董忠司教授六秩祝壽語文集出版暨論文學術討論會 逢甲大學中國文學系 2011/3/5

Excel 在語料處理上的運用

——以《客語能力認證基本詞彙-中級、中高級暨語料選粹》語料剖析為例

清雲科技大學通識教育中心講師

新竹教育大學臺灣語言與語文教育研究所博士生

駱嘉鵬

【摘要】

語言學的研究,常需處理大量語料。語料的建立,應該遵循資料庫檔案管理的規範, 才能充分發揮其效用。唯在實務處理上,卻經常會接觸到不合規範的語料。此或因資料 建立之初格式設計不當,或因材料原本並非為資料處理而建置,甚至有可能是多方湊合 的素材,此類資料自然較不利於系統化管理運用。但基於特定的需要,有時也無法迴避 此類困擾。

《客語認證語料選粹》經過專家學者選錄、編輯,其語料價值不言可喻。唯在無法 取得原始語料庫的前提之下,掌握公開的 PDF 檔,若要物盡其用,就必須經過一番資 料處理的工夫。將原本不合規範的語料轉為正規化的語料庫,必須先了解語料的特質, 評估資料處理的目標,再擬定處理的程序。按部就班,逐步完成。

本文擬分享處理本項語料的經驗,逐步說明處理程序,再歸納本項作業所使用的指 令與函數,並對其功能與用法略做說明,以提供爾後相關作業的參考,並徵求前輩同道 高見,以期有所精進。

【關鍵詞】客語能力認證、語料庫、Excel、樞紐分析

1

1. 前言

《客語認證語料選粹》公開文件屬於 PDF 檔案格式,經轉為 WORD 格式後,大體 而言,雖然整齊,卻與資料表格欄位分明的性質有所差別。本文運用 Excel 的指令與函 數,將此豐富語料轉換為標準資料表格式,並與個人所製〈中古字音表〉結合,建立中 古與現代客語對應資料表,進而探求兩音系間的音韻對應規律。全文共分7節,除前言 和結論外,分別說明匯入資料、格式化、音字媒合、古今音類對應等處理程序,並對所 使用的指令與函數約略說明其用法和功能。

2. 匯入資料

順序	作業項目	作業內容	相關指令及函數
1		使用 Solid Converter PDF Professional 軟體將 PDF	
1		檔轉為 DOC 檔	
2	赫	收 VIS 描下你走供方校校子机为「六字	全選, 儲存格格
2	平守 7 笛	府 ALS 福工作衣储行恰恰式政為「义子」	式->數值->文字
3		複製 DOC 檔內容,在 XLS 檔工作表 A1 格貼上	複製, 貼上
4		將上下冊內容合併(下冊接於上冊之下)	複製, 貼上

由於直接複製 PDF 檔案內容貼上 Excel 工作表的資料格式,比起透過 Word 轉檔 再貼上工作表,顯得更加凌亂,故先使用相關軟體將 PDF 檔轉為 DOC 檔再行處理。

原文中的分類編號採用「99-999」的格式來標示,某些分類號會被 Excel 誤認為日期,並自動轉為日期格式;為避免發生此類錯誤,故在貼上語料前,先設定儲存格格式為「文字」,如此則不致發生上述錯誤。

原書紀錄僅有 4,959 筆,因此可以合併為一份工作表,以保持資料的完整性,並可 簡化處理流程。

2	16 12	11
-	松开	11
۶.	カル	10

順序	作業項目	作業內容	相關指令及函數
5		插入標題列及「編號」、「處理」欄,以便後續處理	插入列,插入欄
6	設定欄	將 A2 格改回「通用」格式,再設定其「編號」=row()-1	儲存格格式->數 值->通用, ROW()
7	列標題	將 A2 格公式向下複製到檔尾	向下複製公式
8	及編號	在 B2 格凍結窗格	凍結窗格
9		在第1列設定「自動篩選」,以便檢視資料欄位範圍	自動篩選
10		設定資料欄位範圍內之欄位名稱為「欄1」至「欄7」	向右填满
11	删除空	將 B2 格改回「通用」格式,再設定其值為其右各欄	儲存格格式->數
11	白列	併合結果 =C2&D2&E2&F2&G2&H2&I2	值->通用,&
12		複製 AB 雨欄後,在A1 格執行「選擇性貼上值」	複製,選擇性貼
12	_		上->值
13		取消選取設定後,按 B 欄遞增排序	遞增排序

