

《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*

駱嘉鵬
清雲科技大學

辭典常做為釋疑辨異的主要依據。辭典的字形、字音和釋義，理當遵照嚴格的系統化標準。就臺灣閩南語而言，腔調的複雜和用字的混亂，是民間長期存在的客觀事實。閩南語辭典字形和字音的釐定工作，因而更形困難。

教育部主編的《臺灣閩南語常用詞辭典》，歷經編輯群的撰稿、審核，在2005年大體定稿。在審稿過程中，偶爾仍會發現字形、字音前後不一的現象，個人乃提議進行本辭典音字系統一致性的校正工作。

個人運用資料庫管理程式，先對辭典內容進行初步剖析，校正部分體例；接著進行辭典內部閩南語用字和字音的媒合，並整理出本辭典閩南語字音的音韻系統；完成不合理音節的校正作業之後，再歸併各個用字的所有音讀，交由編輯小組進行人工檢校。透過電腦軟體的輔助，提高了本辭典的校對速度和精確性。

本文擬分享資料庫管理程式在辭典校對工作上的使用經驗，以做為爾後相關作業的參考。

關鍵詞：辭典學、閩南語注音、閩南語用字、一致性、資料庫管理

1. 前言

方言辭典的編纂，可以分為兩大類型：一是方言標準辭典，一是方言比較辭典。前者以教學為預設目標，在各個次方言間，選取一個代表點，抑或取得一個折衷的標準，

* 本文於「第七屆臺灣語言及其教學國際學術研討會」宣讀時，得到特約討論人曾榮汾教授及湯廷池教授的指正，會後得到本刊匿名審稿委員提供修正意見，獲益良多，謹在此致謝。

而不羅列所有次方言的差異；後者以研究為主要目的，重在方言差異的實錄，以做為比較研究的依據。兩者的編輯旨趣既有區別，對於資料的取舍和編排的方式，自然也就有所不同。

教育部國語會主編的《臺灣閩南語常用詞辭典》，為了照顧大多數的台灣閩南語使用者，因此在音讀方面酌收第二優勢音；在用字方面，也附註異體字。但基本上，本辭典仍屬於方言標準辭典，故其注音和用字系統的一致性，也就顯得格外重要。

然而，就實際執行效果而言，由於編輯成員母語腔調的差別、閩南語用字標準初定、再加上難以避免的人工手誤等因素，在本辭典審校結束之後，仍然存在部分音字系統不一致的問題。為提升本辭典音字系統一致性，並降低錯誤率，乃由本人負責整理本辭典相關語料，歸納全典閩南語用字和注音資訊，以做為進一步檢視、修正的參考依據。

本文擬分享相關拙見和經驗，並藉此拋磚引玉，切磋琢磨，以做為爾後類似作業的參考。

至於執行本校正作業的電腦軟硬體設備，由於筆者接手本案時，本辭典已可上線使用，因此並無特殊需求。由於編輯小組原本校對作業，皆在 Excel 工作表中進行，在執行音字系統一致性校對時，為配合使用習慣，主要是以 Excel 基本功能輔以巨集程式處理。至於較為複雜的處理程序，則以資料庫管理程式 Visual FoxPro 處理之後再提交於 Excel 中使用。

2. 音字系統一致性的意義

方言辭典的音字系統一致性，包含兩個層面：一是注音系統，一是用字系統。所謂注音系統的一致性，是指在一部辭典中，對於任一詞素的注音，不因腔調的不同而分注不同的音讀。例如「雞」字在台灣閩南語主要有「ke、kue」兩種次方言讀音，辭典既定「ke」音為主音讀，則在全文中的所有「雞」字，除了附註又音之外，理當全部注為「ke」音，而不可雜用「ke、kue」，漫無準則。用字系統的一致性，則是指任一詞素的字形，不因習慣、風格……等差異而有所不同。例如對應華語「人」義的「lâng」，通行寫法有「人、儂、郎……」等字形，一部辭典除了附註異體字形外，原則上也只能擇一使用。這種一致性的原則，應該貫徹在整部辭典的所有內容，如詞目、用例和附錄之中。

