

語言分佈 GIS 系統建置數位化工作流程簡介

更新日期：2005/12/05

計畫單位：中央研究院語言學研究所

計畫名稱：語言典藏計畫

計畫簡介：

閩南語和客家話是漢語的主要方言，是重要的語言資產，主要分佈於福建南部、廣東、台灣與東南亞，但受到學校教育、媒體大量使用國語（普通話）的影響，這二種語言能使用的人口有越來越少的傾向，成為相對的弱勢語言，亟待研究與保存。然而目前學界尚未有詳細標記的閩客語文獻語料庫，致使研究上，往往事倍功半。

臺灣人口流通量大而頻繁，語言接觸日益密切，語言生態丕變，方言中的「地區變體」與「社會變體」之消長分合，變化快速。近年方有學者開始積極調查繪製臺灣地區語言地圖，然電子語言地圖的繪製還在起步階段，展現語言分佈情況的語言地圖更付之闕如。

該計畫為中央研究院「語言典藏」分項計畫「漢語典藏與典藏架構」的五個子計畫之一，擬以大眾文學之劇本、歌仔冊二種文體為範圍，建立閩南語、客家語語料庫。並以閩客雜居的新竹縣新豐鄉為對象，調查居民用語，研究閩客用語交互之影響。從歷史語言與語言分佈兩點切入，結合文獻語言與生活語言，進行語言標誌，建置閩客語語料庫、詞彙庫與語言分布地理資訊系統，為學界提供有力的研究工具。

由於新竹縣新豐鄉是閩客雜居的鄉鎮，所以該計畫以新竹新豐鄉為範圍，進行語言分佈的調查研究，發展語言分佈地理資訊系統。

數位化工作流程說明：

該計畫的數位化作業，大致依照下列六項步驟進行，依序分別為：一、訪談及錄音；二、街路定位及數位化地圖；三、問卷輸入；四、錄音備份；五、圖層製作；六、匯入資料庫。茲分別介紹如次。

一、訪談及錄音

以新竹縣新豐鄉為範圍，進行語言分佈的調查研究，調查以訪談的方式進行，藉由問卷記錄訪談者平日對談所使用的語言，並且製作圖卡及字卡，請受訪者依照圖卡及字卡發音並錄音。但因在訪談過程中，受訪者不易同意訪談，故先拜訪鄰長，再由鄰長陪同前往，取得受訪者信任，進行訪談與錄音。（圖一至圖四）



圖一：訪談及錄音之情形（示範者：呂奇蓉、鄭月霞、郭彧岑）



圖二：訪談所使用之圖卡（範例）

寫字	關門
查某	迤迤
血	桃園

圖三：訪談使用之閩南語字卡（範例）

朋友	講話
胃	肥
時	四

圖四：訪談使用之客家語字卡（範例）

在問卷設計上細分為受訪者對不同對象所使用之語言，例如與長輩及晚輩談話的語言或有不同，問卷調查以『一戶家庭』為語言分佈調查的單位，以家庭最常使用的語言作為該家庭的語言。（圖五及圖六）

