

台灣學術網路的閩客語資源現況

首屆海峽兩岸現代漢語問題學術研討會

主辦單位：天津南開大學

2005年11月5-6日

羅鳳珠·台灣元智大學中國語文學系

台灣學術網路所建置的語文資源以國語文佔大宗，閩客語資源的建置起步較晚，數量也比較少，但正因為起步較晚，建置閩客語資源網站時少了很多摸索的時間。

台灣近年所執行的數位典藏國家型科技計畫，迄今已邁入第五年，所建置的閩客語數位資源，內容方面包括小說、散文、戲曲劇本、歌謠、諺語；語音方面包含音標系統、語音資料；語料方面包括小說與戲劇語料、生活用語語料；網路教學方面，包括民歌諺語、臺灣文學、生活會話、語音；字辭典工具書方面包含閩客語字辭典；語言調查之空間資訊方面有新竹新豐鄉閩客語分布地理資訊系統。在網站功能方面，包含典藏、研究、教學與推廣四種。

本文將介紹台灣學術網路所建置的閩客語資源以及建置的理念與方法。

全文

台灣學術網路所建置的語文資源以國語文為大宗，閩客語資源的建置，起步比較晚，資源比較少。閩客語資源的建置者，或來自政府單位國家計畫的推動，或由閩客語研究者、教學者、民間個人、民間團體自發性建置。正因為起步比較晚，所以建置的過程省去很多摸索的時間，但也因為投入的人力、經費等資源比較少，成果遠不及國語文資源。

閩客語建置的內容包含小說、散文、戲劇、歌謠、諺語等文本、語音、語料、研究著作、教學網站、語言調查地理資訊、字辭典工具書等資料；網站的功能包含典藏、研究、教學、推廣等方面。今舉其要者介紹如後。

一、推動及建置者

- (一) 數位典藏國家型科技計畫：因感於數位典藏的重要，臺灣政府行政院於 2000 年 7 月通過「國家典藏數位化計畫」，之後由國家科學發展委員會整編為國家型計畫，於 2002 年 1 月 1 日正式展開，並將計畫名稱訂為「數位典藏國家型科技計畫」，在這個國家型計畫之下，建置了「荔鏡姻，河洛源—閩南語第一名著《荔鏡記》數位博物館」、「閩南語典藏—歷史語言與分佈變遷資料庫」、「不惜歌者苦，但傷知音稀：詩詞曲文三語吟唱讀教學網站」等三個網站。
- (二) 行政院僑務委員會：由僑委會建置的「全球華文網路教育中心」提供臺灣閩南語及客語教材。
- (三) 行政院客家委員會：建置「哈客語言館」、「(客家)文史天地館」，內有「哈客網路學院」，提供客語網路教學教材。
- (四) 行政院文化建設委員會：建置國家文化資料庫，內有「鍾理和數位博物館」、「賴和數位博物館」。
- (五) 閩客語研究者：成功大學臺灣文學研究所呂興昌教授主持的「臺灣文學研究工作室」、大漢技術學院資訊工程學系楊允言教授主持的「台語文 e 網站」、國立交通大學外文所暨語言學研究所鄭良偉教授主持的「臺灣語文資訊網站」。
- (六) 民間個人及團體：臺灣大學客家研究社以及呂興昌、楊允言、王順隆三位教授都曾投入多年心力。

二、網站內容

- (一) 文本網站：以文本內容為網站主要建置內容，文本的類別包含：
 1. 語體文：吳濁流、龍瑛宗、鍾理和、賴和、鍾肇政、李喬、鍾鐵民等作家之全部或部分文學作品。
 2. 戲曲劇本：荔鏡記（萬曆、光緒、順治、嘉靖本）、同窗琴書記、蘇六娘、金花女、陳三五娘。
 3. 歌謠：閩南語歌仔冊、客家歌本。
 4. 諺語：臺灣諺語
- (二) 教學網站：
 1. 行政院僑務委員會「全球華文網路教育中心」：
「全球華文網路教育中心」的網路教學資源，台語教學包含台語開講、俚諺俗語、鄉土歌謠等文字及語音資料。客語教學包含四線與海陸腔客語日常用語教材之文字及語音、詞語解釋。

