

《臺灣閩南語常用詞辭典》 暨《閩南語字彙》網路版擴增連結計畫

駱嘉鵬 李良一

清雲科技大學通識中心講師 國立中央大學資工系博士後研究員

luo@cyu.edu.tw lihenry12345@gmail.com

摘要

教育部國語會於 2008 年初公告《臺灣閩南語常用詞辭典》網路試用版提供各界使用，其後陸續發現諸多問題。此外，由楊秀芳教授主持的「閩南語本字研究專案計畫」，研究成果經出版為《閩南語字彙》一至四集，雖曾轉換為網路版，但因未以資料庫處理，而僅逐字建立靜態網頁，故未能充分彰顯研究成果之價值。筆者駱嘉鵬參與教育部《閩南語常用詞辭典》編輯計畫期間，陸續提供個人意見，並協助處理資料問題，故於當年 9 月接受教育部委託，承辦「《臺灣閩南語常用詞辭典》暨《閩南語字彙》網路版擴增連結計畫」。

本計畫執行內容，主要在於修正《臺灣閩南語常用詞辭典》網路版系統資料及程式的錯誤，提高其使用效能；並建置《閩南語字彙》檢索網站，進而建立兩網站之間的連結。

本文係執行本計畫的成果報告。內容針對《臺灣閩南語常用詞辭典》原系統主要問題及其修正結果，新建《閩南語字彙》網站之主要功能，及兩網站相互連結的處理模式略加說明。藉此提高與會學者對本系統的認識，以加強推廣之功效。並期待聽取各方賢達意見，以做為後續改進之參考依據。

關鍵字：臺灣閩南語常用詞辭典、閩南語字彙、辭典學、資料庫管理、網路辭典

1. 前言

《臺灣閩南語常用詞辭典》網路試用版自 2008 年初公告後，陸續發現諸多問題。由於筆者駱嘉鵬參與本辭典編輯計畫，期間曾多次提供本辭典修正意見，並陸續協助處理資料問題，為使本辭典更臻完善，並結合「閩南語本字研究專案計畫」專案研究成果，乃由教育部委託清雲科技大學承接本計畫案。

本計畫案由駱嘉鵬擔任主持人，並負責系統分析與資料轉檔作業；李良一負責程式設計。自 2008 年 9 月起，執行期程一年，完成後並負責保固一年。

關於執行本計畫的電腦軟硬體設備，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。編輯小組編校作業，原本皆在 Excel 工作表中進行，在新版上線前的資料處理作業，為配合業務單位使用習慣，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。

改版後的系統採用 Java 及 JSP(Java Server Page)開發，Web Server 為 Apache，JSP Server 為 Tomcat，資料庫為 MySQL。資料校正作業亦經由網頁後端管理程式直接登錄於系統資料庫。

本計畫內容分為三部分，一是《臺灣閩南語常用詞辭典》網路試用版的修正，二是《閩南語字彙》網站的建置，三是上述兩系統相關詞目的連結查詢。本文依此順序，略述本案執行成果，以提高學界對辭典與字彙系統網站的認識，加強其使用效能。並藉此聽取各方高見，以做為後續改進的參考。

2. 《臺灣閩南語常用詞辭典》系統維護

2.1 資料處理

2.1.1 資料庫設計

本系統程式原由「財團法人中文資訊推廣基金會」負責設計，系統並未遵循正規化設計原則處理。例如：所有資料全部集中在同一份表單，並未分類、分表存放。表單之中，一個閩南語詞目的所有相關資訊全部集中在同一筆，欄位屬性不夠明確。例如「釋義」欄除了包括詞目的華語釋義，還包括閩南語例句、例句的閩南語注音、例句的華語翻譯。此外，還包括關於該義項的近義詞、反義詞等。若該詞目包括義項不超過四個，則每個義項各佔一欄；至於超過四個以上義項的詞條，則將兩個(甚至三個)義項的上列各種相關資訊全部併入同一欄中儲存。

