

台湾学术网路的闽客语资源现况

首届海峡两岸现代汉语问题学术研讨会

主办单位：天津南开大学

2005年11月5-6日

罗凤珠·台湾元智大学中国语文学系

台湾学术网路所建置的语文资源以国语文占大宗，闽客语资源的建置起步较晚，数量也比较少，但正因为起步较晚，建置闽客语资源网站时少了很多摸索的时间。

台湾近年所执行的数位典藏国家型科技计划，迄今已迈入第五年，所建置的闽客语数位资源，内容方面包括小说、散文、戏曲剧本、歌谣、谚语；语音方面包含音标系统、语音资料；语料方面包括小说与戏剧语料、生活用语语料；网路教学方面，包括民歌谚语、台湾文学、生活会话、语音；字典辞典工具书方面包含闽客语字典；语言调查之空间资讯方面有新竹新丰乡闽客语分布地理资讯系统。在网站功能方面，包含典藏、研究、教学与推广四种。

本文将介绍台湾学术网路所建置的闽客语资源以及建置的理念与方法。

全文

台湾学术网路所建置的语文资源以国语文为大宗，闽客语资源的建置，起步比较晚，资源比较少。闽客语资源的建置者，或来自政府单位国家计划的推动，或由闽客语研究者、教学者、民间个人、民间团体自发性建置。正因为起步比较晚，所以建置的过程省去很多摸索的时间，但也因为投入的人力、经费等资源比较少，成果远不及国语文资源。

闽客语建置的内容包含小说、散文、戏剧、歌谣、谚语等文本、语音、语料、研究著作、教学网站、语言调查地理资讯、字典辞典工具书等资料；网站的功能包含典藏、研究、教学、推广等方面。今举其要者介绍如后。

一、推动及建置者

- (一) 数位典藏国家型科技计划：因感于数位典藏的重要，台湾政府行政院于2000年7月通过「国家典藏数位化计划」，之后由国家科学发展委员会整编为国家型计划，于2002年1月1日正式展开，并将计划名称订为「数位典藏国家型科技计划」，在这个国家型计划之下，建置了「荔镜姻，河洛源—闽南语第一名著《荔镜记》数位博物馆」、「闽南语典藏—历史语言与分布变迁资料库」、「不惜歌者苦，但伤知音稀：诗词曲文三语吟唱读教学网站」等三个网站。
- (二) 行政院侨务委员会：由侨委会建置的「全球华文网路教育中心」提供台湾闽南语及客语教材。
- (三) 行政院客家委员会：建置「哈客语言馆」、「(客家)文史天地馆」，内有「哈客网路学院」，提供客语网路教学教材。
- (四) 行政院文化建设委员会：建置国家文化资料库，内有「钟理和数位博物馆」、「赖和数位博物馆」
- (五) 闽客语研究者：成功大学台湾文学研究所吕兴昌教授主持的「台湾文学研究工作室」、大汉技术学院资讯工程学系杨允言教授主持的「台语文e网站」、国立交通大学外文所暨语言学研究所郑良伟教授主持的「台湾语文资讯网站」
- (六) 民间个人及团体：台湾大学客家研究社以及吕兴昌、杨允言、王顺隆三位教授都曾投入多年心力。

二、网站内容

- (一) 文本网站：以文本内容为网站主要建置内容，文本的类别包含：
 1. 语体文：吴浊流、龙瑛宗、钟理和、赖和、钟肇政、李乔、钟铁民等作家之全部或部分文学作品。
 2. 戏曲剧本：荔镜记（万历、光绪、顺治、嘉靖本）、同窗琴书记、苏六娘、金花女、陈三五娘。
 3. 歌谣：闽南语歌仔册、客家歌本。
 4. 谚语：台湾谚语
- (二) 教学网站：
 1. 行政院侨务委员会「全球华文网路教育中心」：
「全球华文网路教育中心」的网路教学资源，台语教学包含台语开讲、俚谚俗语、乡土歌谣等文字及语音资料。客语教学包含四线与海陆腔客语日常用语教材之文字及语音、词语解释。