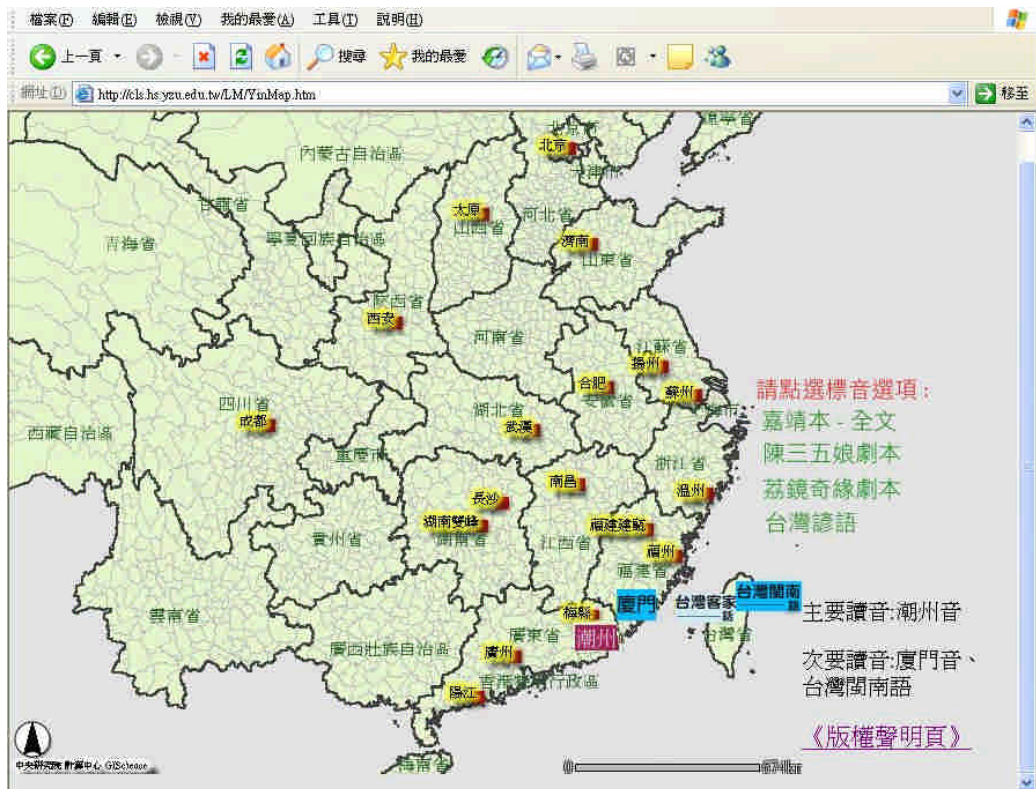
2. 行政院客家委員會「哈客網路學院」：
網路教學資源包括「哈客語言館」及台灣饒平、大埔、詔安客語教學資源中心，多數教材只提供聲音資料，未提供文字資料。
3. 九年一貫鄉土語言台語（閩南語）教材資源中心：
這是由楊允言教授整理的閩南語教材入口網站，提供教材相關訊息。
4. 九年一貫母語（閩南語）教材：
這是由蕭平治老師編寫建置的教學網站，內容分爲：幼稚園、國小、臺灣囡仔歌、囡仔故事、俗語等單元，教材內容包含文字以及生難詞語的音標、註解。
5. 客語 e 學堂：
這是由台北市政府客家事務委員會所建置的網站，「客語 e 學堂」是網站內容的一部份，教學內容以客語會話爲主，分爲：客語 e 學堂（初級、中級課程）、日常問候語（飲食、生活、稱讚三類）、客語師傅話，以上三類教材均提供文字及聲音資料，聲音資料含國語及海陸、饒平、四縣、詔安四種腔調。
6. 不惜歌者苦，但傷知音稀：詩詞曲文三語吟唱讀教學網站：
這個網站是筆者受教育部及國科會委託所建置，主要目的是以歷代詩詞曲文爲素材建立閩客語讀書音吟讀教學，教材內容包含文字及聲音資料，聲音資料提供閩客語的吟讀唱，書面教材提供洪澤南老師撰寫的閩南語傳統八音，包括：說吟調、說平仄四聲八音、說句法、說文音、說勾破、說韻目、說轉調、說四聲遞用、說雙聲疊韻等，是學習讀書音不可多得的教學資源。