新豐鄉閩客方言地理資訊系統建置基本資料表

訪問時間： 94 年 月 日 時 分 GIS地理資訊：
地址： 村 鄰 路/街 段 巷 弄 號之 樓之
姓名： 出生： 年 月 日 性別： <input type="checkbox"/> 1. 男； <input type="checkbox"/> 2. 女
出生地： <input type="checkbox"/> 1. 新豐鄉； <input type="checkbox"/> 2. _____
職業： <input type="checkbox"/> 1. 農林漁牧； <input type="checkbox"/> 2. 工； <input type="checkbox"/> 3. 商； <input type="checkbox"/> 4. 教； <input type="checkbox"/> 5. 公； <input type="checkbox"/> 6. 其他； <input type="checkbox"/> 8. 無（退休）
教育程度： <input type="checkbox"/> 1. 無； <input type="checkbox"/> 2. 國小； <input type="checkbox"/> 3. 國初中； <input type="checkbox"/> 4. 高中職； <input type="checkbox"/> 5. 專科； <input type="checkbox"/> 6. 大學以上
族群： <input type="checkbox"/> 1. 客家人； <input type="checkbox"/> 2. 閩南人； <input type="checkbox"/> 3. 大陸各省市； <input type="checkbox"/> 4. 原住民； <input type="checkbox"/> 5. 其他 _____
1. 請問您住的這個地方，地名叫什麼？
2. 您在本地住了多久了？
3. 您住過其他地方嗎？
4. 您會哪些語言？ <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 流利程度： <input type="checkbox"/> 流利 <input type="checkbox"/> 流利 <input type="checkbox"/> 流利 <input type="checkbox"/> 流利 <input type="checkbox"/> 還可以 <input type="checkbox"/> 還可以 <input type="checkbox"/> 還可以 <input type="checkbox"/> 還可以 <input type="checkbox"/> 會聽不會說 <input type="checkbox"/> 會聽不會說 <input type="checkbox"/> 會聽不會說 <input type="checkbox"/> 會聽不會說
5. 您工作時常用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____
6. 您祭祖的時候使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____
7. 您跟長輩對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） 跟爸爸： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟媽媽： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟祖父： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟祖母： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟外祖父： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟外祖母： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____

圖五：使用之問卷範本—第一頁

8. 您跟配偶對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 配偶是： <input type="checkbox"/> 1. 客家人； <input type="checkbox"/> 2. 閩南人； <input type="checkbox"/> 3. 大陸各省市； <input type="checkbox"/> 4. 原住民； <input type="checkbox"/> 5. 外籍 _____
9. 您跟兄弟姐妹對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） 跟哥哥： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟姊姊： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟弟弟： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 跟妹妹： <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____
10. 您跟子女對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____
11-1. 您跟媳婦對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 媳婦是： <input type="checkbox"/> 1. 客家人； <input type="checkbox"/> 2. 閩南人； <input type="checkbox"/> 3. 大陸各省市； <input type="checkbox"/> 4. 原住民； <input type="checkbox"/> 5. 外籍 _____
11-2. 您跟女婿對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____ 女婿是： <input type="checkbox"/> 1. 客家人； <input type="checkbox"/> 2. 閩南人； <input type="checkbox"/> 3. 大陸各省市； <input type="checkbox"/> 4. 原住民； <input type="checkbox"/> 5. 外籍 _____
12-1. 您跟孫子女對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____
12-2. 您跟外孫子女對話使用什麼語言？（若複選，依常用順序標示 1, 2, 3, 4 等） <input type="checkbox"/> 1. 客語 <input type="checkbox"/> 2. 閩南語 <input type="checkbox"/> 3. 國語 <input type="checkbox"/> 4. _____
謝謝！現在有一些簡單的字詞，請您說一說。我們會錄音以便分析。
1. 請從 1 數到 10 2. 請看圖卡或字卡說一說

圖六：使用之問卷範本—第二頁

二、街路定位及數化地圖

因為在地理資訊系統上製作語言分佈圖，需要每個訪談地點的地理座標，因此將資料整理為地理資訊系統所需之檔案，有以下二種方法。

第一種方法為採用衛星自動定位，衛星定位儀器曾經採用 GARMIN eTrex Vista，由於誤差值較大，現在採用更先進設備 Trimble ProXH，精準度在一公尺以內，該儀器是在現場取得 X 及 Y 座標，以無線藍芽傳輸該地點的地理座標至筆記型電腦，電腦藉由程式直接寫到該家戶的通訊地址。但此方法在建築物密集的地方，難取得衛星訊息。(圖七及圖八)