因原系統的資料庫格式混雜，對於程式設計造成嚴重的限制，故本案雖名為擴增、維護，其實除操作畫面大體沿用舊系統架構之外，資料庫檔案及程式皆屬重新設計。

本案重新設計之後，依照資料屬性，分別拆解至數份工作表存放，並建立必要的關聯，讓使用者能夠設定更精確的檢索條件，而在瀏覽檢索結果時，仍可得到完整的資訊。

2.1.2 資料校正

「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，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。」¹

主持人接手本案之後，即由原資料庫中抽離出所有閩南語用字和注音資訊，包括詞目、詞目注音，以及釋義欄中的台語例句和例句注音。其次對這些台語用字和注音進行音字媒合處理，整理出所有用字在辭典中的音讀和用例，並製作音字媒合表、字音彙總表、聲韻調配合表、聲韻配合表、韻末配調表、字聲配合表、字調配合表、同音字表、例句與詞目字音比對表等各式統計分析表，提供編輯小組，以便發覺問題字音，進而加以校正。

2.2 系統管理

原系統提供詞目、音讀、對應華語及全文檢索四種查詢對象，並搭配精確查詢與模糊查詢兩種檢索模式；精確查詢指檢索對象與輸入檢索值完全相同，模糊查詢則指檢索對象中包含輸入值；由使用者依需求自由搭配。唯其處理邏輯發生諸多錯誤，或雖無錯誤，亦有諸多不便，僅將修正內略述如下：

2.2.1 詞目檢索

本系統詞目，除做為主撰條者外，尚包括地名、姓氏、外來詞等各項附錄中之詞彙。原系統之詞目檢索，僅能查詢辭典主詞目，而不能查得附錄詞目。本案處理之後，已涵蓋辭典所有詞目，皆可查詢。此外，單字不成詞之字音，如「耀」字並非台語詞彙，但人名中含「耀」字者屢見不鮮；又如「上」字在台語雖極為常用，唯其必須搭配其他詞素方能成詞，而不單獨使用。為使一般使用者能查得該類字音，系統亦加以收錄；唯僅列示音讀，不做任何解釋，以與語言學上較為精確的台語詞彙做區別。

2.2.2 音讀檢索

2.2.2.1 拼音系統

本系統採用教育部公告的「臺羅拼音正式版」為標音系統，唯在民間通行者，還有「臺羅拼音數字版」、「臺灣語言拼音方案(通稱 TLPA)」及「教會羅馬字(通稱白話字)」等拼音。原系統只能接受一種拼音，本案將各種拼音方案詳加比較，在使用者輸入拼音時，即將其轉為系統內定格式，故能兼容上列各式拼音使用者；又為推廣部定拼音，故

¹ 引自駱嘉鵬，2008。

在輸出檢索結果時，僅列示「臺羅拼音正式版」。

2.2.2.2 連字符

臺羅拼音正式版以詞為單位，詞素與詞素之間須以連字符連結，如「風吹 hong-tshue」、「風咧吹 hong teh tshue」。原系統在輸入音讀檢索時，其連字符必須輸入準確方可查得。唯就一般使用者而言，何時該連、何時該空往往不易判別，因此常會造成檢索結果不如預期的窘境。本案處理之後，對於使用者輸入的音節間符號，不論是全形空白、半形空白或連字符，一律改為一個半形空白，再進行資料檢索，如此即可免除使用者連字符輸入不當之困擾。

2.2.2.3 對應華語檢索

原系統將各詞目所對應的華語詞全部彙集在一欄之中，並以頓號區隔各詞，如「一世人」的對應華語欄為「一世、一生、一輩子」。如此則當以「精確查詢」模式檢索華語詞為「一生」時，即無法查得此筆紀錄，而僅能以「模糊查詢」模式檢索。但模糊查詢的結果又往往超出使用者需求之外，因此容易造成使用上的不便。本案將對應華語另立一表處理，每筆紀錄僅包含單一詞彙，再以索引鍵建立與原詞目的關聯，如此則不論以精確查詢或模糊查詢方式檢索，皆能查得最適切的結果。