2. 行政院客家委员会「哈客网路学院」：
网路教学资源包括「哈客语言馆」及台湾饶平、大埔、诏安客语教学资源中心，多数教材只提供声音资料，未提供文字资料。
3. 九年一贯乡土语言台语（闽南语）教材资源中心：
这是由杨允言教授整理的闽南语教材入口网站，提供教材相关资讯。
4. 九年一贯母语（闽南语）教材：
这是由萧平治老师编写建置的教学网站，内容分为：幼稚园、国小、台湾囡仔歌、囡仔故事、俗语等单元，教材内容包含文字以及生难词语的音标、注解。
5. 客语 e 学堂：
这是由台北市政府客家事务委员会所建置的网站，「客语 e 学堂」是网站内容的一部份，教学内容以客语会话为主，分为：客语 e 学堂（初级、中级课程）、日常问候语（饮食、生活、称赞三类）、客语师傅话，以上三类教材均提供文字及声音资料，声音资料含国语及海陆、饶平、四县、诏安四种腔调。
6. 不惜歌者苦，但伤知音稀：诗词曲文三语吟唱读教学网站：
这个网站是笔者受教育部及国科会委托所建置，主要目的是以历代诗词曲文为素材建立闽客语读书音吟读教学，教材内容包含文字及声音资料，声音资料提供闽客语的吟读唱，书面教材提供洪泽南老师撰写的闽南语传统八音，包括：说吟调、说平仄四声八音、说句法、说文音、说勾破、说韵目、说转调、说四声递用、说双声迭韵等，是学习读书音不可多得的教学资源。

（三）语音、文字及语料网站：

1. 字辞典语料：
「台文 / 华文线顶辞典」收有 62,035 笔台语词条，台北市客委会「现代客语词汇汇编」收有 22,600 余笔客语词条，台湾大学客家研究社建置的「在线客语有声字典」共收中文字 9,837 个，四县腔 11,784 个音，大埔腔 1,119 个音，「客语小词典」共收大埔腔 1,333 笔、四县腔 1,098 笔、南部四县腔 33 笔。
2. 教育部闽南语字汇网站：
教育部委托台湾大学中文系杨秀芳教授搜集整理，严谨考证闽南语本字，建置「闽南语字汇」网站，提供部首、音序、词例索引，词例分为天文地理、人体、人事活动、物类、其它等五类，依类别列出国语用词、通俗写法、闽南语用词、词例出处对照表，每一笔词例列出衍生之词汇、读音、词义。
3. 海陆注音检定词汇：
收有 2,148 笔客语海陆注音检定词汇，每一笔词条都加上国语词汇字义。

4. 闽南语读书音资料库：

闽南语及客语都可以分为读书音（文音、孔子白）及说话音，多数人能说说话音，但不熟悉读书音。笔者所建置的「荔镜姻，河洛源：闽南语第一名著《荔镜记》多媒体教学网站」，所收 131 段歌仔戏陈三五娘剧本歌词，以纯正闽南语读书音录制语音资料。笔者所建置的「不惜歌者苦，但伤知音稀：诗词曲文三语吟唱读教学网站」，以闽南语及客语之读书音录制之诗词曲文吟读资料，计有闽南语诗词 605 篇，客语诗词 60 篇。

5. 语料库建立搜集计划：

这是由大汉技术学院杨允言教授所主持的计划，广泛搜集台语语料，应用于台语文 Concordance、台语文词频统计、编写辞典(Lexicography)，并公开提供台语文研究者使用。