圖七：使用之衛星定位儀器（左：為現在使用之儀器）



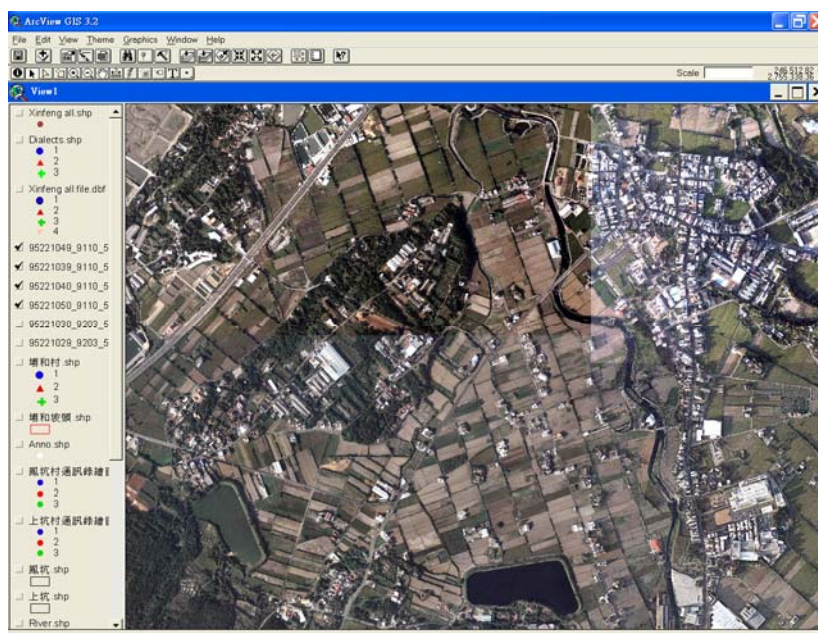
圖八：現場實地進行定位（示範者：鄭錦全、張智傑）
（拍攝者：黃菊芳）

第二種方法為利用已經向量化之航照地圖來尋找坐標，其工作是由二位助理擔任，一位助理使用 ArcView 軟體開啓已向量化之航照圖，於航照圖上找出受訪者住家位置，點選取得該位置 X 及 Y 座標。然後將該地點之 X 及 Y 座標唸給另一位助理聽，再由該助理負責將其資訊輸入另一電腦。(圖九及圖十)

第二台電腦所顯示的是從戶政機關取得的受訪者通訊錄所製作成的 Excel 檔案，藉由航照圖定位，將 X Y 座標輸入。之後 Excel 檔案再轉成 DBF 檔案，以 ArcView 軟體開啓，以便製作語言分佈之空間資訊系統。(圖十一)



圖九：確認受訪者門牌地址及地理位置（示範者：林千慧、郭彧岑）
（拍攝者：黃菊芳）



圖十：使用 ArcView 軟體開啓之航照圖

	A	B	C	D	E	F	I	J	R	S	T
1	總編號	編號	村	鄰	路	巷	號	樓	語言代碼	x	y
705	547	22	重興村	11	建興路二段			766	1	247867.65	2754912.44
706	548	23	重興村	11	建興路二段	768		2	1	247871.92	2754906.95
707	549	24	重興村	11	建興路二段	768		4	2	247881.68	2754908.78
708	550	25	重興村	11	建興路二段			770	2	247860.94	2754919.76
709	551	26	重興村	11	建興路二段			770	3		
710	552	27	重興村	11	建興路二段			772	2	247863.38	2754927.69
711	553	28	重興村	11	建興路二段			774	2	247862.77	2754930.74
712	554	29	重興村	11	建興路二段			774	2		
713	555	30	重興村	11	建興路二段			776	2	247860.94	2754936.83
714	556	31	重興村	11	建興路二段			778	2	247860.34	2754944.15
715	557	32	重興村	11	建興路二段			778	2		
716	558	33	重興村	11	建興路二段			780	1	247860.34	2754950.25
717	559	34	重興村	11	建興路二段			782	1	247856.68	2754950.86
718	560	35	重興村	11	建興路二段			782	1		
719	561	36	重興村	11	建興路二段			784	1	247854.85	2754960.61