對於方言標準辭典而言，音字系統的一致性，攸關整體的成敗。一部注音和用字系統紊亂的辭典，將會讓使用者無所適從。果真如此，則其實用價值勢必大打折扣。

3. 音字系統一致性的檢校需求

本辭典的編輯過程，是台灣閩南語學界的空前大合作。由於編輯和審稿委員囊括學界菁英，南腔北調、各放異彩。本辭典編輯體例雖約定音讀以通行腔為標準，但因各委員的自然習用，即使多加留心，百密一疏，勢必難以完全避免。

再者，由於閩南語用字向來混亂，爭議用字言人人殊；人工校稿，難以兼顧全局；加上眾人合作，更難一以貫之。教育部「推薦用字」公布於本辭典大致底定之後，¹ 緩不濟急。因此，本辭典的用字系統性實有必要加以全面清理。

此外，由於資料內容的龐雜，人工手誤，自然也在情理之中。凡此種種，都需要對整部辭典的音字系統一致性做一次全面性的整理。

在辭典的編輯過程中，注音和用字系統不一致的形成原因錯綜複雜。理論上雖可分別探究，但在實際檢校時，僅需掌握任一詞素均使用固定字形、固定字音的原則，即可達到音字系統一致的目標。因此，校正辭典注音系統和用字系統一致性的方法，即可綜合探究，而不必分別處理。

4. 問題剖析

在筆者接觸本案時，系統已開始進行線上測試，唯其檔案與程式架構，並未依照正規化處理原則設計。根據個人對本辭典的理解，要提高本辭典的音字系統一致性，必須先將本辭典所有閩南語用字和注音資訊加以媒合，歸納出所有用字的所有注音之後，再以人工檢視問題字音，並視實況予以更正。

本辭典正文中的閩南語用字和注音，出現在詞目、詞目注音和釋義欄中的例句² 之中。關於詞目和詞目注音欄，因為內容單純，處理方式相對簡單許多。至於釋義欄中的例句和注音，則在進行音字媒合之前，還需要執行一些前置處理程序，如此方能精確掌握辭典用字的注音，以做為後續處理的依據。

¹ 教育部，《臺灣閩南語推薦用字》（第一批）公告於 2007 年 5 月 23 日，（第二批）公告於 2008 年 5 月 1 日，都在本辭典複審完校之後。

² 本文所指「例句」，實際上也包括不成句的例詞或詞組。

4.1 釋義體例

本辭典釋義欄的編輯體例如下：³

詞性 華語釋義。例：台語規句的例句。Kui-kù lē-kù ê im-phiau。（整句華語解釋。）；台語詞 im-phiau（華語對譯）、台語詞 im-phiau（華語對譯）。**近**
近義詞**反**反義詞

若一條詞目具有不同義項，則在**詞性**之前以阿拉伯數字標示項次。例如詞目「猛」的釋義⁴如下：

1. **形**行動迅速。例：你跤手有夠猛，隨予你提去。Lí kha-tshiú ū-kàu mé, suí hō lí thèh--khì。（你動作真快，馬上被你拿去。）**近**緊
2. **形**旺盛而強烈的樣子。例：日頭足猛足熱。Jit-thâu tsiok mé tsiok juáh。（太陽很豔很熱。）；火真猛 hué tsin mé（火勢很猛）。**近**旺

4.2 資料剖析

根據上述體例，要將釋義欄中的閩南語用字和注音精確地搭配起來，必須進行以下幾項前置處理程序：

4.2.1 標記統一

在釋義欄中，我們需要的資訊是例句的閩南語用字和注音。依照編輯體例，這些資訊是以「例：」領頭，而與前面的「詞性」和「華語釋義」相區隔；注音之後則在「（）」中附註該例句的華語解釋。該義項若有近、反義詞，則分別在**近**、**反**標誌之後註記。