順序	作業項目	作業內容	相關指令及函數
14		刪除 B 欄空白之所有紀錄	選取, 刪除列
15		按「編號」欄還原排序	遞增排序
16		設定「編號」欄格式為「文字」,並將欄名改為「分	儲存格格式->數
16		類號」	值->文字
		將 A2 格內容設為原書「分類號」	
17	撷取「分	=IF(FIND("-",C2&" -")<4,LEFT(C2,FIND(" ",C2&"	IF(), FIND(),
	類號」	")-1),A1)	LEFI()
18		將 A2 格公式向下複製到檔尾	向下複製公式
10		治制 ∧ 烟,西劫行「避埋此卧上佐	複製, 選擇性貼
19		後表 A 陳,丹轨11 送择住船上值」	上->值
		重設 B2 格公式	
	擷取「例	(若 C2 格起始值為分類號則設為分類號,「例」或「譯」	
20	句」及	則設為「例」或「譯」,否則空白)	IF(), OK(),
	「翻譯」	=IF(A2=A1,IF(OR(LEFT(C2,2)="例",LEFT(C2,2)="譯	
		"),LEFT(C2,1),""),A2)	
21		將 B2 格公式向下複製到檔尾	向下複製公式
22		海制 R 调, 再劫行「深摆州肚上店」	複製,選擇性貼
		後衣 D 佩 丹扒们 这样任知上值」	上->值
23		篩選「欄6」,「欄7」,修正欄位錯亂資料 (53 筆)	手動修正
24		篩選「欄5」大於"一"者,刪除其內容(類名誤入資料)	自動篩選
25		篩選「處理」欄包含"-"且「欄5」為空白者,修正欄	白動結躍
23		位錯亂資料	日刧帥达
26		篩選「欄1」開頭以「例」之紀錄,複製到新資料表	自動篩選,複製,
20		Sheet1	貼上
27		將例句缺字暫補在「欄 3」(E 欄)	手動修正
		删除例句標示字眼「例 」,並修正因缺字而誤分多欄	
28		之例句	RIGHT(), &
		B2 = RIGHT(C2, LEN(C2)-2)&E2&D2	
20		泊制 B 欄, 再劫行「選擇此肚上值」	複製, 選擇性貼
2)			上->值
30		修正 B 欄標題為「例句」, 並刪除其右各欄	刪除欄
31		回原資料表篩選「欄1」開頭以「例 」之紀錄,複製	自動篩選, 複製,
51		到 Sheetl 資料表「例句」欄之右	貼上
32		在 Sheetl 資料表 C 欄前插入「處理」欄	插入欄
33		將例句分類號與翻譯分類號不一致之紀錄做上標記	ΙΕΩ ΕΧΔΟΤΩ
55		C2 = IF(EXACT(A2,D2),"",1)	п (), БЛАСТ()
34		將 C2 格公式向下複製到檔尾	向下複製公式