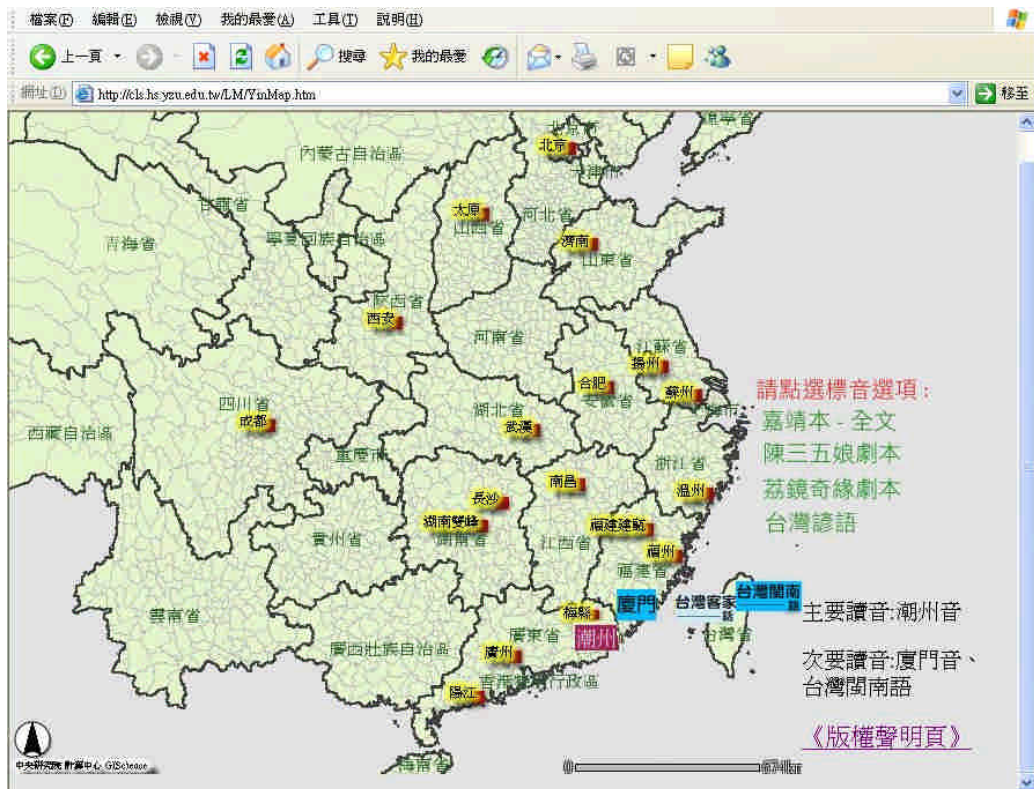
6. 文学作品语料：

由笔者与学界学者共同主持的研究计划，建立了一些闽客语文学文本，并且从这些文学文本里抽取闽客语语料，客语语料包含：〈渡台悲歌〉、钟理和文学作品客语语料（钟怡彦整理）；闽南语语料包含：《荔镜记》全文词汇种数 17,589 笔，其中闽南语词汇种数 4,541 笔。