「例：」標記既為截取資訊的重要依據，故其標記的一致性，顯得格外重要。在檢校之前，我們先篩選包含「例」字的紀錄，發現有數筆標誌並未統一；其中包含「例」字之後多一「如」字的，「例」字之後未加冒號的，以及冒號內碼不同等等類型，在截取例句和注音之前，就必須先把這些格式統一，以免遺漏相關例句的字音資訊。

4.2.2 釋義彙總


本辭典的釋義內容分列四欄。釋義時，依各義項先後順序分別填入各欄；義項不足四個的，填完為止；超過四項的，則會將兩個（甚至三個）義項填在同一欄。在執行音字媒合作業之前，為簡化處理程序，應先過濾空白欄位，以及不含例句的釋義欄位，而僅篩選出包含例句的欄位內容做為後續處理的材料。

³ 此為編輯委員會所提供的說明。其所使用之符號、音標、文字及其排列順序，直接代表其體例。意即先標註詞性，標註方式係在詞性文字外加方框；其次以華語文句解釋詞義；其次以「例：」引領例句，先以台語文句舉例，再以台羅音標注音。台語例句最多僅舉一例，不成句之例詞或詞組則不限一例。

4.2.3 欄位拆解

篩選出包含例句的欄位之後，接著必須對其內容加以剖析。首先應將「例：」標記及其左邊的「項次編號」、「詞性」和「華語釋義」等資料清除；若其末端包含「近義詞」或「反義詞」資訊，也應一併清除。接著再根據編輯體例，將每個義項之中所包含的各個例句（含「閩南語用字」、「閩南語注音」和「華語解釋」）分別拆開。例如：

詞目：不孝

釋義：對長輩不孝順。例：伊真不孝。I tsin put-hàu。（他很不孝順。）；不孝囡 put-hàu kiánn（不孝子）。

應將例句拆解為下列兩筆：

伊真不孝。I tsin put-hàu。（他很不孝順。）

不孝囡 put-hàu kiánn（不孝子）。

各個例句分別拆開之後，再依照編輯體例，分別將「閩南語用字」、「閩南語注音」和「華語解釋」拆解，並存放到新的資料表中，以便後續分析處理。如同上例，應將兩筆例句分項拆解如下：

表一 例句之閩南語用字、注音及華語解釋欄位拆解表

閩南語用字	閩南語注音	華語解釋
伊真不孝	I tsin put-hàu	他很不孝順
不孝囡	put-hàu kiánn	不孝子

4.2.4 例句音字媒合

將各例句中的「閩南語用字」和「閩南語注音」拆解開來之後，原則上依序將每個漢字對應到每個音節即可完成音字媒合作業。但因本資料庫的屬性特殊，仍有幾點必須特別處理。否則，若例句字數與音節數不同，勢必導致媒合錯誤。

4.2.4.1 字數統計

由於本辭典尚有部分用字須待造字，這些待造之字在本辭典資料檔中是以大括弧「{}」包上四個字碼來標示。例如：tshit-thô 的 thô，現以{8BB5}標示。為避免統計錯誤，在計算字數前，先將每個待造之字的字碼連同括弧分別以一個漢字替代。例如本辭典未使用「籊」字，因此可先以「籊」字取代本辭典資料庫中的「{8BB5}」，等到音字

⁴ 參見教育部國語推行委員會（2008b），後列辭典引文仿此。

媒合完成之後，再將「籐」字還原成「{8BB5}」。⁵

此外，例句漢字間的標點符號，以及手誤多出來的空白，也應一併加以濾除，才能算出精確的字數。例如：「你著細膩，毋通公親變事主。」本句中的閩南語字數應為 11，而不可將「，。」兩個標點符號誤計在內，算成 13 個字。

4.2.4.2 音節數統計

例句注音中的音節基本上並不連寫，可以做為區分音節、計算音節數的依據，但音節間的區隔符號，則又分為三類：即空白、連字符和標點。例如：「Lí tiòh sè-jī, m̄-thang kong-tshin piàn sū-tsú。」為方便統計，統一將這三類符號改為一個半形空白，而後再統計各個例句的音節數。