順序	作業項目	作業內容	相關指令及函數
35		檢視例句公輝號陶翻譯公輝號不一致者,並加以修正	IF(), 自動篩選,
55		1次他内可力积加兴翻开力积加不 又有"亚加以修正	手動校正
36		刪除翻譯標示符號「譯 」,並將修正結果存入 C2	RIGHTO I ENO
50		C2 = RIGHT(F2, LEN(F2)-2)	KIOIII (), LLIN()
37		將 C2 格公式向下複製到檔尾	向下複製公式
38		複製 C 欄, 再執行「選擇性貼上值」	複製, 選擇性貼
50	-		上->值
39		修正 C 欄標題為「翻譯」, 並刪除其右各欄	刪除欄
	撷取「客	回 〈 人 併 〉 咨 料 耒 , 左 Δ 嫻 挵 λ 「 法 水 胩 , 嫻 , 并 依	插入欄, 輸入前
40	家語」、	口(口历) 具杆衣, 在A 佩捆人 加小弧」佩·亚欣 克绝驰	兩筆編號, 向下
	「客語		填滿
	標音」、		插入欄,儲存格
41	「華語	在 D 欄插入「處理 2」欄,並設為「通用格式」	格式->數值->通
	詞義」、		用
42	「英語」	篩選「處理」欄包含「-」號之紀錄(各筆首列之分類號)	自動篩選
		在 D2 欄輸入下列公式,將「欄1」中分類號與客語用	
43		字誤合之紀錄加上標記	IF(), EXACT()
		=IF(EXACT(B2,E2),"",1)	
44		將 D2 格公式向下複製到檔尾(僅複製已篩選之紀錄)	向下複製公式
4.5		取消篩選,改依「處理2」欄(E欄)遞增排序	诉讼此方
45		(將分類號與客語用字誤合之紀錄接續在一起)	远增排户
		將「處理2」欄標1(分類號與客語用字誤合)之各筆紀	
46		錄「欄1」至「欄7」各欄複到到新工作表〈sheet2〉	複製,貼上
		處理	
47		19-1	儲存格格式->數
4/		將 sneet2 衣之欄 0,7 兩種設為一週用」格式	值->通用
		將「欄1」之分類號與客語用字拆解至「欄6」,「欄7」	
48		F2 = LEFT(A2, FIND("", A2)-1)	LEFI(), FIND(),
		G2 =TRIM(RIGHT(A2,LEN(A2)-LEN(F2)))	TRIM(), LEN()
49	1	將 F2,G2 兩格公式向下複製到檔尾	向下複製公式
		複製 Sheet2 表 B2:E3832 (欄 2-5),再貼到〈合併〉表	· 生 生 目 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -
50		G2:J3832 (欄 3-6)	稷 聚, 貼上
	1	複製 Sheet2 表 F2:G3832 (欄 6-7),再到〈合併〉表	複製, 選擇性貼
51		B2 格「選擇性貼上值」	上->值
52	1	按「流水號」欄遞增排序(還原)	遞增排序

順序	作業項目	作業內容	相關指令及函數
		將「處理」欄空白且「欄1」包含「-」之紀錄,其「處	
		理2」欄內容標註「1」	
52		篩選「處理」欄等於空白之紀錄	经避 石下插进
55		篩選「欄1」包含「-」之紀錄	即进,问下填闲
		在篩選結果首筆「處理2」欄輸入「1」,並向下	
		複製到檔尾	
		將「處理2」欄標註1之上一筆「處理2」欄標註「2」	
54		篩選「處理2」欄「等於空白」者	篩選, IF(), 向下
54		在篩選結果首筆(第2列)「處理2」格輸入	填滿
		=IF(D2=1,2,""),並向下複製到檔尾	
		將「處理2」欄標註2之紀錄末欄補上下一筆(「處理	た、肥 0-
22		2」欄標註1)首欄內容	師进, X
50			複製, 選擇性貼
56		複聚〈合併〉表,冉孰行「選擇性貼上值」 	上->值
57		刪除「處理2」欄標註1之紀錄	刪除列
58		按「處理」欄遞增排序	遞增排序
59		刪除「處理」欄內容為「例」或「譯」之紀錄	刪除列
60		按「流水號」欄遞減排序	遞減排序
		將分列上下筆且「分類號」相同之紀錄,串聯為一筆	
61		D2 = E2 & "\$" & F2 & "\$" & G2 & "\$" & H2 & "\$" & I2	&
		& "\$" & J2 & "\$" & K2 & "\$" & IF(C1="","\$" & D1,"")	
62		將 D2 格公式向下複製到檔尾	向下複製公式
	•		複製,選擇性貼
63		複聚 ' 處埋 2」 欄, 再執行 ' 選擇性貼上值」	上->值
64		按「處理」欄遞增排序	遞增排序
65		删除「處理」欄空白之紀錄	刪除列
66		刪除欄 1-7	刪除欄
		反覆將「\$\$」全部取代為「\$」,直到無可取代為止(共	
67		執行四次)	取代
			資料剖析->分隔
68		將「處理2」欄以「\$」為界, 剖析為數欄	符號->\$
69		取消自動篩選再重設 (以便查看資料欄位範圍)	自動篩選
70		重設欄名「欄 01」-「欄 14」	向右填滿
71	1	將相同屬性資料合併於同一欄	&
	1	 將「欄 01」-「欄 04」複製後,插入 Sheet1 表 B 欄之	複製, 插入複製
72			的儲存格