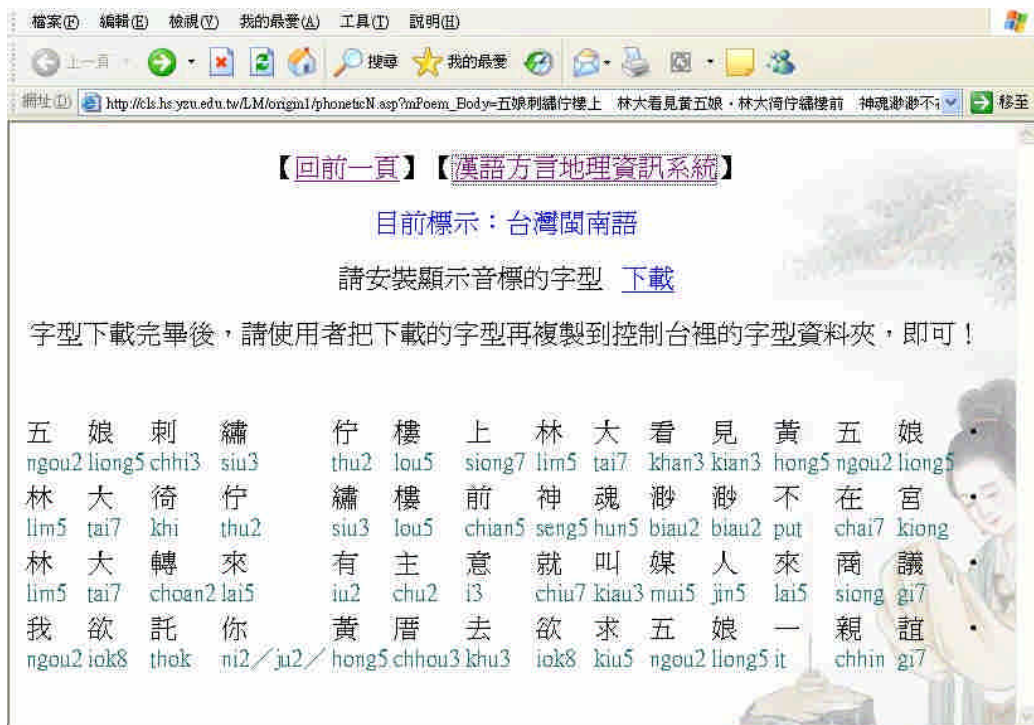
7. 读音自动标注系统：

笔者引用郑锦全院士「方言读音采集计划」所采集的厦门、潮州等闽南语读音及笔者研究计划所建置的客语读音，建立读音自动标注系统，提供在任何文本上自动标注闽客语读音音标的功能，以「陈三五娘歌仔戏剧本」为例，读音自动标注平台如图一及读音标注结果如图二：

图一：读音自动标注平台介面图：



图二：读音标注结果（陈三五娘剧本）



(四) 台、客字辞典网站：

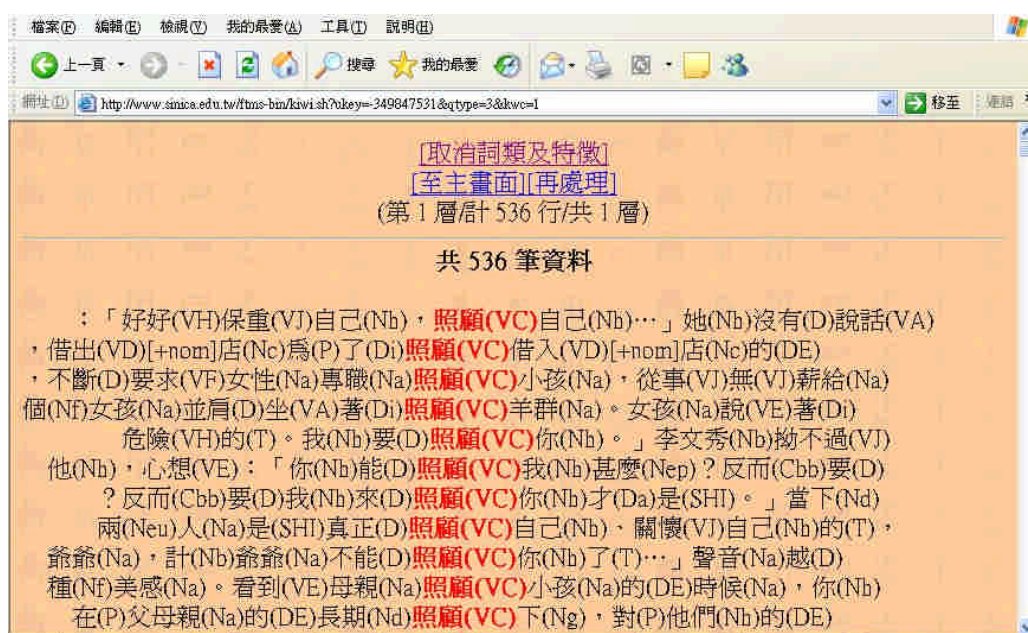
1. 台文 / 华文线顶辞典：

总计搜集 62,035 笔词条，其中使用者增加的词条有 195 笔，提供的查询栏位包括台文（汉罗、罗马字）、华文、罗马字（无声调）三种；提供的查询方式包含：完全比对、部分比对、词头、词尾等四种。查询结果显示台语罗马字、汉罗、华文，所查询的华文词汇可以进一步连结中央研究院平衡语料库显示该词汇之出处，并标示其词类。以华文的「照顾」为例，选择「完全比对」，查询结果如下：