4.2.4.3 音字數量比對

計算出各例句的漢字個數和音節數之後，接著比對其值是否相同。若發現字數與音節數不同，則必須以人工檢視其原因，並視實際狀況予以更正。根據實際作業的經驗，這些數量不合的原因，包括缺漢字、缺注音以及兩音節誤合為一等狀況，在修正這些錯誤之後，才能進行準確的音字媒合作業。

4.2.4.4 例句音字媒合

在確認各例句的漢字個數與注音音節數完全吻合之後，即可依序將每個漢字與注音搭配起來，建立本辭典例句字音資料表，請見表二：

⁵ 關於此處所談的字碼替換方法，筆者另行設計一個整批替換的轉換工具；使用者只要將要替換的原碼（文數字、漢字詞或編碼）和新碼對照表依照約定的格式存放，再按一下執行鍵，即可將所有待轉字碼全部替換。

表二 例句音字媒合表

閩南語用字	閩南語注音
你	Lí
著	tióh
細	sè
膩	jī
毋	m̄
通	thang
公	kong
親	tshin
變	piàn
事	sū
主	tsú

4.2.5 詞目音字媒合

仿照 4.2.4 節所述方法，執行詞目及詞目注音欄的音字媒合作業。本項作業與前項差別在於部分詞目注音包括第二優勢音，但依體例，此類注音僅供參考，不宜用來標注例句字音，因此必須在字音檔中獨立一欄註明字音來源屬性，以做為後續判別例句是否誤注的依據。例如：

詞目：天然

注音：thian-jiân / thian-liân

應將其分析為：

表三 詞目音字媒合表

字	音讀來源	音讀
天	詞目主音	thian
然	詞目主音	jiân
天	詞目二優	thian
然	詞目二優	liân

然後再將「詞目二優」與「詞目主音」相同者（如本例的「天 詞目二優 thian」）刪除。

此外，因詞目及詞目注音欄內容皆為單一屬性資料，因此不必執行 4.2.1 至 4.2.3 各節所述的分析、拆解作業，相較之下，處理程序較為單純。

5. 音字媒合結果分析與音讀核校

經過上述音字媒合作業處理之後，即可針對媒合結果進一步分析其音韻結構，製作下列各式字音表。各字音表的製作，在每次校對前執行，以便篩選出問題字音，提供人工校正參考；校正完成後再重新製作，反覆執行，直到完校為止。以下 5.1 至 5.8 節，以例句字音為例，分別說明各字音表格式，詞目字音表亦仿此格式製作。5.9 節則將詞目字音與例句字音表合併，製作「例句與詞目字音比對表」，以便篩檢例句異常字音。

5.1 音字媒合表

依序媒合各例句漢字與注音的初步結果，請見表四：

表四 例句音字媒合表 (2)

句次	詞目序號	例 句 注 音
1	1	一 tsit 蕊 lúí 花 hue
2	1	紅 Âng 嬰 enn 仔 á 哭 khàu 甲 kah 一 tsit 身 sin 軀 khu 汗 kuānn
3	1	一 tsit 暈 khùn 仔 á
4	1	一 tsit 絲 si 仔 á
5	2	一 it 流 liû
6	2	{8CAE}In 兩 nāng 个 ê 生 senn 做 tsò 一 it 模 bôo 一 it 樣 iūnn
7	2	一 it 生 sing
8	2	一 it 向 hiòng

5.2 字音彙總表

先將上述「音字媒合表」逐字拆解，再將相同字音彙整，統計其使用頻次，並分析其音節結構。完成之後，再據以製作以下各表，請見表五：

表五 字音彙總表

字音 ⁶	字	音	聲母	韻母	聲調	韻調 ⁷	韻末 ⁸	次數
一 it	一	it		it	1	it	t	150
一 tsit	一	tsit	ts	it	8	it	t	1002
乙 it	乙	it		it	4	it	t	3
丁 ting	丁	ting	t	ing	1	ing	ng	9
七 tshit	七	tshit	tsh	it	4	it	t	31
九 káu	九	káu	k	au	2	áu	u	15
九 kiú	九	kiú	k	iu	2	iú	u	6