順序	作業項目	作業內容	相關指令及函數
73		將 B,C,D,E 欄更名為「客家語」、「客語標音」、「華 語詞義」、「英語」	直接輸入欄名
74		在 A 欄前插入一欄「編號」,並依序編定流水號	插入欄, 向下填 滿

本節處理的目標,是要將凌亂的各欄資料還原成正規化的資料表格式。主要程序如 下:

- (1) 設定欄列標題及編號:此為正規化資料表的主體架構,從 PDF 轉來的文件檔不 包含這些訊息,故在此優先添加。
- (2) 刪除空白列:原檔上下筆紀錄間,甚至不同欄位間攙雜許多空白欄位,在資料剖 析之初,即可將其刪除。
- (3) 摘取「分類號」:分類號均出現在欄位開頭,以兩個數字中間加上「-」符標示, 編號右側並有空白與其後資料區隔,故可透過公式運算取得。
- (4) 摘取「例句」及「翻譯」:例句與翻譯均獨佔一行,資料擷取相對容易,故可優 先擷取,存放於新資料表之後,再將原檔紀錄清除。
- (5) 撷取「客家語」、「客語標音」、「華語詞義」、「英語」:此為格式化資料階段相對而言最為複雜之程序。大體而言,處理方式是先將同一編號之各欄資料串連在一起,然後再將原屬同一欄位之語料分別拆解到特定欄位。處理到每一編號分別正確區隔「客家語」、「客語標音」、「華語詞義」與「英語」等四個欄位後,再與前述分類號、例句與翻譯等欄位資料併合。

4. 音字媒合

順序	作業項目	作業內容	相關指令及函數
75		清除「客家語」欄之標點符號	取代
76		清除「客家語」欄之空白	取代
77		將「客語標音」欄之標點符號改為空白	取代
78		清除「客語標音」欄多餘之空白	TRIM()
79		插入 Sheet4, Sheet5, Sheet6 三張空白工作表	
80	詞目音字 媒合	將「客家語」欄複製到工作表 Sheet4,再逐字拆解	MID(),向右填
		C2 = MID(\$B2, RIGHT(C\$1, 2), 1)	满,向下填满
01		收「安玩栖立 烟泊制列工价丰 Shoot5,西河立长钢	資料剖析->分隔
81		府 谷祜保宙」 懒凌袭到工作农 Succis, 丹逐百孙辟	符號->空格
01		將 Sheet4 與 Sheet5 音字媒合後存入 Sheet6 表	&,向右填满,
62		C2 =Sheet4!C2&Sheet5!C2	向下填满
83		彙計 Sheet6 表音字媒合結果(統計各字音使用頻率)	樞紐分析表

本節處理的目標,是將「客家語」欄的漢字與「客語音標」欄的注音逐一媒合。亦 即將每一個漢字依序與每一個音節搭配,即可歸納本認證語料所有詞目用字的注音。漢 字的區隔係以固定長度(1)為準則,音節的區隔則是根據音節間的空白。由於轉入語料的 格式並未完全一致,因此須先將「客家語」欄中的所有標點符號和空白清除;「客語音標」欄則除了將標點改為空白之外,並須清除欄位前後及音節間多餘的空白。此外,由於語料雖以北四縣腔為基礎,但亦包括南部美濃和六堆之部分特殊詞彙或音讀,因此也必須另做標記,以便區隔。

5. 古今音類對應

順序	作業項目	作業內容	相關指令及函數
01		分析各筆客語字音的聲母、韻母、調類,及聲母發音	遞增排序, 向下
04	十人留笠	部位、方式、韻母四呼、韻部、韻尾等屬性	複製, IF()
85	占 一 平 平 料 確	彙整漢字中古音類	遞增排序, &
86	封惩	將中古音類填入《認證語料》檔	IF()
87		檢選中古多音字中與客語對應規律最為吻合之音類	
88		中古與客語聲類對應關係表	樞紐分析表
89		中古五音與客語聲母發音部位對應關係表	樞紐分析表
90	十人立粘	中古與客語韻類對應關係表	樞紐分析表
91	白ブ日朔 出産目律	中古韻攝與客語韻部對應關係表	樞紐分析表
92	到應퀐件	中古韻攝與客語韻尾對應關係表	樞紐分析表
93		中古等呼與客語三呼對應關係表	樞紐分析表
94		中古與客語調類對應關係表	樞紐分析表