	台语罗马字	汉罗	华文
1	chhiân	晷	照顾 / 养育
2	chiàu-kò	照顾	眷顾 / 照拂 / 照应 / 照顾
3	kám-kài	感介	照顾 / 关照
4	khòa ⁿ -kò	看顾	看管 / 看护 / 看顾 / 照料 / 照顾 / 护理
5	tá ⁿ -chah	打 chah	支援 / 打点 / 协助 / 照顾 / 帮助

任何一笔华文词汇，都可以连结中央研究院平衡语料库，显示文本出处及词类标记，以「照顾」为例，显示结果如图三：

图三：中央研究院平衡语料库显示查询结果



2. 台语线顶字典：总计收 22,080 个字，其中 2,726 个字没有台语音节。主要的用途是查某一个汉字的台语读音，并可扩充查询该汉字的衍生词，其余用法、功能都与《台文 / 华文线顶辞典》相同。
3. 台大客家社客语有声字典：共收中文字 9,837 个，四县腔 11,784 个音，大埔腔 1,119 个音。提供「中

文单字查其读音或读音查其文字」的查询功能，查询结果提供语音之声音资料。

4. 台北市客委会现代客语词汇汇编：

本词汇汇编收录二万二千六百余词条，编写方式为四县词汇、标音、海陆词汇、标音、国语词汇的对应，提供客语声韵表、客家造字表以及客语互动搜寻等功能。

(五) 自动翻译系统

1. 国客语翻译系统：

输入一句国语的读书音或说话音，可以翻译为教会罗马音标或台湾客语音标的客语。

2. 台湾本土语言互译及语音合成系统：

这是台湾大学资讯工程研究所陈信希教授主持的「自然语言处理实验室」所建置的网站，以国、台、客语输入句子，即可自动发出国语读音、台语的漳州或泉州音的读音，客语的说话音或读书音之音标。

3. 华台转换系统：

由台语信望爱网站所开发的华台转换系统，可以汇入整段文字或整篇文章，系统自动转换为台文。

(六) 语言调查地理资讯网站：

由中央研究院郑锦全院士主持的「闽南语典藏：历史语言与分布变迁资料库」计划，全面调查郑院士故乡新竹新丰乡闽客语之分布与变迁，以新丰乡作为调查的目标，主要的考量，是当地是闽客语及外省籍族群杂居的乡镇，随着经济环境的变迁、教育程度的普及等等因素，人口结构产生变化，连带使生活语言互有消长。该计划收录闽客语语音，并根据调查结果建立语言分布之地理资讯系统，以作为未来全面调查语言分布之基础，这是目前在台湾将语言调查与地理资讯结合的最大型计划。目前已完成上坑、凤坑、埔和、中仑、松林、员山、福兴、坡头、后湖、新丰、青埔、瑞兴、重兴等十三个村落 **13,218** 户的调查，使用客语计有四县 **225** 户，海陆 **5,801** 户，总计 **6,026** 户，占全乡的百分之五十。调查结果绘制语言分布地理资讯系统如图四（以不同颜色的圆点区别不同语言）：

图四：语言分布地理资讯网页



(七) 综合及全文资料库型网站：

1. 台语文 e 网站：

这是由杨允言教授整理维护的台语文入口网站，有自建的资源，有连外的资源。网站内容分为字典、语料库、生活、个人网站、blog 网志、教学、社团、台语运动、台语输入软件、研究、刊物等。