5.3 聲韻調配合表

根據「字音彙總表」製作的音節統計表，見表六：

⁶ 音、字合在一欄的目的，在於後續篩檢到問題字音時方便複製，再到「音字媒合表」尋找出處，再根據「句次」編號訊息修改原始資料內容；分別為兩欄的目的則在於製作後續各種統計表。

⁷ 本欄係將字音去除聲母而成，目的為便於製作下述「聲韻調配合表」。

⁸ 本欄係截取音節末尾音段而成，目的為便於製作下述「韻末配調表」，以發掘入聲韻調的不合理組合紀錄。

表六 聲韻調配合表

韻調	b	g	h	j	k	kh	l	m	n	ng	p	ph	s	t	th	ts	tsh	零聲母	總計
a			5		199	281	127	6			14	33	12	47		164	25	1225	2138
á					21	11		178	169	4	39			6		102	33	1743	2306
à			3		29	10		4	4		12		3	11	2	7		3	88
â	10	3	3				4	18	27		2			1			50		118
ā	7		41		576		11	142	245		2	5						1	1030
ah	93		2		556	404	9	10	5		53	189		44	6	24	20	47	1462
áh			19			2	9						3	29	13	13			88
ai			7		26	63	337		2		4		43	11	37	153	11	14	708
ái	85		56		78	1		1			111		165	5	3	10	40	1	556
ài					57	5	1	280			66	22	15	23	22	61	127	663	1342
âi	8		5				551				52		3	75	27	79	10		810
âi		13	62				128	15	12		21		16	536	18	136	6		963

本表為 Excel 樞紐分析表格式。依照樞紐分析表內定功能，在任一統計數字儲存格中雙按滑鼠左鍵，即可以上述「字音彙總表」格式另開新表，以顯示該數字所代表的原始資訊。例如「m」行「á」列交錯的數字 178，表示「má」音共有 178 筆，雙按該數字即可另開新表，見表七：

表七 同音紀錄分析表

字音	字	音	聲母	韻母	聲調	韻調	韻末	次數
碼 má	碼	má	m	a	2	á	a	2
媽 má	媽	má	m	a	2	á	a	52
馬 má	馬	má	m	a	2	á	a	124

如此一來，則當在分析表中發現錯誤或異常數據時，即可根據「字音」欄查詢字音彙總表，進而查閱原始資訊，進行修正。以下四種配合表皆仿此方式製作。

5.4 聲韻配合表

忽略聲調的音節統計表，便於查閱異常聲韻組合的音節，例如：鼻音聲母與鼻化韻或閉尾韻相配等問題；請見表八：

表八 聲韻配合表

韻母	b	g	h	j	k	kh	l	m	n	ng	p	ph	s	t	th	ts	tsh	零聲母	總計
a	17	3	52		825	302	142	348	445	4	69	38	15	65	2	273	108	2972	5680
ah	93		21		556	406	18	10	5		53	189	3	73	19	37	20	47	1550
ai	93	13	130		161	69	1017	296	14		254	22	242	650	107	439	194	678	4379
aih			1																1
ainn			11		4	2						205		3		9		1	235
ainnh			1																1
ak	163	11	93		82	37	60				110	27	9	180	52	10	9	20	863
am		7	44		264	37	76						49	69	35	27	41	127	776
an	131	22	105		172	44	240				118	7	89	131	93	44	62	297	1555
ang	58	1	106		297	164	1413				136	58	48	291	376	52	21	112	3133
ann			7		79	5						9	315	82	7	1	1	7	513
annh			3										6						9

5.5 韻末配調表

各音節末尾音段與聲調配合關係表，藉以發掘誤植聲調或韻母，以及入聲韻調的異常組合。⁹ 請見表九：

⁹ 本表非第一次製作之原始統計表，初次製作此表之後，已將入聲韻調異常結合之紀錄校正，故在此呈現者已無該類錯誤搭配類型。

5.7 字調配合表

各字聲調統計表，藉以發掘誤打聲調或漢字，例如「乜」字第 8 聲及「二」字第 2 聲等問題，請見表十一：

表十一 字調配合表

字	1	2	3	4	5	7	8	9	總計
一				150			1002		1152
乙				3					3
丁	9								9
七				31					31
乜		140					1		141
九		21							21
了	2	254							256
二		1				46			47
人	55				1421				1476