本節處理的重點,分為兩個階段:一是建立單筆現代客語字音與中古音類的對應關 係,以組成字音對應資料總表;二是根據此字音對應資料總表,利用樞紐分析功能歸納 出各式古今音類對應關係表。

5.1 單音對應

現代客語與中古漢語都存在一字多音的現象,就個別字音的學習而言,固然應該了 解任何客語字音在中古的各種音類,抑或了解任何中古字音在現代客語的所有音讀。例 如:對學習中古音的客語使用者而言,應了解客語讀為「zung1,zung3,tung1」音的「中」 字,中古有「通合一平東知」和「通合一去送知」兩種音讀;反之,對熟悉中古音類的 客語學習者而言,也應了解中古「通合一平東知」和「通合一去送知」的「中」字,現 代客語有「zung1,zung3,tung1」三種音讀。亦即「中」字的現代客語與中古音類的對應 關係如下:

字	客語音讀	中古音類
中	zung1	通合一平東知
中	zung3	通合一平東知
中	tung1	通合一平東知
中	zung1	通合一去送知
中	zung3	通合一去送知
中	tung1	通合一去送知

如此對應關係的羅列固然詳盡,但卻缺乏系統性。若將上述相乘的六種對應關係,

别除其中聲調不合的部分,再區分其文白屬性,則可簡化如下:

字	客語文白	客語音讀	中古音類
中	文	zung1	通合一平東知
中	白	tung1	通合一平東知
中	文	zung3	通合一去送知

如此則不止可以簡化其對應關係,在語言的學習或語音對應規律的探求方面,更富 有系統性的意義。因此,我們在歸納出認證語料所有客語字音之後,先將各筆漢字的所 有中古音類填註在其右側欄位,然後再根據初步歸納的客語與中古音類的對應規律,採 用優選理論的方法,汰除不合理的對應紀錄,由此來建立所有的現代客語與中古音類的 對應關係表。

5.2 樞紐分析

完成所有認證語料的客語字音與中古音類對應資料表之後,即可以 Excel 的樞紐分 析功能,製作各種古今音類對應規律統計表。利用樞紐分析表功能,只要依照預期的格 式,將適當欄位拖曳到「列欄位」、「欄欄位」、「資料欄位」和「分頁欄位」所示區塊內, 即可完成。此類統計表又可細分為兩類:一是逐類對應,一是依據音韻特徵分組對應。 5.2.1 逐類對應

我們可以從聲、韻、調的角度,分別製作客語與中古音類的對應關係統計表。茲以 聲類為例,利用樞紐分析表功能,製作中古聲類與客語聲母的對應關係統計表的程序如 下:

(1) 在資料表的範圍內,點選功能表列的「資料->樞紐分析表及圖報表」。

(2) 一般而言,可以接著點選「完成」紐,進入如下畫面,逕行設定各種欄位。

	將欄欄位拖曳到這裡
將列欄位拖曳到這裡	將資料欄位拖曳到這裡

將分頁欄位拖曳到這裡

- (3) 依據預期模式,分別將「列欄位」、「欄欄位」與「資料欄位」拖曳到適當區塊。 例如將「中古聲類」拖曳到「列欄位」,「客語聲母」拖曳到「欄欄位」,「字」 拖曳到「資料欄位」。
- (4)如需設定篩選條件,亦可將適當欄位拖曳到「分頁欄位」處。例如依序將「分區」 和「重複」拖曳到「分頁欄位」。

經過上述設定後,即可得到以下樞紐分析表:

分區	(全部)
重複	(全部)

計數 的字	客聲																				
中古聲類	b	c	d	f	g	h	j	k	1	m	n	ng	р	q	s	t	v	x	z	(空白)	總計
	47	24	38	7	50	9	4	35	64	45	29	10	30	8	20	15	8	5	28	177	653
Ś		1									1			1	<mark>286</mark>			281	5		575
日									1	1	2	277								45	326
匣				129	18	309		11		1	5	2	3				83			6	567
床		20												1	46			1	4		72
見					1247	21	1	26				3					1			2	1301
邪		7						1	5					15	11			22	1		62
並	32												251								283
來			2		6				717		22		1			3					751
奉	2			107									22								131
定		6	26									1			1	587				1	622
明										473	1						2				476
泥											94	77									171
知		5	135				7												193		340
初		59														2		1			62
非	64			146									5				1				216
為				1		4											44			125	174
穿		144												4							148
娘									2		23	24								9	58
神			1												127			5			133
從		97					9							89	17			10	3		225
清		116												128	2			3			249
疏		4							4					1	279			2	1		291
莊							15								2				50		67
透			24			1			3							248					276
喻					2	5						4		1	1		2	1		235	251
微										102		1					71			26	200
滂	6			25									104								135
溪		3		39	5	71		255						2			2			1	378
照			1		4		12	1							2				259		279
群					12			153				1		1							167
徹		35	1											1		4					41