2.2.2.4 全文檢索

原系統雖提供「全文檢索」功能，但其檢索對象實未及於全文，只是整個釋義欄位。本案將其擴大到辭典全文，不論是詞目、音讀、釋義、對應華語、近義詞、反義詞、附錄等，均完全納入。唯附錄中的俗諺、語音差異表與詞彙比較表性質特殊，故未納入。如此已能大幅提高檢索效能，對使用者觸類旁通頗有助益。

2.2.2.5 排序

檢索結果列表的順序，原系統純以音讀為依據。新系統擴大檢索範圍，資料性質較為繁雜，故依據詞目性質類別，先列辭典主詞目、次列不成詞單字音，最後再依序列示各種附錄中之詞目，同一類中再依音讀順序排列。如此較能區分辭典主從架構，不致混雜。

2.2.2.6 造字

對於系統缺字部分，原系統以造圖形字來呈現。唯在資料庫內部代碼與圖形字檔案之間的對應關係處理不當，往往造成畫面空缺的錯誤。本案修正上列錯誤，所有圖形字皆可正常呈現。

2.2.2.7 拼音字形

臺羅拼音須使用特殊字形(如 Charis SIL)方能正確顯示。原系統在列示檢索結果之音讀欄能正確呈現，但在其他字、音並列的欄位中，則為遷就中文字型，而犧牲拼音，以致許多拼音符號出現錯誤。本案將欄位內容單一化，已無字、音並列於一欄之問題，故能在畫面呈現時，依資料性質使用正確字形，而無此類缺失。

2.2.2.8 詞目錄音檔連結

為讓使用者實際體會詞目發音，本系統新增詞目錄音檔連結功能。在各筆詞目瀏覽畫面，點按詞目右側的播放按鈕，即可聆聽該筆詞目的台語發音。對於正音和音標的學習，應可發揮積極正面的功效。

2.2.2.9 靜態網頁

舊系統設計之初，尚處臺羅拼音研擬階段；其後雖已正式公告，但有許多網頁未及修正。故有 TLPA 與臺羅數字版、臺羅正式版雜用的問題，其說明網頁亦多語誤。新系統全面檢視所有靜態網頁，已盡可能將原有錯誤修正。

2.2.2.10 後端管理

舊系統後端管理程式存在諸多問題。例如修改單筆紀錄時，會造成部分欄位內容的漏失；無法上傳、下載整批資料等。新系統重新設計後，已可精確處理各種資料維護問題。

3. 《閩南語字彙》網站建置

《閩南語字彙》為臺大楊秀芳教授主持計畫之研究成果；原本係以紙本刊印，後為提高使用效能，而將其數位化，提供網路版瀏覽功能。唯該版僅將原書內容逐字鍵入靜態網頁，並未提供資料庫檢索功能。原計畫考證閩南語本字數百筆，該網站即逐筆建立數百個靜態網頁，不但不易管理，查詢時亦多所不便。

本案即以上述資料為本，將其轉為資料庫檔案，並設計網站檢索系統，提供使用者快速檢索相關語料。

3.1 資料處理

3.1.1 資料庫設計

本案原始資料係以 Word 文檔格式儲存，其後轉換為靜態網頁，僅在各單字頁面之外另建檢索連結頁面，並未依正規化資料庫格式儲存處理。筆者分析資料屬性後，即重新加以規畫，並將所有文本資訊轉入新資料庫相對應的表單、欄位之中，以利後續程式設計之用。