2. 台湾语文资讯网站：

这是由台湾交通大学外文所暨语言学研究所郑良伟教授建置的网站，内容包括台湾语言研究、台湾语言资料库、台语教育资料库等单元，有自建的内容，也有连外的内容。

3. 台湾文学研究工作室网站：

这是由成功大学台湾文学研究所吕兴昌教授自力自费建置的网站，网站内容包含：台语软件、母语文学、民间文学、古典文学、原住民文学、郑辖时期、清领时期、日治时期、国统时期、台湾文学研究学者、学位论文、研究书目、研究刊物、学术会议、文学通论、出土资料。

4. 客家人母语教学百花园：

这是一个综合性的客家资源入口网站，从网站无法看到网站主持人。搜集的内容非常丰富，分为：教材园地、注音课程、词汇对照、字音研究、注音谚语、昔时贤文、三字经、注音四句、渡台悲歌、创作新诗、客家童谣、创作歌谣、网路客家、客语典藏等，是一个非常详备完整的客家资源入口网站。

5. 荔镜姻，河洛源：闽南语第一名著《荔镜记》多媒体教学网站：

这是由笔者所主持的数位典藏国家型科技计划所建置的网站，以现存最早以闽南语撰写的剧本《荔镜记》为建置的内容，提供教学、研究、典藏的功能，内容包括原始文献、改写剧本、语料、语音、研究资料等。

6. 闽南语典藏：历史语言与分布变迁资料库

这是由郑锦全院士所主持的数位典藏国家型科技计划所建置的网站，延续「荔镜姻，河洛源：闽南语第一名著《荔镜记》多媒体教学网站」扩充建置，文本内容从嘉靖本《荔镜记》扩充到万历、顺治、光绪本，增加了同窗琴书记、苏六娘、金花女、闽南语歌仔册、客家歌本以及语言（闽客语）分布调查资料库

7. 闽南语俗曲唱本「歌仔册」全文资料库：

这是由王顺隆教授建置，收有台湾及大陆地区出版的歌仔册闽南语俗曲唱本六百余种，资料库需经申请方可使用。

三、建置的理念与方法

文前已提及闽客语资源的网站建置，起步晚，数量少，台湾推行闽客语教学，所遭遇的最大困难是教学素材不足，多数老师只能以闽客语的歌谣及俗谚语当作教材，加上音标体系的混乱、用字无统一标准，更增添了教学及推广的困难。广泛阅读文学作品是培养语文能力的重要方法，学习中英文是如此，学习闽客语也是如此，但是以闽客语书写的文学作品比较少，建置在网路上的资源更少。以下以《荔镜记》网站为例，说明建置的理念与方法。

笔者与郑锦全院士、黄居仁、杨秀芳、曾淑娟三位教授选择《荔镜记》作为建置闽南语教学网站的素材，主要的考量是戏曲剧本保留了大量的口语语料，是建立方言教学最好的素材。《荔镜记》是现存文献之中最早以闽南语（潮州话、泉州话）写成的剧本，成书的年代在十六世纪。除了潮剧之外，《荔镜记》还被改编成高甲戏、黄梅戏、歌仔戏、梨园戏、南管等地方戏曲，不同时代的改编剧本，又加入了该时代、该语种的口语，因此保留了十六世纪以来闽南地区丰富的口语资料。

好的教学需要有好的研究作基础，《荔镜记》网站的建立兼顾典藏、研究、教学、传播的功能，主要的架构包含：原剧文献、表演艺术、儿童剧场、研究资料、语文教学区、使用者回馈区、系统简介等七部分。原剧文献分原剧书影和嘉靖本全文，并提供检索查询、标注字词意义、标注闽南语 / 潮州话 / 厦门话的读音等功能。表演艺术分剧本资料和演出资料，搜集了歌仔戏《陈三五娘》和南管汉唐乐府《荔镜奇缘》的演出资料。剧本资料也能标注上述各种读音，而且也请专人录制闽南语读书音。儿童剧场