5.8 同音字表¹⁰

本表係仿照「聲韻調配合表」排列方式，但在表格內不呈現統計頻次，而將各音節對應漢字填入其中。內容分為六種：詞目主音讀、詞目第二優勢音、單字（不成詞）主音讀、單字（不成詞）第二優勢音、例句字音及所有字音彙總表，藉以直接查看問題字音。以「例句同音字表」為例，請見表十二：

¹⁰ 具體製作方法請參見駱嘉鵬（2006）。

表十二 例句同音字表

韻調	p	ph	b	m	t	th	l	n	k	kh	g	ng	h	ts	tsh	s	j	零
a	吧巴 疤 {89E 4}	拋葩 脬			大焦 礁礁		拉		交家 加佳 嘉咖 膠咳 傀絞 鉸鷄	咳跤 尻			哈	查昨 贖	差	沙柵 捎		阿鴉 腌
á	把飽			馬碼 媽	打			那若 哪攪	假絞	巧		雅		早	吵炒			仔
à	霸豹			嘛	罩	拈			教較 駕激	敲			孝	炸		嘎		
â	爸		貓麻 峇媯 猫	麻			垃喇 鯪	林籃 籠			牙		結			查柴		
ā	罷	泡	覓峇 攞	嘛罵			拈	但若 焦吶	共咬				下夏					
ah	百爸	拍	肉	嘛媽	答貼 搭	塔	啦拉	凹隘	較甲 蓋恰 胛胛 枹	較卡 盞			哈	扎紮	插			啊壓 矣鴨 押
ah					踏	疊	曆獵 蠟臘			卡闔			合箬	闌		燂		盒曷

5.9 例句與詞目字音比對表

本表係由「例句字音彙總表」與「詞目字音彙總表」組合而成，組合之後，再比對例句字音與詞目字音重出狀況，請見表十三：

表十三 例句與詞目字音比對表

分類	字音	字	音	區分	次 數
主音	一 it	一	it		
例句	一 it	一	it	1	150
主音	一 tsit	一	tsit		
例句	一 tsit	一	tsit	1	1002
主音	乙 it	乙	it		
例句	乙 it	乙	it	1	3
主音	丁 ting	丁	ting		
例句	丁 ting	丁	ting	1	9
主音	七 tshit	七	tshit		
例句	七 tshit	七	tshit	1	31
主音	九 káu	九	káu		
例句	九 káu	九	káu	1	15
主音	九 kiú	九	kiú		
例句	九 kiú	九	kiú	1	6
主音	了 liáu	了	liáu		
例句	了 liáu	了	liáu	1	254
例句	二 jí	二	jí	3	1
主音	二 jī	二	jī		
例句	二 jī	二	jī	1	46
二優	二 lī	二	lī		

「分類」欄標示音讀來源，分別為詞目主音、詞目二優及例句字音。「區分」欄僅針對例句字音做分析，其中 1 代表該例字音已見於詞目主音讀，2 代表已見於詞目第二優勢音，3 代表不見於詞目音讀。凡是「區分」欄標 2 者，理應改為主音讀；「區分」欄標 3 者，則應逐筆檢視其注音與用字，再視實況修正，抑或考慮增補詞目音讀。