疑			1	2	1	1 30	3							65	373
端	2	447		4	1	2	4			2			8		468
精	9	1	2	170 1				5				2	225	533	946
審	18							1	340			12	4	6	381
影			1	1			4				88			312	406
敷		21	1				5				1				28
澄	193	2						4		2			2		203
曉	1	143	1 175	2	3	1	0				4	4		4	347
封吊	471	3					17								491
禪	8			1					263	1		13	2		288
幽計	623 749	578 621 1	348 597	223 488	801	622 180 72	1 438	262	1397 9	864	307	360 7	785	1547	13611

上表主體部分,每一列代表一個中古聲類,每一欄代表一個客語聲母,縱橫交錯處 即為中古與客語聲類對應的筆數。例如「心」列、「S」欄交錯的數字「286」,即表示在 本客語認證語料的詞目音讀中,屬於中古「心」類且客語聲母讀「S」者共有 286 筆。 聲母「(空白)」者,表示零聲母;中古聲類空白者,則表示該類字未見於廣韻,屬於後 起字,或為筆者廣韻字音表所漏收。

此外,表格中所有統計數字,若欲進而查詢其所對應之原始語料,只需直接點按該 數字,即可完整列出;如此則不僅便於搜尋實例,在資料校對的過程中,也是一大利器。

分頁欄位中的「分區」,係指「北四縣、美濃、六堆」等三區,當選擇「(全部)」時, 所呈現之數據為三區合計筆數;若僅欲查單區對應數據,可在「(全部)」字眼處分別點 選。「重複」則是提供另一種統計模式,若選「(全部)」,不論該字音出現多少次,均分 別計算;若該欄選「1」,則將古今字音對應內容重複者合併為一筆來統計。所得結果即 如下表所示:

分區	(全部)
重複	1

計數	的字	客聲																					
中古	聲類	b		c	d	f	g	h	j	k	1	m	n	ng	р	q	s	t	v	x	Z	(空白)	總計
			22	11	19	5	17	6	4	18	33	9	14	6	17	8	15	12	6	5	13	24	264
ن،	:			1									1			1	57			50	1		111
日											1	1	2	24								15	43
匣	-					28	7	52		5		1	2	2	1				17			3	118
床				6												1	8			1	4		20
見							272	6	1	15				1					1			2	298
邪	3			4						1	1					6	8			14	1		35
並	<u>.</u>		20												59								79
來	2				2		2				212		14		1			2					233
奉	L.		2			28									9								39

定		3	12									1			1	86				1	104
明										103	1						2				106
泥											28	9									37
知		2	13				4												26		45
初		17														1		1			19
非	11			23									4				1				39
為				1		1											13			21	36
穿		26												4							30
娘									2		11	4								2	19
神			1												15			3			19
從		25					4							20	3			5	3		60
清		26												25	2			2			55
疏		2							1					1	42			1	1		48
莊							5								1				19		25
透			6			1			1							50					58
喻					1	1						3		1	1		1	1		52	61
微										13		1					11			1	26
滂	4			2									43								49
溪		1		6	4	13		69						2			1			1	97
照			1		2		7	1							1				73		85
群					8			48				1		1							58
徹		13	1											1		2					17
疑						1		1	1		1	60								11	75
端			90				2		1		2	2				1			2		100
精		2	1		1		40	1						3					47	2	97
審		9												1	46			5	1	3	65
影					1			1				2					21			83	108
敷				10		1							4				1				16
澄		48	2											3		1			2		56
曉	1			14	1	42		2	1			1					2	3		3	70
封吊	93			1									10								104
禪		5					1								41	1		8	2		58
總計	153	201	148	118	316	124	68	162	254	127	76	117	148	78	241	156	77	99	195	224	3082