（三）語音、文字及語料網站：

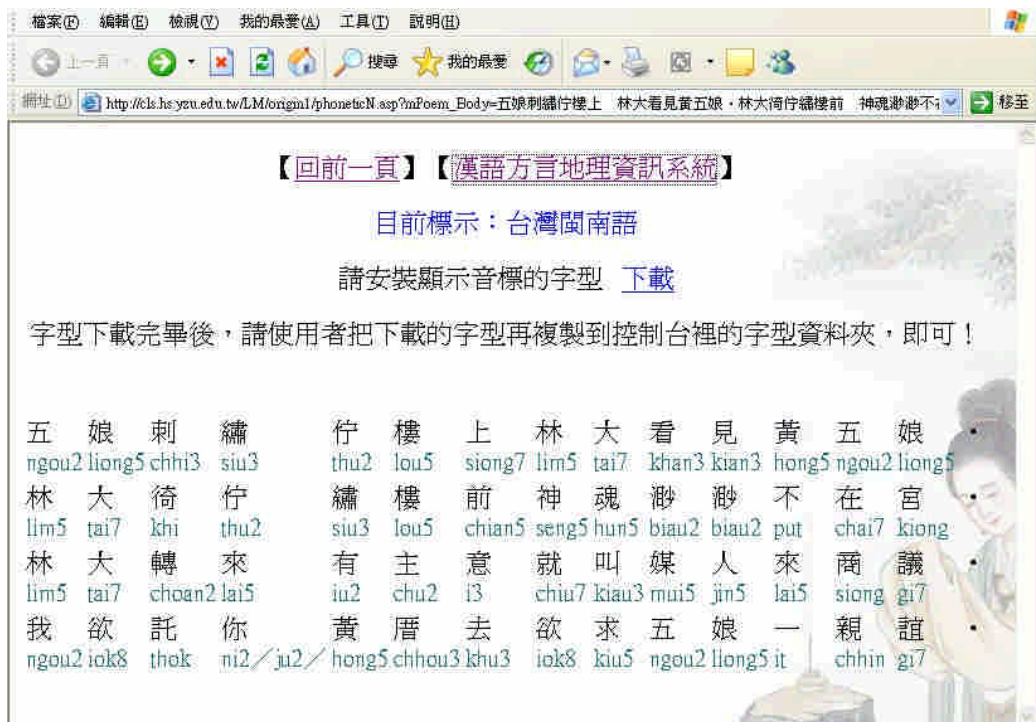
1. 字辭典語料：
「台文／華文線頂辭典」收有 62,035 筆台語詞條，台北市客委會「現代客語詞彙彙編」收有 22,600 餘筆客語詞條，臺灣大學客家研究社建置的「線上客語有聲字典」共收中文字 9,837 個，四縣腔 11,784 個音，大埔腔 1,119 個音，「客語小詞典」共收大埔腔 1,333 筆、四縣腔 1,098 筆、南部四縣腔 33 筆。
2. 教育部閩南語字彙網站：
教育部委託臺灣大學中文系楊秀芳教授蒐集整理，嚴謹考證閩南語本字，建置「閩南語字彙」網站，提供部首、音序、詞例索引，詞例分爲天文地理、人體、人事活動、物類、其他等五類，依類別列出國語用詞、通俗寫法、閩南語用詞、詞例出處對照表，每一筆詞例列出衍生之詞彙、讀音、詞義。
3. 海陸注音檢定詞彙：
收有 2,148 筆客語海陸注音檢定詞彙，每一筆詞條都加上國語詞彙字義。