3.1.2 拼音轉換

原始資料所用拼音，屬 TLPA 系統。為配合部定政策，乃將其完全轉為臺羅拼音正式版。

3.2 系統管理

關於檢索方式，本系統分別採用分層瀏覽與全文檢索兩種方式處理。分述如下：

3.2.1 分層瀏覽

關由於本資料庫紀錄筆數不多，對於不熟悉台語音讀、用字者層層引導，有助於使用者掌握全文內容。本系統提供的分層檢索方式如下：

3.2.1.1 部首索引

點選「部首索引」按鈕後，列出部首筆畫數以供進一步點選，然後即依序呈現符合該部首筆畫數之所有部首，並在各部首之下列出《閩南語字彙》中包含的詞目。使用者點按所需詞目，即可瀏覽該詞目相關音讀、釋義及例句等內容。

3.2.1.2 筆畫索引

仿照「部首索引」操作模式，點按「筆畫索引」按鈕後，再點按筆畫數，即分部首列出該總筆畫數之所有詞目以供點閱內容。

3.2.1.3 聲母索引

點選「聲母索引」按鈕後，列出所有聲母以供進一步點選，然後即依韻、調順序呈現資料庫中符合該聲母之所有韻調，並在各韻調之下列示相關詞目，以供使用者瀏覽該詞目內容。

3.2.1.4 韻母索引

點選「韻母索引」按鈕後，列出所有韻母以供進一步點選，然後即依聲母及聲調順序呈現資料庫中符合該音讀的所有詞目，以供使用者瀏覽內容。

3.2.1.5 詞例索引

此為依照詞義屬性所做的分類索引。先分天文地理、人體、人事活動、物類及其他等五大類，點按大類之後再呈現該大類內部之細類，再點按細類之後即表列該細類所有詞目，包括國語用詞、通俗寫法、閩南語用詞及詞例出處等項目，點按「詞例出處」中的連結，即可瀏覽該詞目詳細內容。例如點按「天文地理」大類中的「天文」類時，即列示該細類所有詞目簡表如下：。

國語用詞	通俗寫法	閩南語用詞	詞例出處
天	天	天 ◎參看詞條 「天頂」	頂 1
太陽	日頭	日頭	頭 22
日蝕	蝕日	蝕日	蝕(二)
月蝕	蝕月	蝕月	蝕(二)
空中	空中	空中	空(三)1

(下略)

圖1 《閩南語字彙》網站詞例細類瀏覽畫面

3.2.2 全文檢索

本系統亦提供全文檢索功能，大抵與閩南語辭典類似。唯因系統資料較少，故相對亦較單純。本系統提供字彙、音讀及全文三種檢索方式：

3.2.2.1 字彙檢索

仿照閩語典系統的詞目檢索功能，以《閩南語字彙》的詞目為檢索對象。由於字彙詞目皆為單字詞，故僅能輸入單字檢索，內定檢索模式為「精確查詢」。

3.2.2.2 音讀檢索

仿照閩語字典系統的音讀檢索功能，以《閩南語字彙》詞目的音讀為檢索對象。亦僅能輸入單字音檢索，內定檢索模式亦為「精確查詢」。

3.2.2.3 全文檢索

仿照閩語字典系統的全文檢索功能，以《閩南語字彙》的所有內容為檢索對象，內定檢索模式為「模糊查詢」。

4. 辭典與字彙網站連結

根《閩南語字彙》係閩南語本字考證研究的成果。其具體做法，簡而言之，係在閩南語的音義及古今音類對應規律的基礎上所做的本字考證研究；獲得的成果，有利於閩南語史的研究、古文的理解以及閩南語用字的釐訂。《臺灣閩南語常用詞辭典》的製作，則主要係為廣大台灣閩南語使用民眾而設計，目標著重在教學和一般使用上的便利、合理性。辭典的音讀、用字和釋義，固應依循學術研究的成果，唯在實用上，則須加入更多的考量。以用字為例，字彙研究的成果，固然可以考證出方言詞彙的本字，但考證出來的本字，卻可能因為太過冷僻、太過多義或尚無定論，甚至無字可考等因素，不得不捨棄本字而另覓它字來使用。此在教育部〈臺灣閩南語漢字之選用原則〉中已有說明，不必贅述。