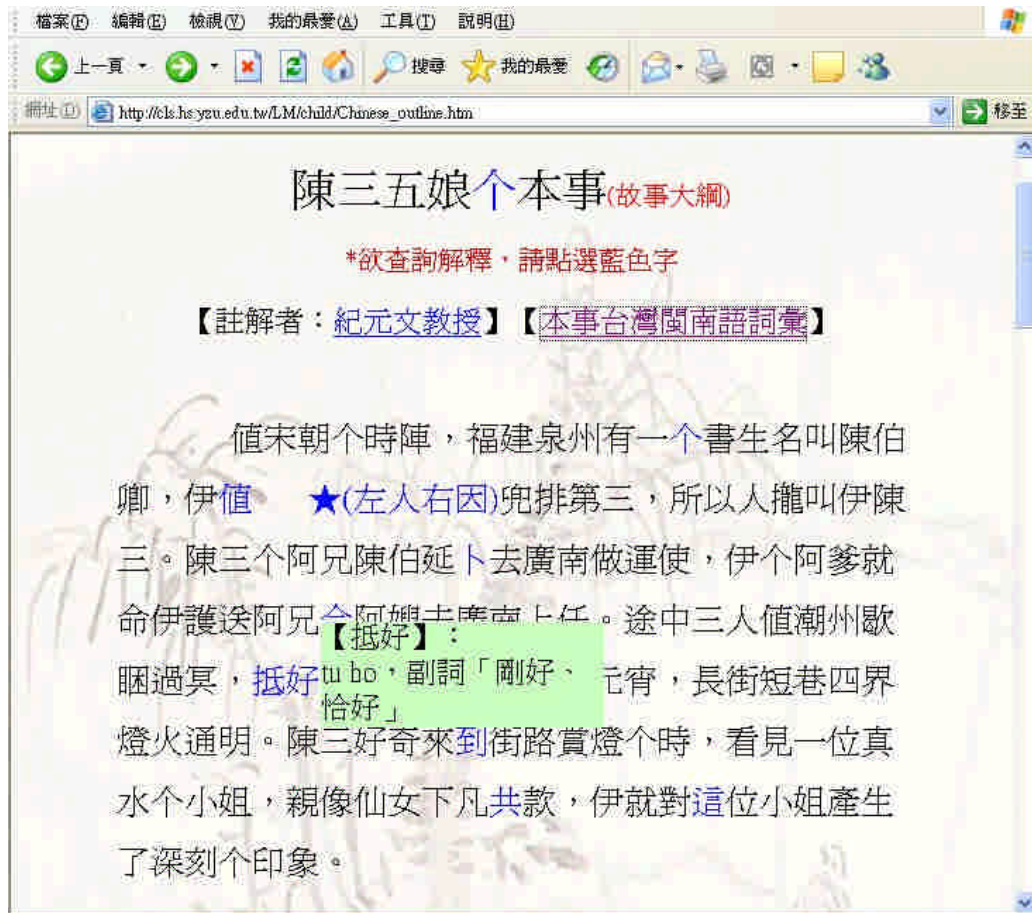
是陈三五娘的儿童故事改编本，计有国语、闽南语、英文、西班牙文等四种语文版。研究资料中包含相关研究论着目录资料、相关研究论着全文资料、相关研究网站、乡土语文教学参考工具书等，方便读者从事进一步的研究。语文教学区包含国语 / 闽南语同义词的对照、陈三五娘故事的语音教学、语言演进、潮州语言辞典、俗谚语等单元，提供非同步远距多媒体语文教学环境，使用者回馈区提供使用者回馈资料的平台、系统简介等介绍系统相关资讯。

这个网站建置的目的是为了提供闽南语研究、教学及阅读闽南语文本使用，因此从文本整理出闽南语词汇，建立闽南语与国语同近义词对照、国语词义解释、原文例句之索引表，以供查检，如图五。在系统功能的设计上，提供了闽南语词汇随选随译的功能如图六，以降低阅读时因为不懂闽南语词汇语义而产生的阅读障碍。

图五：闽南语与国语同近义词网页



图六：闽南语词汇随选随译功能介面



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，內容難免重迭，縱有不重迭處，也因各自建置，未能建立一些共同性的標準，致使資料無法兼容、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。