圖十一：輸入之 Excel 檔案部分信息

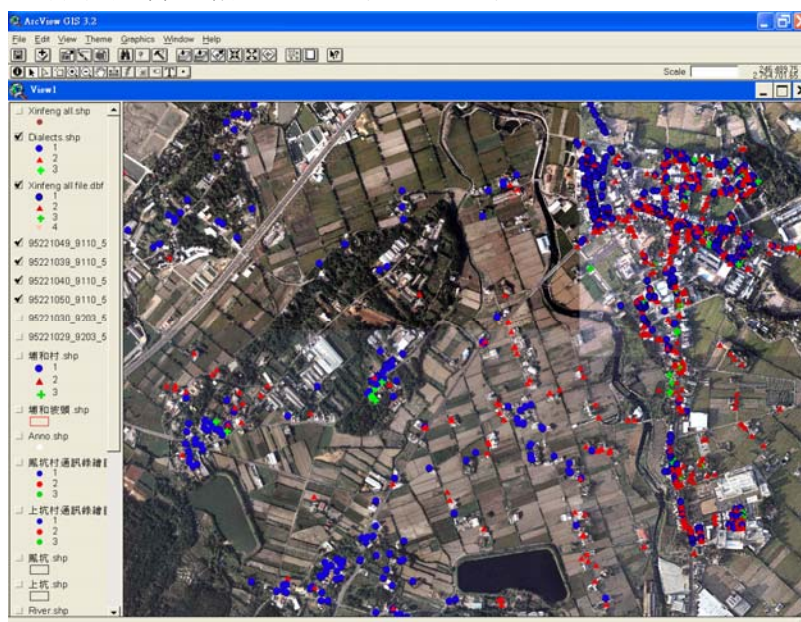
三、問卷輸入

將問卷輸入電腦，並建置檔案，便於使用者在使用語言分佈空間資訊系統時，藉由點選資料，能更了解新竹新豐鄉語言分佈之情況。(圖十二及圖十三)

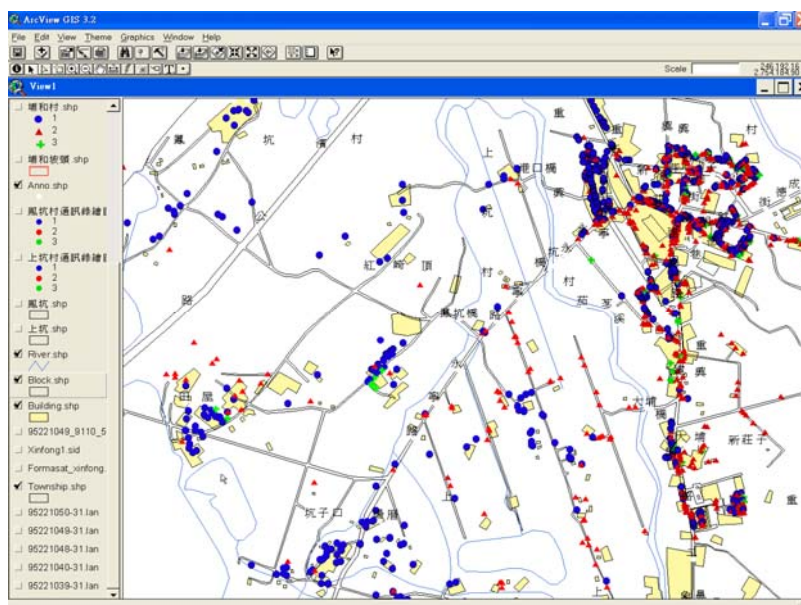


圖十二：問卷輸入之工作狀況（示範者：黃菊芳）

將語言分佈狀況製作成點狀圖，可分為客家語－四縣腔、客家語－海陸腔、閩南語及其他語言，分別以不同符號作為標示；再與其他相關資料分別製作成圖層，其他相關資料包含：已調查地點、建物、道路、河流、土地利用、鄉鎮界及航照影像等資料。分別製作圖層目的是為了讓不同的使用者，能依據自身的使用需求，選擇點選所需的圖層介面。使用者可以藉由下載免費的 ArcRead，加以使用已製作好的語言分佈的情況。（圖十五及圖十六）