為便於字彙研究者得以參見其在台語使用之實例，以及辭典的使用者得以進一步探究本字，本案除進行辭典網站功能的擴充與修正和字彙網站的建置外，並建立兩網站相關詞彙的連結。為讓使用者便於參閱閩南語字彙相關同源詞及異體字，並建立《閩南語字彙》網站與《教育部異體字字典》相關詞目的連結。簡述如下：

4.1 字彙網站連結辭典

《閩南語字彙》網站的各項檢索功能，最終都可導入單筆詞目的瀏覽畫面，而若該字有多種音讀，則會分音呈現其各種釋義和例句。在兩系統完成之後，本案並進而連結相關詞目。對於任一詞目的任一音讀，若在辭典中有相對應的詞目，則在音讀右側，即會呈現「[連結閩語辭典](#)」訊息，點按該提示文字，即可另開視窗呈現辭典相應詞條內容。例如查詢「分」字出現如下結果：

【分】 刀—2—4 

(一)pun	(二)hun 	(三)hūn 
--------	--	---

(一)pun [連結閩語辭典](#)

1. 分開。

「頭毛分做兩片。」Thâu-m̄ng pun tsò n̄ng-pîng.：將頭髮向兩邊分開。

「分隨儂」pun suí lāng：兄弟分家，各自為食。

2. 分配、發放。

「老師在分簿仔。」Lāu-su teh pun phōo-á.：老師在發放簿本。

3. 給予。

「提錢分乞食。」Theh tsinn pun khít-tsiah.：拿錢施捨給乞丐。

「我糖仔分汝一半。」Guá thng-á pun lí tsit-puànn.：我把糖果分一半給你。

「我糖仔分汝食。」Guá thng-á pun lí tsiah.：我把糖果分給你吃。

4. 乞求施捨。

「乞食在分。」Khít-tsiah teh pun.：乞丐在乞求施捨。

5. 收養子女。

「分兩個後生來飼。」Pun n̄ng ê hāu-senn lâi tshī.：領兩個兒子來養。

「阿英是分的。」A-ing sī pun--ê.：阿英是領養的。

6. 從全部中取得部分。

「兄弟仔分家伙。」Hiann-tī-á pun ke-hué.：兄弟分祖產。

「我分著三粒。」Guá pun-tioh sann liap.：我分到三顆。

「對半分」tuì-puànn pun：從中各分一半。

(二)hun [連結閩語辭典](#)

(下略)

圖2 《閩南語字彙》網站單詞瀏覽畫面

點按上列畫面詞目下一列的各種音讀，即可連結到該音讀的各個義項說明及例句；而若點按音讀右側的「[連結閩語辭典](#)」字眼，即可連結到閩語典相應的「分 pun」詞條，顯示如下內容：

詞目	分 部首：刀 018-02-04
音讀	 pun 連結閩南語字彙
釋義	1.  發出。例：分帖仔 pun thiap-á(發帖子)。 2.  分配。例：對半分 tuì-puànn pun(平均分成一半一半)。 3.  施捨。例：好心的頭家！一碗飯來分我好無？Hó-sim ê thâu-ke! Tsit uánn pn̄g lâi pun--guá hó--bô? (好心的老闆！可以施捨一碗飯給我嗎？) 4.  領養。例：家己若袂生，規氣去分別人的囡仔來飼。Ka-kī nā bē senn, kui-khì khì pun pat-lâng ê gín-á lâi tshī. (自己若是生不出孩子，乾脆去領養別人家的孩子。); 分人做