4. 閩南語讀書音資料庫：
閩南語及客語都可以分爲讀書音（文音、孔子白）及說話音，多數人能說說話音，但不熟悉讀書音。筆者所建置的「荔鏡姻，河洛源：閩南語第一名著《荔鏡記》多媒體教學網站」，所收 131 段歌仔戲陳三五娘劇本歌詞，以純正閩南語讀書音錄製語音資料。筆者所建置的「不惜歌者苦，但傷知音稀：詩詞曲文三語吟唱讀教學網站」，以閩南語及客語之讀書音錄製之詩詞曲文吟讀資料，計有閩南語 605 篇，客語 60 篇。
5. 語料庫建立蒐集計畫：
這是由大漢技術學院楊允言教授所主持的計畫，廣泛蒐集台語語料，應用於台語文 Concordance、台語文詞頻統計、編寫辭典(Lexicography)，並公開提供台語文研究者使用。
6. 文學作品語料：
由筆者與學界學者共同主持的研究計畫，建立了一些閩客語文學文本，並且從這些文學文本裏抽取閩客語語料，客語語料包含：〈渡台悲歌〉、鍾理和文學作品客語語料（鍾怡彥整理）；閩南語語料包含：《荔鏡記》全文詞彙種數 17,589 筆，其中閩南語詞彙種數 4,541 筆。
7. 讀音自動標註系統：
筆者引用鄭錦全院士「方言讀音採集計畫」所採集的廈門、潮州等閩南語讀音及筆者研究計畫所建置的客語讀音，建立讀音自動標註系統，提供在任何文本上自動標註閩客語讀音音標的功能，以「陳三五娘歌仔戲劇本」爲例，讀音自動標註平台如圖一及讀音標註結果如圖二：

圖一：讀音自動標註平台介面圖：



圖二：讀音標註結果（陳三五娘劇本）



(四) 台、客字辭典網站：

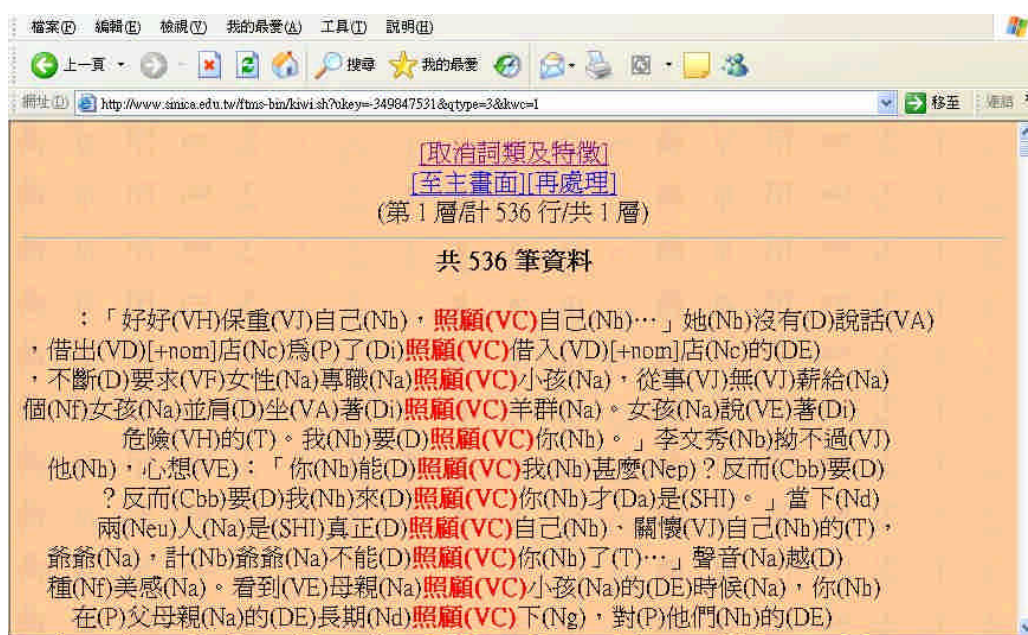
1. 台文／華文線頂辭典：

總計蒐集 62,035 筆詞條，其中使用者增加的詞條有 195 筆，提供的查詢欄位包括台文（漢羅、羅馬字）、華文、羅馬字（無聲調）三種；提供的查詢方式包含：完全比對、部分比對、詞頭、詞尾等四種。查詢結果顯示台語羅馬字、漢羅、華文，所查詢的華文詞彙可以進一步連結中央研究院平衡語料庫顯示該詞彙之出處，並標示其詞類。以華文的「照顧」為例，選擇「完全比對」，查詢結果如下：