5.2.2 分組對應

上列逐類對應雖可羅列各類對應的詳情,唯對於音韻規律的探究或相關語言的教學 而言,則難免失之瑣碎而缺乏系統性。在實際運用時,分類對應所呈現的數據,往往更 富有教學、研究的價值。欲進行分類對應的統計,自然應先對各筆紀錄的各類屬性進行 分析;分析完成之後,即可仿照逐類對應的統計程序,製作各種分類對應統計表。茲以

中古五音與客語發音部位對應關係表為例,統計如下:

分區	1
重複	1

計數 的字	客部位							
五音	舌尖	舌尖前	舌面	舌根	唇齒	喉	雙唇	總計
牙音		1	2	434	7	33		477
正齒二	-	2 91	8					101
正齒三	-	3 206	6	3		3		221
舌上	28	8 82	5	4		2		121
舌齒	200	6		23		12	2	243
舌頭	259	9 5		10		1		275
重脣	-	1			5		309	315
喉音	-	3 1	5	25	94	241	2	371
輕脣				1	67	2	41	111
齒頭	4	2 171	154	3		2		332
(空白)	70	0 32	16	40	11	27	45	241
總計	574	4 589	196	543	184	323	399	2808

上表分頁欄位中,「分區」格右側選「1」,表示只統計「北四縣」語料;「重複」格 右側選「1」,表示中古同音且客語同音之字併為一筆計算。由此可知:中古牙音所對應 的現代客語主要是舌根音,其次是喉音。若點按次要對應的數字 33,則可另開視窗呈現 該 33 筆紀錄,進而可以發現此類喉音多為「溪」類今讀「h」聲母者。

6. Excel 相關指令與函數

Excel 提供相當多樣的指令與函數,且可自行設定巨集程式,執行批次運算。唯巨 集通常比較適合已經過正規化處理的資料表,用來反覆執行相同模式的處理運算;對於 類似本文所處理的繁雜語料而言,必須隨時因勢制宜,甚至需要人工逐筆校正,則設計 巨集程式來處理,反而事倍功半。因此本文採用 Excel 基本指令與函數,相對而言,其 邏輯亦較巨集程式單純。本節歸納前述處理流程中使用的所有指令與函數,統計其使用 次數,並略述其使用方式及其功能。

6.1 指令

編號	相關指令	使用 次數	語法或操作方式	功能
1	&	7	х & у	將「&」左右參數值串接成新字串
2	向下複製	0	在儲存格右下侧「填滿控點」	將儲存格公式依照相對位址往下
2	公式	9	點兩下	複製到資料末端

編號	相關指令	使用 次數	語法或操作方式	功能
3	向下填满	4	連續選取上下多個儲存格後, 再按 Ctrl+D 控制鍵	將選定範圍中最上端的儲存格內 容和格式複製到以下所有選取儲 存格中
4	向右填满	4	連續選取左右多個儲存格後, 再按 Ctrl+R 控制鍵	將選定範圍中最左側的儲存格內 容和格式複製到右方所有選取儲 存格中
5	自動篩選	8	在資料區中的標題列時,點選 資料->篩選->自動篩選	呈現篩選紐以便篩選,或隱藏篩 選紐
6	刪除列	4	選取刪除列範圍後,在選取區 間內按右鍵->刪除	删除選定範圍,並將以下紀錄往 上移
7	刪除欄	3	選取刪除欄範圍後,在選取區 間內按右鍵->刪除	刪除選定範圍,並將右側紀錄往 左移
8	取代	3	編輯->取代,設定尋找目標和 取代值,按全部取代紐	替换字串
9	凍結窗格	1	選定凍結起始位置,視窗->凍 結窗格	保留工作表上凍結起始位置上方 和左側訊息
10	插入列	1	選取一列或數列,插入->列	在游標所在列插入選定列數,並 將原列紀錄往下移
11	插入複製 的儲存格	1	複製,選取一格或連續數格, 按右鍵->插入複製的儲存格	貼上複製內容並將現有儲存格往 右移或往下移
12	插入欄	5	選取一欄或數欄,插入->欄	在游標所在欄插入選定欄數,並 將原欄紀錄往右移
13	貼上	6	编輯->貼上,或按右鍵->貼上	將剪貼簿內容貼在指標及其右下 同等範圍內,並覆蓋該範圍內舊 資料
14	資料剖析 ->分隔符 號->\$	1	將選定範圍內儲存格內容依 「\$」號拆解,並覆蓋其右側內 容	將合併在一欄之資料內容拆解至 不同欄位存放
15	資料剖析 ->分隔符 號->空格	1	將選定範圍內儲存格內容依照 空格拆解,並覆蓋其右側內容	將合併在一欄之資料內容拆解至 不同欄位存放
16	遞減排序	1	點按遞減排序工具紐	依照游标所在栏值将资料表遞減 排列
17	遞增排序	7	點按遞增排序工具紐	依照游标所在栏值将资料表远增 排列

