

台灣客語語言典藏的現況與資源整合：

從學術研究角度看客語語言典藏的未來展望

江敏華

中央研究院語言學研究所

一、前言

語言政策在臺灣一直是一個相當敏感的問題，不管是對於過往歷史的得失評價，或是對於傳承現況的樂觀程度，以及對於展望未來所應投注的積極心力，在在都觸動著每一個關心臺灣語言發展者的神經，也關係著每個人心中對於語言政策那把尺的定位。臺灣本土語言的活力與傳承，儘管有著重重的危機與障礙；關心臺灣本土語言發展的人，在政策面上也或許仍有許多分歧，短期之內無法達成完全的共識，然而，放任眼前險阻而一籌莫展，畢竟不是解決問題的正面態度。現階段我們所能做、也必須做的，便是尋求如何在剪不斷、理還亂的歷史糾葛之外，找到一個語言政策光譜兩端的彼此都可以共同努力的切入點，為增強臺灣語言活力並維持其傳承，找到一個深耕實作的平台。如此，才是臺灣語言發展之幸，也才能真正展現有容乃大的多元意涵。以學術研究為立足點的語言典藏，或許便是提供這樣一個平台的可能方向，也或許，無為而為，共識反而於焉形成。

二、語言典藏與語言數位典藏

自古以來，大凡人類對於他們認為有價值的、超越其時代成就的人、事、物，都有相當程度的「保存」、「流傳」的慾望，所謂「藏諸名山，傳之其人」。屬人者，大至影響深遠的哲學家、思想家、政治家、軍事家、藝術家，或立史傳書之，或建廟祭祀之，或保存故居紀念之；小至己身所出之父母祖先，或撰家譜明其源流，或立牌位祭祀感念，或保留近身之物恆久追思。屬事者，則如智慧的傳承與累積，或是宗教文化與傳統習俗的保存與發揚。屬物者，舉凡鬼斧神工的建築與工藝、展現人類心靈巧思的文學與藝術創作等等，更是人類亟欲珍藏的目標。乃至近現代，針對新發現而尚不知如何最佳運用的地球景觀資源，而有了國家公園

的設置；聯合國教科文組織（UNESCO）針對足以表徵人類文明進程與地球生命進化歷程的文化與自然遺產，通過了《保護世界文化與自然遺產公約》，從觀念上與實際作為上，對世界文化遺產進行保存與維護。凡此種種，都可以視為是人類對於自身所生存世界的物質與精神文明加以「典藏」所作的努力。

「語言典藏」最基本的出發點亦是如此。「語言」在很長一段時間被視為「人之異於禽獸者」以及「人之所以為人」的特質之一，而今我們雖然不能非常肯定地如此堅稱，但語言仍是人類擁有如此豐碩的精神文明的重要因素之一。世界上每一種語言，都代表一個文化生命，聯合國教科文組織「世界遺產委員會」於二〇〇一年在自然遺產與文化遺產之外，新增「口述與無形人類遺產」一類，其中「語言」赫然在列。教科文組織於二〇〇三年亦決議，將通過國際共同合作，保護以「語言多樣性」為優先之人類文化多樣性。由此可見，無論在觀念上或實質上，語言之必須且值得「典藏」，乃是世界時勢之所趨。

其實早在「語言典藏」這個概念具體化之前，也並不乏將語言的某些形式流傳下來的方法，例如透過文字，雖不全面，但我們仍得以觀察到古代語言在語音、詞彙、語法甚至語言流通使用的概貌。以客語而言，透過漢字或羅馬字的紀錄方式，我們有以客語為基礎語言創作的文學作品，也有各種流傳海外的傳教士客語文獻，如客語聖經、客語讀本、客語醫書、客語文法等，還有各種不同類型的客語字詞典工具書等等，可以稱為語言的文本典藏。然而，語言的特性之一在於「聲音」，話一出口，便隨風而逝¹，要「典藏」這些看不到、摸不著的語言元素，便需要比文本典藏更具技術門檻的工具。語言的「數位典藏」，便具有文本典藏所無法達到的功能。

語言的數位典藏，簡單地說，就是儘可能地蒐集語言的各種文字、聲音、圖片影像的素材，加以數位化，並妥善地保存。它不但可以取代文本典藏原有的功能，同時還能兼顧文本典藏所無法儲存的聲音或影像資料。因此，就典藏內容而言，一個語言數位典藏計畫，可以涵括：一、文獻典藏：包括各種以該語言書寫

¹ 即使是手語，雖不是以聲音為媒介，但仍具有「隨風而逝」的特性。

的歷史文獻、文學作品、童蒙教材、長篇語料與字詞典工具書等。二、口語典藏：包括各種不同口音的語言音檔、各種目的型導向的大型有聲語料庫、平衡語料庫等。三、語言的社會資訊：包括語言的社會使用、地理分布與歷史變遷等等。

然而，語言數位典藏，若僅止於「收藏」資料，而未能充分利用數位化資料存取快速、整合容易，以及便於作各種「後設」加工的特性，實在是辜負了現代科技發展所帶來的便利。隨著現代社會透過網路的無遠弗屆，以及科技人性化的設計與互動式學習，語言的數位典藏，事實上可以讓語言典藏的效益不僅僅在於「保存」，它還可以「活化」這些蒐集而來的語言資訊，透過典藏資料庫與使用者之間的互動，發揮母語教學、社會推廣或是學術研究等等的綜效。如此，便可以消除某些認為語言一旦需要典藏，便是行將就木、逐漸步入死亡階段的疑慮。