	台語羅馬字	漢羅	華文
1	chhiáⁿ	晷	照顧 / 養育
2	chiàu-kò	照顧	眷顧 / 照拂 / 照應 / 照顧
3	kám-kài	感介	照顧 / 關照
4	khòaⁿ-kò	看顧	看管 / 看護 / 看顧 / 照料 / 照顧 / 護理
5	táⁿ-chah	打 chah	支援 / 打點 / 協助 / 照顧 / 幫助

任何一筆華文詞彙，都可以連結中央研究院平衡語料庫，顯示文本出處及詞類標記，以「照顧」為例，顯示結果如圖三：

圖三：中央研究院平衡語料庫顯示查詢結果



2. 台語線頂字典：總計收 22,080 個字，其中 2,726 個字沒有台語音節。主要的用途是查某一個漢字的台語讀音，並可擴充查詢該漢字的衍生詞，其餘用法、功能都與《台文／華文線頂辭典》相同。
3. 台大客家社客語有聲字典：共收中文字 9,837 個，四縣腔 11,784 個音，大埔腔 1,119 個音。提供「中

文單字查其讀音或讀音查其文字」的查詢功能，查詢結果提供語音之聲音資料。

4. 台北市客委會現代客語詞彙彙編：

本詞彙彙編收錄二萬二千六百餘詞條，編寫方式為四縣詞彙、標音、海陸詞彙、標音、國語詞彙的對應，提供客語聲韻表、客家造字表以及客語互動搜尋等功能。

(五) 自動翻譯系統

1. 國客語翻譯系統：

輸入一句國語的讀書音或說話音，可以翻譯為教會羅馬音標或臺灣客語音標的客語。

2. 臺灣本土語言互譯及語音合成系統：

這是臺灣大學資訊工程研究所陳信希教授主持的「自然語言處理實驗室」所建置的網站，以國、台、客語輸入句子，即可自動發出國語讀音、台語的漳州或泉州音的讀音，客語的說話音或讀書音之音標。

3. 華台轉換系統：

由台語信望愛網站所開發的華台轉換系統，可以匯入整段文字或整篇文章，系統自動轉換為台文。

(六) 語言調查地理資訊網站：

由中央研究院鄭錦全院士主持的「閩南語典藏：歷史語言與分布變遷資料庫」計畫，全面調查鄭院士故鄉新竹新豐鄉閩客語之分布與變遷，以新豐鄉作為調查的目標，主要的考量，是當地是閩客語及外省籍族群雜居的鄉鎮，隨著經濟環境的變遷、教育程度的普及等等因素，人口結構產生變化，連帶使生活語言互有消長。該計畫收錄閩客語語音，並根據調查結果建立語言分布之地理資訊系統，以作為未來全面調查語言分布之基礎，這是目前在台灣將語言調查與地理資訊結合的最大型計畫。目前已完成上坑、鳳坑、埔和、中崙、松林、員山、福興、坡頭、後湖、新豐、青埔、瑞興、重興等十三個村落 13,218 戶的調查，使用客語計有四縣 225 戶，海陸 5,801 戶，總計 6,026 戶，佔全鄉的百分之五十。調查結果繪製語言分布地理資訊系統如圖四（以不同顏色的圓點區別不同語言）：

圖四：語言分布地理資訊網頁



(七) 綜合及全文資料庫型網站：

1. 台語文 e 網站：

這是由楊允言教授整理維護的台語文入口網站，有自建的資源，有連外的資源。網站內容分為字辭典、語料庫、生活、個人網站、blog 網誌、教學、社團、台語運動、台語輸入軟體、研究、刊物等。

2. 臺灣語文資訊網站：

這是由臺灣交通大學外文所暨語言學研究所鄭良偉教授建置的網站，內容包括臺灣語言研究、臺灣語言資料庫、台語教育資料庫等單元，有自建的內容，也有連外的內容。

3. 臺灣文學研究工作室網站：

這是由成功大學臺灣文學研究所呂興昌教授自力自費建置的網站，網站內容包含：台語軟體、母語文學、民間文學、古典文學、原住民文學、鄭轄時期、清領時期、日治時期、國統時期、台灣文學研究學者、學位論文、研究書目、研究刊物、學術會議、文學通論、出土資料。