編號	相關指令	使用 次數	語法或操作方式	功能
18	樞紐分析 表	9	資料->樞紐分析表及圖報表	依照選定欄位,交叉統計各項對 應數值
19	複製	15	選取->右鍵->複製	將所選取內容複製到剪貼簿
20	盆瑕	2	按篩選紐->選定篩選方式->設	篩選符合特定條件的紀錄,並隱
20	雪皮	5	定篩選值	藏其他紀錄
	選擇性肚			將儲存格現值存入,並清除公
21	运行任知	8	編輯->選擇性貼上->值	式,可避免因相對位址異動而改
	上-7 但			變儲存格值
	儲存格格			設定儲存格為文字格式,可防止
22	式->數值	2	儲存格格式->數值->文字	系統將看似日期的字串資料轉為
	->文字			日期
	儲存格格			机宁健方故为涌田故半,可滥久
23	式->數值	4	儲存格格式->數值->通用	成大帕什哈匈迪用哈式, 9 避免 么结收八十祖五十宁
	->通用			亦 航 枏 公 式 税 局 义 子

6.2 函數

編號	相關函數	使用 次數	語法或操作方式	功能
24	EXACT()	2	EXACT(字串 1,字串 2)	檢查兩個字串是否完全相同
25		2	FIND(待查字串,查詢對象,開始	計算待查字串在查詢對象中出現
23	FIND()	2	查詢位置)	的位置
26	IFO	7	IF(條件式,條件成立時傳回值,	根据低件识完甘估
20	IF()	/	條件不成立時傳回值)	松稼味什設足共值
27	LEFT()	3	LEFT(字串,長度)	從字串左側擷取指定長度內容
28	LEN()	2	LEN(字串)	計算字串長度
20	ΜΙΦΟ	1	MID(宁史 扣払位罢 長庄)	從字串左邊指定位置截取指定長
29	MID()	1	IMID(于中,起始位重,衣皮)	度內容
30	OR()	1	OR(P,Q)	傳回「P 或 Q」的邏輯值
31	RIGHT()	2	RIGHT(字串,長度)	從字串右側擷取指定長度內容
32	ROW()	1	ROW()	傳回所在列的列號
33	TRIM()	2	TRIM(字串)	删除字串中多餘的空格

7. 結論

語料庫語言學是二十世紀新興的語言科學,大量而精確的語料有助於語言理論的探 求與驗證。唯在具體運用時,則須妥適規劃語料的儲存格式,方能發揮較大的參考價值。 在處理語料時,難免會遇到未經正規處理的素材,如此則必須耗費一番轉檔處理的工 夫。本文以客語認證語料為例,說明利用 Excel 指令與函數整理語料的經驗;除將原始 文書檔案形式的語料轉為資料表之外,並與中古音類資料表結合,最後再利用樞紐分析 表的功能,呈現古今音類的對應規律。對於中古漢語或客語的教學與研究,應可發揮執 簡馭繁的功效。

【參考文獻】

行政院客家委員會. 2009. 《客語能力認證基本詞彙-中級、中高級暨語料選粹-四縣》。 http://elearning.hakka.gov.tw/Kaga/Kaga Attach QDM.aspx?xItem=47689。

兔兔電腦教室. Excel 試算表教學網站。<u>http://www.bonny.idv.tw/www/exexcel/index.htm</u>。 駱嘉鵬. 2010. 〈《臺灣閩南語常用詞辭典》音字系統一致性的校正——資料庫管理程式 在辭典校對作業上的用例〉,《臺灣語文研究》, 5.2:43-62。