三、從學術研究角度看客語語言典藏的努力方向

臺灣雖小，卻是世界資訊大國，資訊數位化的程度與網路普及化的程度，在世界各國中名列前茅。現有的網路資源中，事實上也不乏將臺灣客語的文本或音檔數位化並供使用者查詢使用者。檢視這些網路資源，可以發現，在內容方面，資料尚稱豐富，詞彙、諺語、童謠、詩詞、歌詞、短文、民間故事、字詞典等，頗能面面俱到；惟童謠諺語等重複性亦頗高，廣度有餘，深度不足。在典藏性質方面，則以母語教學和社會推廣為主，以學術研究為目的的語言典藏亟待整合相關資源並投注心力。

以學術研究為目的的客語語言典藏，必須包含研究客語（包含各種次方言）所必須具備的語音、詞彙、語法的資料，大量足供分析的口語素材，以及語言的社會使用、地理分布與歷史變遷等資訊。更重要的是，這些語言素材必須組織良好（well-organized）、提供足夠的後設語言標記，建置成易於存取與整合的資料庫，並提供可供瀏覽與查詢檢索的界面。

在這樣的理想下，臺灣客語語言典藏的目標應包括：

一、 建立臺灣各地客語次方言的語音及詞彙資料庫

臺灣客語的次方言，經過過去幾年不斷地調查、整理與比較，可以歸納為北部四縣、南部四縣、海陸話、大埔話、詔安話與饒平話等。每一個次方言，都應有一個語音（包含聲音檔案及語音文字紀錄）資料庫與詞彙資料庫。這些資料庫，還必須建立其網路使用介面與查詢系統等資源分享機制。

二、 建立各式客語辭典的資料庫

客語現有各種類型的紙本字典，包括以特定方音為基礎的辭典，以解釋客語特殊詞彙語意的辭典，以及以雙語對譯為主的客華、客英、客法辭典等。字典、辭典的功能在於檢索，將這些資料數位化並建置成資料庫，不但可加速檢索的效能，更可以整合各種不同類型的辭典資料庫，使之得以相互聯結為更大的客語辭典資料庫。

三、 建立流傳海外的傳教士客語文獻

早年傳教士來到東方，為了更有效的推廣宗教，努力學習當地人的語言，編寫讀本或文法書、翻譯聖經、流傳醫書等，留下了不少描述當地語言或以當地語言為書寫基礎的文獻。客家話亦有不少這樣的傳教士文獻。以筆者所知，巴色會教會（Basel Mission, 或 Basel Evangelical Missionary Society）便收藏了為數不少的客家話出版物，包括客語聖經、客語讀本、客語醫書以及客語文法等。蒐集各種流傳海外的傳教士客語文獻，解讀、數位化並加以標記，使之成為可進一步分析的珍貴語言材料，便是語言數位典藏的重大貢獻。

四、 建立臺灣客語的長篇語料庫

語言的分析，不論是某特定詞彙的使用特性，或是某項語法功能的使用情況，在在都需要大量的口語或文本資料。然而，長篇語料取得不易，往往需要長時間的累積才能達到可觀的數量。此外，長篇語料取得後，還需要經過解讀、整理，加入語音標記、詞彙關聯與詞義解釋等，才能供研究之用。因此，整合各方資源，集眾人之力來建立臺灣

客語的長篇語料庫，才能收事半功倍之效。

五、 建立臺灣客語的社會資訊系統

要了解一個語言，除了研究其語音、詞彙、語法的規則外，還必須了解它在社會上的使用情況，包括使用人口、分布地區及其變化、歷史變遷、世代差異、共時變異或語言接觸等問題。這樣的研究同樣也必須耗費極大的心力才能獲得非常微小的成果，而所獲得的微小成果還必須累積起來，諸如歷史變遷、地域變化、世代差異等等的課題才能持續進行。因此，一個臺灣客語的社會資訊系統，是「典藏」與「研究」相輔相成的，社會資訊系統讓我們得以進行需時間累積才能進行的研究，而利用社會資訊系統所進行的研究，又進一步豐富社會資訊系統的內容。

無可諱言，上述每一項理想目標的背後，都是一個非常龐大而艱鉅的工作，必須有陣容堅強的工作團隊，並且可能耗時數年才能完成。其中的工作內容包括廣泛的資料蒐集、嚴謹的校對，需要客語語言學專業人士進行的語言標記工作，需要資訊技術人士支援的數位化工作；此外，還有資料庫與檢索系統的建置、程式的開發以及原始材料智慧財產權的授權等等問題。然而，臺灣並非沒有這樣的條件來成就此事。我們有專責推動客家事務的行政院一級單位「行政院客家事務委員會」，客語使用者仍有相當數量，尚未達到瀕臨滅絕的程度；各大專院校也有為數不少的語言系所與臺灣語言相關系所，培養出若干客語語言學專業人士；在資訊技術上，臺灣資訊數位化的程度、網路普及化的程度，以及資訊業人才的質與量，都讓我們有信心地認為，客語語言典藏，尤其是以學術研究為目的的客語語言典藏，是一個可以努力的方向。更重要的是，透過這樣的資源整合，臺灣還可能成為世界客語與客家文化研究中心，在本地化的努力之外，也同時尋求國際化的契機。