4. 客家人母語教學百花園：

這是一個綜合性的客家資源入口網站，從網站無法看到網站主持人。蒐集的內容非常豐富，分為：教材園地、記音課程、詞彙對照、字音研究、注音諺語、昔時賢文、三字經、注音四句、渡台悲歌、創作新詩、客家童謠、創作歌謠、網路客家、客語典藏等，是一個非常詳備完整的客家資源入口網站。

5. 荔鏡姻，河洛源：閩南語第一名著《荔鏡記》多媒體教學網站：

這是由筆者所主持的數位典藏國家型科技計畫所建置的網站，以現存最早以閩南語撰寫的劇本《荔鏡記》為建置的內容，提供教學、研究、典藏的功能，內容包括原始文獻、改寫劇本、語料、語音、研究資料等。

6. 閩南語典藏：歷史語言與分布變遷資料庫

這是由鄭錦全院士所主持的數位典藏國家型科技計畫所建置的網站，延續「荔鏡姻，河洛源：閩南語第一名著《荔鏡記》多媒體教學網站」擴充建置，文本內容從嘉靖本《荔鏡記》擴充到萬曆、順治、光緒本，增加了同窗琴書記、蘇六娘、金花女、閩南語歌仔冊、客家歌本以及語言（閩客語）分布調查資料庫

7. 閩南語俗曲唱本「歌仔冊」全文資料庫：

這是由王順隆教授建置，收有臺灣及大陸地區出版的歌仔冊閩南語俗曲唱本六百餘種，資料庫需經申請方可使用。

三、建置的理念與方法

文前已提及閩客語資源的網站建置，起步晚，數量少，臺灣推行閩客語教學，所遭遇的最大困難是教學素材不足，多數老師只能以閩客語的歌謠及俗諺語當作教材，加上音標體系的混亂、用字無統一標準，更增添了教學及推廣的困難。廣泛閱讀文學作品是培養語文能力的重要方法，學習中英文是如此，學習閩客語也是如此，但是以閩客語書寫的文學作品比較少，建置在網路上的資源更少。以下以《荔鏡記》網站為例，說明建置的理念與方法。

筆者與鄭錦全院士、黃居仁、楊秀芳、曾淑娟三位教授選擇《荔鏡記》作為建置閩南語教學網站的素材，主要的考量是戲曲劇本保留了大量的口語語料，是建立方言教學最好的素材。《荔鏡記》是現存文獻之中最早以閩南語（潮州話、泉州話）寫成的劇本，成書的年代在十六世紀。除了潮劇之外，《荔鏡記》還被改編成高甲戲、黃梅戲、歌仔戲、梨園戲、南管等地方戲曲，不同時代的改編劇本，又加入了該時代、該語種的口語，因此保留了十六世紀以來閩南地區豐富的口語資料。

好的教學需要有好的研究作基礎，《荔鏡記》網站的建立兼顧典藏、研究、教學、傳播的功能，主要的架構包含：原劇文獻、表演藝術、兒童劇場、研究資料、語文教學區、使用者回饋區、系統簡介等七部分。原劇文獻分原劇書影和嘉靖本全文，並提供檢索查詢、標注字詞意義、標注閩南語／潮州話／廈門話的讀音等功能。表演藝術分劇本資料和演出資料，蒐集了歌仔戲《陳三五娘》和南管漢唐樂府《荔鏡奇緣》的演出資料。劇本資料也能標注上述各種讀音，而且也請專人錄製閩南語讀書音。兒童劇場