6. 檢校程序

根據上文的說明，以下依序臚列所有執程序，並略述各程序的作業目的及其處理筆數，請見表十四。

表十四 檢校程序表

順序	作業項目	作業內容	目的	處理筆數
1	前置作業	總檔新增「編號」欄，依序填入紀錄編號	做為後續處理的索引	17,243
2	標記統一	將所有「例:」、「例如:」、「例;」替換為「例:」	統一例句識別標記，做為後續截取例句及音讀之依據	14
3	標記統一	將詞目、釋義、異體及近反義詞欄中所有「v」替換為「-v」	統一造字識別標記，將「v」前漏標「-」號者補足	98
4	標記統一	將詞目、釋義、異體及近反義詞欄中所有「--v」替換為「-v」	統一造字識別標記，將「-v」前多標「-」號者刪除	27
5	釋義彙總	將各釋義欄中非空白紀錄及編號填入暫存工作表	以便後續截取例句及音讀	19,635
6	例句篩選	將暫存表包含「例:」之紀錄轉存於「例句」工作表	以便後續截取例句及音讀	12,885
7	例句篩選	將暫存表不含「例:」之紀錄轉存於「非例句」工作表	以備不時之需	6,750
8	欄位拆解	使用「例句」工作表，將釋義欄中「例:」標記之右方文字存入「例句」欄	清除與例句音、字和華語翻譯無關之訊息	12,885
9	欄位拆解	將「例句」欄中「()」號之間的文字拆置於「華語」欄	分離出華語釋義(但一欄內存多例者仍待分析)	10,541
10	欄位拆解	校正例句紀錄，將每條例句分列一筆紀錄	校正前項之疏漏	14,885

11	欄位拆解	將閩南語例句、注音及華語釋義分欄存放	以便後續比對閩南語用字及注音	14,885
12	欄位拆解	清除「例：」之右多餘的空白	統一體例，以便音字媒合	9
13	欄位拆解	清除閩南語例句之末多餘的全形空白	統一體例，以便音字媒合	2
14	欄位拆解	清除閩南語字與音之間多餘的空白	統一體例，以便音字媒合	158
15	漢字拆解	將例句欄中各漢字代碼分別以辭典未使用之漢字代替	以便後續之音字媒合時，精確計算字數	1,672
16	漢字拆解	清除例句用字欄中之標點	以便後續之音字媒合作業	10,728
17	漢字拆解	離析所有例句之所有漢字	以便後續之音字媒合作業	95,802
18	注音拆解	將例句注音中之半形逗號改為全形逗號	統一體例，以便音字媒合	2
19	注音拆解	依空白與連字符將例句注音分音節，並計算各句音節數	以便後續之音字媒合作業	95,787
20	音字媒合	比對各例句漢字個數與注音音節數之異同	篩檢出字數與音節數不同之紀錄，以便檢視比較其差異，減少音字媒合之錯誤率	14,885
21	音字媒合	逐筆檢視例句漢字個數與注音音節數有別之紀錄，並視實況校正	校正例句中缺字、缺音及音節誤黏之紀錄	17
22	音字媒合	音字媒合	彙計各例句用字字種、讀音，及其使用次數	95,802
23	音字媒合	補足整句注音末端漏標之句號	統一體例	24
24	音字媒合	將例句注音首碼小寫者改為大寫	統一體例	33
25	音字媒合	詞目及注音欄音字媒合	做為比較例句及詞目注音異同之依據	17,243
26	音讀核校	例句音讀分析	做為製作各式例句字音表之資料來源	95,802

27	音讀核校	詞目音讀分析	做為製作各式詞目字音表之資料來源	17,243
28	音讀核校	不合法音節檢校	篩選不合法音節，並據以校正主檔及例句注音檔，重計用字注音使用次數	12
29	音讀核校	製作各式例句字音分析表	做為核校例句注音之參考依據	95,802
30	音讀核校	製作各式詞目字音分析表	做為核校詞目注音之參考依據	17,243
31	音讀核校	例句中複音詞二優音讀篩檢	依據詞目音字彙總檔主音、二優音讀分析，篩選出例句中複音詞採用二優之紀錄	11
32	音讀核校	例句字音歸類	依據詞目音字彙總檔主音、二優音讀分析結果，區分所有例句字音的類別，以便篩選錯誤注音	95,802
33	音讀核校	例句中之非主音讀核校	檢視例句注音未採用主音讀之紀錄，並依實況予以校正	421

