級。依原民會公告之「107 年度原住民族瀕危語言復振補助計畫」，列入瀕危語言的語別有卑南語、撒奇萊雅語、噶瑪蘭語、邵語、拉阿魯哇語、卡那卡那富語、賽夏語、茂林魯凱語、萬山魯凱語、多納魯凱語。

因此除了振興語言之外，亦需要積極的保存語言，正如 2016 年 6 月公布實施的原住民族語言發展法，也將「建置原住民族語言資料庫」視為一項重要的工作。Crystal (2010:349) 提到保存語言的方式，即是從語言記錄開始，以錄音 (聲音與影像)、分析、文字記錄來保存語言，也能夠用於教育。而當語料被記錄後，我們需要有一個整合性的平臺來收錄這些語料，整理、分類後，讓更多人能夠再應用，達到語料保存的目的。

我們查詢國內外族語保存相關的資料庫發現，國外早已有像是 AILLA、PARADISEC 這種網站，平臺提供收集多媒體語言資料及檔案，一般人都可以將所持有的媒體語料提供給該站，就會以數位化的方式被建置，甚至有分級訪問權限的制度，因此吸引許多研究人員或是族人願意提供資料。而這些資料也歡迎其他研究者協助分析、轉譯，讓資料庫收錄的越來越豐富。反觀國內，目前尚未有這種形式的資料庫，因此這也是我們努力的方向。

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、野島本泰為臺灣南島語言研究貢獻良多 (李壬癸, 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 1992 年則由教育部出版，委託由中央研究院李壬癸教

授編訂臺灣南島語言的語音符號系統。臺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，特別是將早期的族語資料彙整，並以數位化的方式保存，是首要之務。將影音的部分加以轉譯成文字，以利後續研究或教學使用，便是我們需要努力的課題。

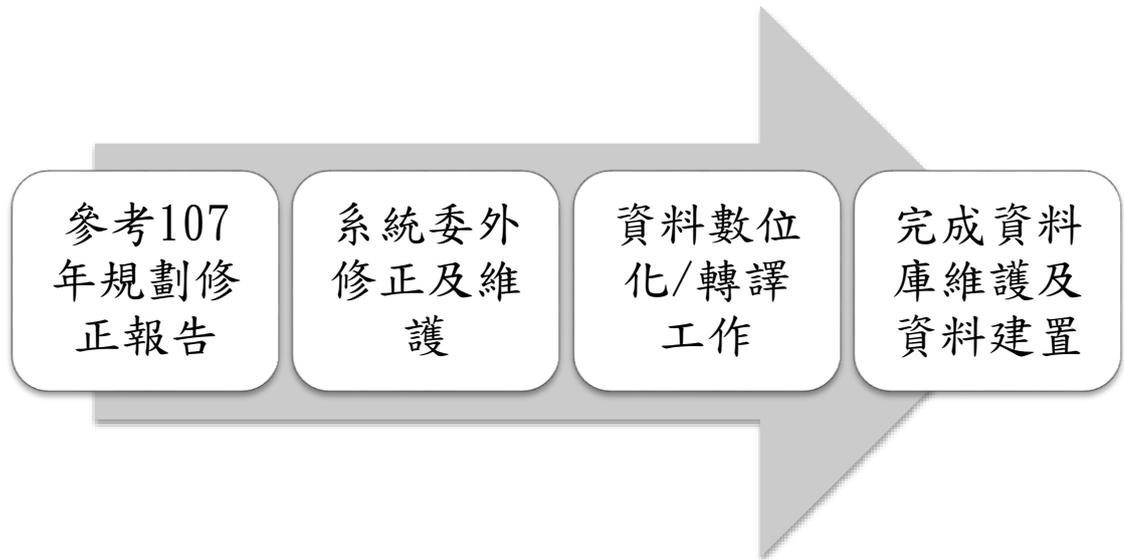
去（107）年語發中心已完成「數位族語資料庫第一期建置」，其中包含基本頁面、檢索功能、系統後臺，並完成逐字稿、新聞編輯、典藏資料及臺灣原住民族研究書目建置等功能，另一方面也進行部分資料數位化、族語新聞轉譯為文字等資料建置工作，更於期末辦理資料庫系統及介面的檢視及修正會議，最後提出「第一期建置內容及規劃修正報告」。

而本（108）年度我們主要的工作是依去（107）年資料庫的修正規劃，除了系統的調整及檢索功能優化外，亦開始轉譯及建置所收集的族語資料，並繼續進行族語新聞的文字轉譯和原住民族語言研究目錄的收錄。預計將分階段完成錄音帶 5 卷及 12 篇族語新聞的轉譯，也正積極尋求與語推人員合作族語新聞轉寫的可行性，以利產出更多族語文字，供未來建置語料庫使用。亦期望此族語資料庫能做為「財團法人原住民族語言研究發展基金會」典藏原住民族語料及建置原住民族語言資料庫工作的基礎。

二、執行目標

1. 數位族語資料庫維護及修正：依 107 年資料庫修正報告為基礎，進行系統系頁面的修正及優化，並維護資料庫的運作。
2. 數位族語資料的建置：進行 5 卷語料及 12 則族語新聞之建置。

三、執行方式：



(一) 參考 107 年規劃修正報告

1. 依修正報告內容提出系統修改細節需求書。
2. 辦理研商諮詢及工作會議。

(二) 委外進行系統修正及維護工作

1. 系統修正需求書。
2. 委外進行資料庫介面及系統修正工作。

(三) 資料數位化及進行轉譯工作

1. 資料進行數位化及修復。
2. 請專業人才協助族語新聞進行轉譯及編輯工作。

(四) 完成資料庫維護及資料建置

1. 確保計畫期間資料庫之運作。
2. 建置語料。
3. 提出本年度建置內容報告。