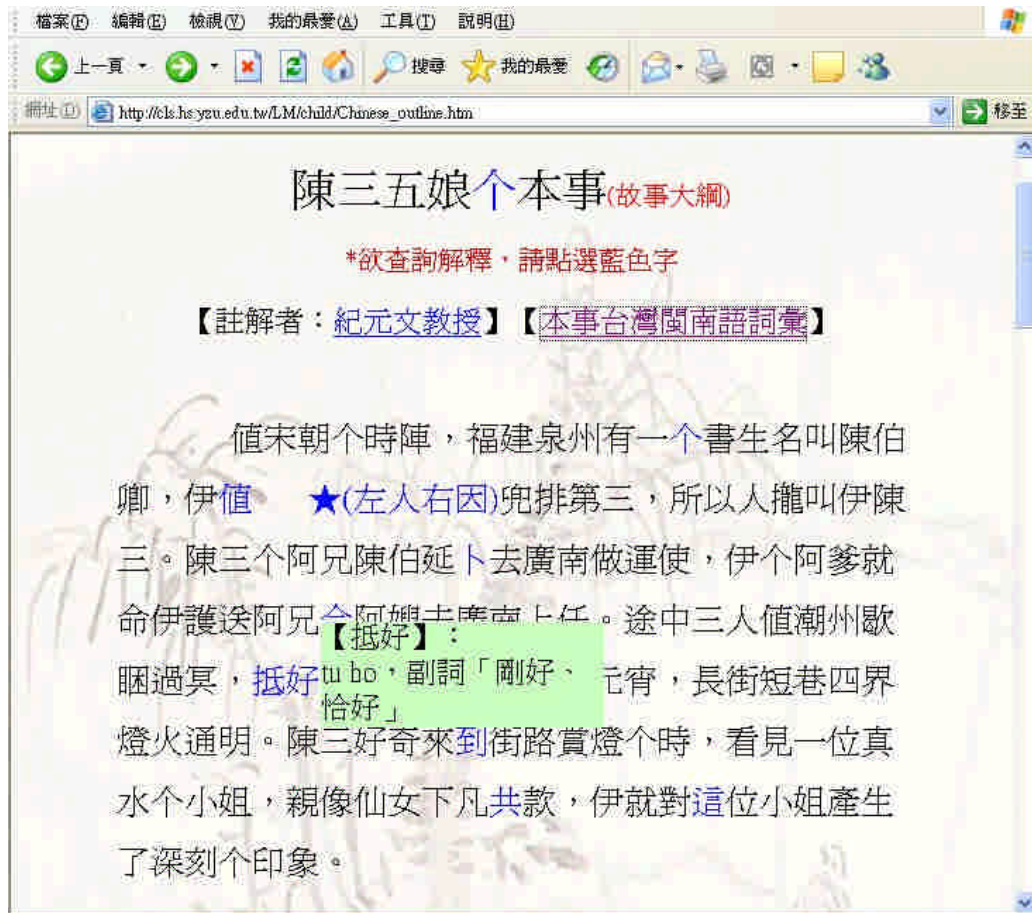
是陳三五娘的兒童故事改編本，計有國語、閩南語、英文、西班牙文等四種語文版。研究資料中包含相關研究論著目錄資料、相關研究論著全文資料、相關研究網站、鄉土語文教學參考工具書等，方便讀者從事進一步的研究。語文教學區包含國語／閩南語同義詞的對照、陳三五娘故事的語音教學、語言演進、潮州語言辭典、俗諺語等單元，提供非同步遠距多媒體語文教學環境，使用者回饋區提供使用者回饋資料的平台、系統簡介等介紹系統相關訊息。

這個網站建置的目的是為了提供閩南語研究、教學及閱讀閩南語文本使用，因此從文本整理出閩南語詞彙，建立閩南語與國語同近義詞對照、國語詞義解釋、原文例句之索引表，以供查檢，如圖五。在系統功能的設計上，提供了閩南語詞彙隨選隨譯的功能如圖六，以降低閱讀時因為不懂閩南語詞彙語義而產生的閱讀障礙。

圖五：閩南語與國語同近義詞網頁



圖六：閩南語詞彙隨選隨譯功能介面



四、結言

如前所言，閩客語相關資源的網站，多數由民間或學者、教師自費建置，因此首先遭遇的困難是經費人力的不足，只能建立簡單的網頁、提供簡單的功能，且難以長期維持網站運作；再者，無論是政府單位、學校或民間個人，各自建置，內容難免重疊，縱有不重疊處，也因各自建置，未能建立一些共同性的標準，致使資料無法相容、無法交換、無法從不同的使用需求作不同的加值應用。三者，難有整體性的規劃，資源與人力均無法整合，事倍功半。凡此種種，都使得有限的資源與人力無法作最好的使用，使成果資源的累積更緩慢，更困難。

閩客語數字化資源的建置，無論從語言的學習、文獻典藏、學術研究等任何角度來看，都需要有完整的規劃。例如，有音無字是所有方言典藏所遇到的共通性問題，閩客語亦是，因此，建立閩客語網路資源時，如何結合學術研究成果，將語料、音標、本字考訂、缺字處理等基本問題解決，是當務之急。