閩客語數位化資源的建置，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。

由國家型計劃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位化典藏標準，優先建立字辭典等工具性的資料庫，文

字输入等工具软体，结合语言、文学、资讯科技、图书资讯四个领域的专业人才，结合政府与民间，学界与业界的力量，订定轻重缓急的顺序，以国家资源有计划的建立闽客语资源，才是长久之计。

附注：

- 一、本文完成于 2006 年 4 月，所有网站资料以这个时间所见为准。
- 二、本文述及台湾话时，以「闽南语」或以「台语」叙述，完全依照所介绍网站所使用的名称。
- 三、本文所述及之各网站名称及网址，依行文顺序分列如下：
 1. 数位典藏国家型科技计划：<http://www.ndap.org.tw/>
 2. 荔镜姻，河洛源—闽南语第一名著《荔镜记》数位博物馆：
http://cls.hs.yzu.edu.tw/LM/Lm_home.asp
 3. 闽南语典藏—历史语言与分布变迁资料库：<http://southernmin.sinica.edu.tw/>
 4. 不惜歌者苦，但伤知音稀：诗词曲文三语吟唱读教学网站：
http://cls.hs.yzu.edu.tw/CL_POEM/
 5. 行政院侨务委员会全球华文网路教育中心：<http://edu.ocac.gov.tw/>
 6. 行政院客家委员会哈客网路学院：<http://www.hakka.gov.tw/>
 7. 钟理和数位博物馆：<http://cls.hs.yzu.edu.tw/ZHONGLIHE/>
 8. 赖和数位博物馆：<http://cls.hs.yzu.edu.tw/laihe/>
 9. 台湾文学研究工作室：<http://www.ncku.edu.tw/~taiwan/taioan/>
 10. 大汉技术学院台语文 e 网站：<http://203.64.42.21/taigu.asp>
 11. 交通大学台湾语文资讯网站：<http://taigu.eic.nctu.edu.tw/>
 12. 台湾大学客家研究社：<http://club.ntu.edu.tw/~hakka/>
 13. 九年一贯乡土语言台语（闽南语）教材资源中心：
<http://203.64.42.21/iug/ungian/POJ/poj.asp>
 14. 九年一贯母语（闽南语）教材：<http://203.64.42.21/iug/Lahjih/kauchai/kauchai.htm>
 15. 客语 e 学堂：http://www.hakka.tcg.gov.tw/e_learning/e_learning.asp
 16. 教育部闽南语词汇网站：
http://www.edu.tw/EDU_WEB/EDU_MGT/MANDR/EDU6300001/allbook/taihtm/f13.html?open
 17. 海陆注音检定词汇：<http://home.kimo.com.tw/hakkangin/frame.htm>
 18. 语料库建立搜集计划：<http://iug.csie.dahan.edu.tw/TG/guliaukhou/>
 19. 台文 / 华文线顶辞典：<http://203.64.42.21/iug/Ungian/soannteng/chil/Taihoa.asp>
 20. 台语线顶字典：<http://203.64.42.21/TG/jitian/tgjt.asp>
 21. 台湾大学客家社客语有声字典：
<http://home.ee.ntu.edu.tw/~r8921044/hdict/hdict.htm>
 22. 台北市客委会现代客语词汇汇编：

- <http://www.hakka-lib.taipei.gov.tw/vocabulary/>
23. 国客语翻译系统: <http://club.ntu.edu.tw/~hakka/>
24. 台湾本土语言互译及语音合成系统:
<http://nlg.csie.ntu.edu.tw/systems/TWLLMT/>
25. 华台转换系统: <http://taigi.fhl.net/ht/index1.html>
26. 台语文ê网站: <http://iug.csie.dahan.edu.tw/taigu.asp>
27. 客家人母语教学百花园: <http://home.kimo.com.tw/hakkangin/frame.htm>
28. 闽南语俗曲唱本「歌仔册」全文资料库:
<http://www32.ocn.ne.jp/~sunliong/kua-a-chheh.htm>