7. 結語

編輯電子辭典，理當依照資料庫系統設計原則，妥善規劃檔案架構，分析編輯與使用流程，進而提供適切的後端管理與前端操作介面，如此才能方便資料建立者與使用者發揮其效能，並減少錯誤發生的機率。唯在實際作業上，難免會有始料未及之處。例如：文本來源並非為資料庫設計而建立，資料庫軟體的更換，抑或系統需求的重大變革等因素，均有可能必須對現有資料進行分析處理。筆者接手本案，係在本辭典已經上線測試之時。本文所探討的，即是這種非正規的資料處理案例。

方言標準辭典的音字系統一致性攸關整體成敗，辭典內含資訊龐雜，集體作品難免鑿柄不入。善用電腦資料庫軟體可以輔助歸納資料整體訊息，方便檢視問題資料，並適當改正疏漏，以降低辭典的錯誤率，提高其品質。

本文所述各項校正作業是在連字符尚未完成校對的條件下所做的處理，因此暫時忽略連字符的功用。若能進一步根據連字符所透露的訊息，以詞為單位來整理辭典的用字和注音，則必能更加精確地篩檢出音字系統不一致的資訊，以提高本辭典的一致性。例

如「重」字有「tiông、tîng、tiông、tāng」等音讀，僅看單字音，則此四音皆有可能正確。但若加上連字符，則「嚴重」必為「giâm-tiông」，「沉重」必為「tîm-tāng」，則可進而察覺誤用又音之例，進而加以校正。

此外，在篩檢音字不一的實例之餘，進一步分析其類型，大略可以歸為以下數種：人工手誤、方音差異、用字不同、定音不同、誤用文白、誤用漢字、誤注字音、誤標調號、誤標變音等等。列舉各類問題的實情，並進一步分析其原因，在辭典體例的制定，以至於方言教學的實務上，也能夠提供一些有效的參考數據。

引用文獻

- 教育部國語推行委員會 2007. 《臺灣閩南語推薦用字（第一批）》，
http://www.edu.tw/files/bulletin/M0001/300iongji_960523.pdf。
- . 2008a. 《臺灣閩南語推薦用字（第二批）》，
http://www.edu.tw/files/bulletin/M0001/100iongji_970501.pdf。
- . 2008b. 《臺灣閩南語常用詞辭典（試用版）》，
<http://twblg.dict.edu.tw/tw/index.htm>。
- . 2008c. 《臺灣閩南語常用詞辭典編輯凡例》，
<http://twblg.dict.edu.tw/tw/editor/03.htm>。
- 駱嘉鵬. 2006. 〈漢字同音字表的新做法——以臺灣閩南語字音表為例〉，第 38 屆漢藏語國際學術研討會。中國廈門：廈門大學。

駱嘉鵬
清雲科技大學
luo@cyu.edu.tw

Improvement of the Consistency between Sound and Character in *The Dictionary of the Most Frequently Used Words of the Min'nan Language*: The Application of Database Management System in Dictionary System Revision

Jia-Peng LUO
Ching Yun University

A dictionary is usually regarded as a standard grammar; however, generally the Taiwan Southern Min lexicography is not considered to be user-friendly due to inconsistent characters and sound variants in Taiwan Southern Min. Although *the Dictionary of the Most Frequently Used Words of the Min'nan Language*, edited by the Ministry of Education, had finally been confirmed in 2005, we still find some mismatches between characters and pronunciations. Hence, this study aims to correct these mistakes according to several procedures, such as using the database management system Visual FoxPro to primarily investigate the dictionary content and to do hand-checking correction. Then, we built up the character and pronunciation mapping system of Taiwan Southern Min by organizing all possible combinations with Excel. Meanwhile, we revised the invalid mappings, and put the sound variants into their corresponding character sets. Finally, all sets containing characters and pronunciations were examined and checked artificially. With the help of the computer software, the rate of speed and accuracy is improved. All in all, we would like to share our experiences in revising dictionary system for your reference.

Key words: lexicography, Taiwan Southern Min character and pronunciation, consistency, database management system