由國家型計畫全面規劃建置閩客語資源，建立文字、語音、音標、影像、圖像等不同媒體資源的數字化典藏標準，優先建立字辭典等工具性的資料庫，文

字輸入等工具軟體，結合語言、文學、資訊科技、圖書資訊四個領域的專業人才，結合政府與民間，學界與業界的力量，訂定輕重緩急的順序，以國家資源有計畫的建立閩客語資源，才是長久之計。

附註：

- 一、本文完成於 2006 年 4 月，所有網站資料以這個時間所見為準。
- 二、本文述及臺灣話時，以「閩南語」或以「台語」敘述，完全依照所介紹網站所使用的名稱。
- 三、本文所述及之各網站名稱及網址，依行文順序分列如下：
 1. 數位典藏國家型科技計畫：<http://www.ndap.org.tw/>
 2. 荔鏡姻，河洛源—閩南語第一名著《荔鏡記》數位博物館：
http://cls.hs.yzu.edu.tw/LM/Lm_home.asp
 3. 閩南語典藏—歷史語言與分佈變遷資料庫：<http://southernmin.sinica.edu.tw/>
 4. 不惜歌者苦，但傷知音稀：詩詞曲文三語吟唱讀教學網站：
http://cls.hs.yzu.edu.tw/CL_POEM/
 5. 行政院僑務委員會全球華文網路教育中心：<http://edu.ocac.gov.tw/>
 6. 行政院客家委員會哈客網路學院：<http://www.hakka.gov.tw/>
 7. 鍾理和數位博物館：<http://cls.hs.yzu.edu.tw/ZHONGLIHE/>
 8. 賴和數位博物館：<http://cls.hs.yzu.edu.tw/laihe/>
 9. 臺灣文學研究工作室：<http://www.ncku.edu.tw/~taiwan/taioan/>
 10. 大漢技術學院資訊工程學系楊允言教授主持的台語文 e 網站：
<http://203.64.42.21/taigu.asp>
 11. 國立交通大學外文所暨語言學研究所鄭良偉教授主持的臺灣語文資訊網站：<http://taigu.eic.nctu.edu.tw/>
 12. 臺灣大學客家研究社：<http://club.ntu.edu.tw/~hakka/>
 13. 九年一貫鄉土語言台語（閩南語）教材資源中心：
<http://203.64.42.21/iug/ungian/POJ/poj.asp>
 14. 九年一貫母語（閩南語）教材：<http://203.64.42.21/iug/Lahjih/kauchai/kauchai.htm>
 15. 客語 e 學堂：http://www.hakka.tcg.gov.tw/e_learning/e_learning.asp
 16. 教育部閩南語字彙網站：
http://www.edu.tw/EDU_WEB/EDU_MGT/MANDR/EDU6300001/allbook/taihtm/f13.html?open
 17. 海陸注音檢定詞彙：<http://home.kimo.com.tw/hakkangin/frame.htm>
 18. 語料庫建立蒐集計畫：<http://iug.csie.dahan.edu.tw/TG/guliaukhou/>
 19. 台文／華文線頂辭典：<http://203.64.42.21/iug/Ungian/soannteng/chil/Taihoa.asp>
 20. 台語線頂字典：<http://203.64.42.21/TG/jitian/tgjt.asp>
 21. 台大客家社客語有聲字典：<http://home.ee.ntu.edu.tw/~r8921044/hdict/hdict.htm>

22. 台北市客委會現代客語詞彙彙編：
<http://www.hakka-lib.taipei.gov.tw/vocabulary/>
23. 國客語翻譯系統：<http://club.ntu.edu.tw/~hakka/>
24. 臺灣本土語言互譯及語音合成系統：
<http://nlg.csie.ntu.edu.tw/systems/TWLLMT/>
25. 華台轉換系統：<http://taigi.fhl.net/ht/index1.html>
26. 台語文e網站：<http://iug.csie.dahan.edu.tw/taigu.asp>
27. 客家人母語教學百花園：<http://home.kimo.com.tw/hakkangin/frame.htm>
28. 閩南語俗曲唱本「歌仔冊」全文資料庫：
<http://www32.ocn.ne.jp/~sunliong/kua-a-chheh.htm>