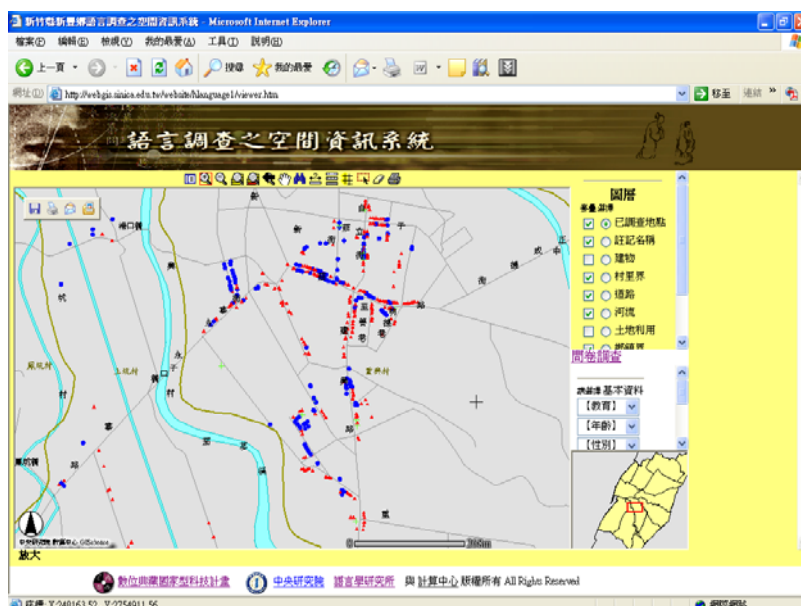


圖十五：使用 ArcView 製作圖層
（藍色表閩南語、紅色表客家語－海陸腔、綠色表其他）



圖十六：以道路交通圖呈現語言分佈

此外，將第二階段街路訂位及數位化地圖和第三階段問卷輸入所搜集整理後的數位化地圖及問卷資料電子檔之 Excel 資料，交由中央研究院計算中心製作圖層並轉檔。轉檔為 WebGIS，讓使用者能連結至 WebGIS 上的語言調查之空間資訊系統。(圖十七)



圖十七：語言分佈空間資訊系統之網路使用介面

六、匯入資料庫

最後，將中央研究院語言學研究所的『閩南語典藏－歷史語言與分佈資料庫』網站與中央研究院計算中心的WebGIS－『語言分佈空間資訊系統』加以連結，讓使用者瀏覽時，能直接連結至語言分佈空間資訊系統使用。

-
- ※ 製作單位：數位典藏國家型科技計畫---內容發展分項計畫
中央研究院語言學研究所---語言典藏計畫
 - ※ 文字撰寫：數位典藏國家型科技計畫---內容發展分項計畫
語言主題小組助理 賴佳旻
中央研究院語言學研究所
語言典藏計畫助理 盧秋蓉、黃菊芳、郭或岑
 - ※ 圖片拍攝：數位典藏國家型科技計畫---內容發展分項計畫
語言主題小組助理 賴佳旻、林淑惠
 - ※ 圖文編輯：數位典藏國家型科技計畫---內容發展分項計畫
語言主題小組助理 賴佳旻、陳秀華
 - ※ 感謝中央研究院語言學研究所---【語言典藏】之計畫主持人 鄭錦全院士與助理盧秋蓉小姐、黃菊芳小姐、郭或岑小姐撥冗指教及協助拍攝與提供資料，特別致謝。
-