	新婦仔。Pun lâng tsò sin-pū-á. (給人領養當童養媳。)
又見音	 hun 、hūn

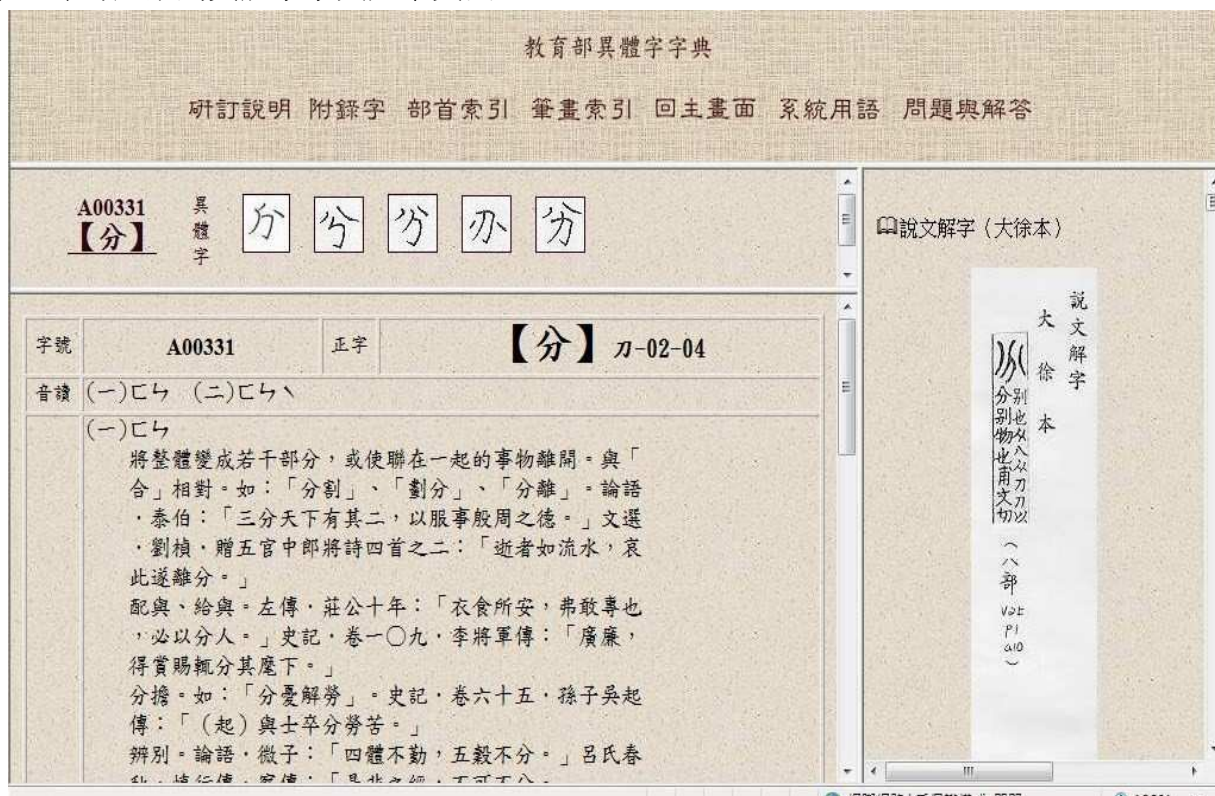
圖3 《臺灣閩南語常用詞辭典》單詞瀏覽畫面

4.2 辭典連結字彙網站

閩語典詞目瀏覽畫面亦設計連結《閩南語字彙》網站相關字彙的功能。例如點按上列圖3音讀欄右側「[連結閩南語字彙](#)」提示訊息，即可另開視窗呈現圖2內容。唯因教育部考量閩語典甫推出不久，正值大力推廣部定用字之際，學術研究成果，似非大眾迫切需求。且若遇部分辭典用字與字彙考證之本字不同者，恐增加使用者困惑。故此功能雖已完成，目前實未正式開啟，而僅以變數控制其開關，將俟未來有需要時再行開啟。

4.3 字彙連結異體字字典網站

《閩南語字彙》係本字考證的成果，在考證過程中，往往會牽涉到許多相關的同源字或異體字。由於教育部原已建置一套《異體字字典》網站，為便於研究者深入探討，本字彙網站亦提供異體字字典相應詞條連結功能。例如點按上列圖2首列右側的書本圖示，即可連結到異體字字典如下頁面：



教育部異體字字典

研訂說明 附錄字 部首索引 筆畫索引 回主畫面 系統用語 問題與解答

A00331 異體字 **分** 𠂔 𠂕 𠂖 𠂗 𠂘

字號	A00331	正字	分 刀-02-04
音讀	(一)ㄈㄣ (二)ㄈㄣˋ		

(一)ㄈㄣ

將整體變成若干部分，或使聯在一起的事物離開。與「合」相對。如：「分割」、「劃分」、「分離」。論語·泰伯：「三分天下有其二，以服事殷周之德。」文選·劉楨·贈五官中郎將詩四首之二：「逝者如流水，哀此遂離分。」

配與、給與。左傳·莊公十年：「衣食所安，弗敢專也，必以分人。」史記·卷一〇九·李將軍傳：「廣廉，得賞賜輒分其麾下。」

分擔。如：「分憂解勞」。史記·卷六十五·孫子吳起傳：「(起)與士卒分勞苦。」

辨別。論語·微子：「四體不勤，五穀不分。」呂氏春秋·德化篇·察傳：「見北之師，不可不八。」

說文解字 (大徐本)

大徐本

分別也。从刀，刀以分別物也。用文切。

(八部 刀部)

100%

(下略)

圖4 《教育部異體字字典》單字瀏覽畫面

由於《異體字字典》除提供各類字形以供參考之外，亦具有一般辭典注音、釋義之功能，並在單字瀏覽畫面右側列示相關辭書的原始頁面，且其收錄字數超過十萬，對於本字考證的研究，自是相當有價值的參考資訊。

5. 結論

本文說明「《臺灣閩南語常用詞辭典》暨《閩南語字彙》網路版擴增連結計畫」主要執行成果。在妥善的資料庫設計及系統規畫下，改進原《臺灣閩南語常用詞辭典》網站諸多缺失，提升辭典內容的精確性及其它使用效能。將《閩南語字彙》研究成果數位化之後，改變原本逐筆製作靜態網頁的做法，而以資料庫軟體重新設計；既便於業務單位管理，更有利於使用者檢索。至於兩網站間的連結，以及《閩南語字彙》與《教育部異體字字典》網站間的連結功能，更有利於研究者的深入探析。

本系統雖經重新設計，改善其使用效能，唯系統程式及資料，仍有許多改進空間。教育部原訂在系統推出後試用二至三年，再參酌各方意見，修訂之後，推出正式版。目前正式版之內容部分，正由本辭典編輯委員會繼續研議；網站功能方面，則仍委託本校承辦，由筆者接續執行，預計八個月之後完成。相信在未來一年之內，必能使閩語典的內含資訊及系統效能更臻完善。

參考文獻

- 教育部國語推行委員會. 2004. 《教育部異體字字典》。
<http://dict.variants.moe.edu.tw/main.htm>。
- . 2007. 〈臺灣閩南語漢字之選用原則〉。
http://www.edu.tw/files/bulletin/M0001/guantsik_960523.pdf。
- . 2008. 《臺灣閩南語常用詞辭典（試用版）》。
<http://twblg.dict.edu.tw/tw/index.htm>。
- . 2009a. 《閩南語字彙》。<http://blgjilui.dict.edu.tw/>。
- . 2009b. 《「臺灣閩南語推薦用字 700 字詞」手冊》。
<http://www.edu.tw/files/download/MANDR/book-one-sided.pdf>。
- 駱嘉鵬. 2008. 〈《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〉，第七屆臺灣語言及其教學國際學術研討會，國立台灣師範大學。