

教育資料與研究雙月刊
第 91 期 2009 年 12 月 131-160 頁

客語文獻分析與數位典藏——以客英、 客法大辭典為例

江敏華* 黃彥菁** 宋柏賢***

摘要

本文以《客英大辭典》與《客法大辭典》這兩部早期客語文獻為例，說明客語文獻分析與數位典藏化的過程，這些文獻是由傳教士到客家地區進行傳教工作時，根據當地的客語方言，採錄漢字及詞條，經過修訂後出版的。本文簡介數位典藏計畫與辭典的歷史背景，並針對辭典的音韻系統及字音、用字等體例作出說明，辭典所收錄的漢字及詞條內容豐富，除具有客家文化及教化意義的成語、諺語外，《客法大辭典》還收錄大量的教會用語詞條。客語文獻數位典藏的資料建構方式說明，主要敘述《客英大辭典》、《客法大辭典》轉成數位化全文資料時所作的特殊處理，修正原件有誤處，並保留原件掃描檔以便對照，最後再將建構完成的資料庫置於閩客語典藏網站中供讀者線上檢索。由於數位典藏計畫目前尚在進行中，仍須持續修改以提供更完整的客語數位文獻資料。

關鍵詞：客語、客英大辭典、客法大辭典、資料建構、數位典藏

*江敏華，中研院語言所副研究員兼副所長

**黃彥菁，中研院語言所研究助理

***宋柏賢，中研院語言所研究助理

電子郵件：mhchiang@gate.sinica.edu.tw；htmyc@gate.sinica.edu.tw；apostle@gate.sinica.edu.tw

來稿日期：2009 年 9 月 14 日；修訂日期：2009 年 12 月 11 日；採用日期：2009 年 12 月 16 日

Literature Analysis and Digital Archives of Hakka Language: A Case Study of A Chinese-English Dictionary Hakka-Dialect and Dictionnaire Chinois-Francais Dialecte Hac-ka

Chiang Min-hua* Huang Yen-Ching** Sung Po-Hsien***

Abstract

Based on two early Hakka texts——A Chinese-English Dictionary Hakka-Dialect (CEDH) and Dictionnaire Chinois-Francais Dialecte Hac-ka (DCFH), this paper indicates the analysis and digitalization process of Hakka text archives. The above texts were compiled and published by the missionaries who collected related corpus while preaching in Hakka areas. The paper introduces "National Digital Archives Program" and the history of CEDH and DCFH, focusing on the style of the two dictionaries, such as phonological system, phonetics and wording. There are abundant Chinese characters, vocabulary entries, as well as phrases and proverbs in CEDH and DCFH, and the latter gathers ample entries of church. The explanation on constructing the data is mainly about the special handling of digitalizing the whole CEDH and DCFH texts, and about the error correction while scanning the original texts in order to compare. When databases are done, we put them on Min and Hakka Language Archives website to provide users with online search services. National Digital Archives Program is still on going and it needs modification to render more digital archives of Hakka texts.

Keywords: Hakka language, a Chinese-English Dictionary Hakka-Dialect (CEDH), Dictionnaire Chinois-Francais Dialecte Hac-ka (DCFH), data construction, digital archives

*Chiang, Min-hua, Associate Research Fellow & Deputy Director, Institute of Linguistics, Academia Sinica

**Huang, Yen-Ching, Research Assistants, Institute of Linguistics, Academia Sinica

***Sung, Po-Hsien, Research Assistants, Institute of Linguistics, Academia Sinica

E-mail: mhchiang@gate.sinica.edu.tw; htmyc@gate.sinica.edu.tw; apostle@gate.sinica.edu.tw

Manuscript received February 14, 2009; Modified: December 11, 2009; Accepted: December 16, 2009

壹、前言

自社會走入資訊時代後，數位資訊、通訊科技以及網路的發展對社會的影響日益深遠。以往，資訊科技在生活各方面的影響是局部的，網際網路普及以後，資訊與數位化對社會的影響已是全面的，短時間內改變了社會上的人際關係、食衣住行、工作與學習研究的態度及方法。亞太經濟合作（Asia-Pacific Economic Cooperation，APEC）認為知識經濟是未來的發展趨勢，有效的利用經過思考或實踐所獲得的智慧來創造價值的活動就是知識經濟。典藏和教育學習的數位化是知識經濟的基礎環節之一，可有效提升知識的保存、傳承及應用。為了讓現代人們能夠更方便接觸資訊，以及更快、更有效率的學習，數位典藏與數位學習正是此等時代下所催生的產物。

數位典藏國家型科技計畫始於2002年，參與的機構（依筆劃序）有中央研究院、國立自然科學博物館、國立故宮博物院、國立台灣大學、國立歷史博物館、國史館、國史館台灣文獻館、國家圖書館、及台灣省諮詢會等九個典藏機構。閩客語典藏計畫為語言典藏計畫的子計畫（以下簡稱本計畫），隸屬數位典藏中央研究院機構計畫。數位典藏內容所涉及的層面很廣，其一主要方式為透過整理與建構早期文獻資料，再藉由數位典藏技術與網路程式設計，最後放上網站供研究交流與資源共享之線上檢索。本計畫彙集與整理一系列客語文獻資料的，目前已收錄《客英大辭典》（以下簡稱《客英》）及《客法大辭典》（以下簡稱《客法》）中所有的漢字及詞條數位化後，典藏成寶貴的語言文化資產，分享給社會大眾做學習及學術研究工具，未來並持續典藏其他客語文獻，擴充客語資料庫以臻客語文獻之研究環境於至善，並促進知識發展。此二部辭典歷經近百年，與現代通行的客語或有差異，但其保存了當代客語的風貌，為人類的智慧留下記錄，甚具歷史文化價值；在研究方面，學者可藉其比較古今客語的異同、改變的規則或差異的原因，以此文獻做為客語數位典藏之開端有其學術價值。

數位典藏可以更方便資訊的使用，其優點如下：一、減少蒐尋大量資料的時間，因為只要在網路上輸入相對應的關鍵字便能找到資料；二、在連行動電話都能上網的當下，資料庫建置後何時何地都可以使用，不用再攜帶笨重的文獻；三、文獻中可能的錯誤，會做初步的註記或修正，減少使用者自己摸索查證的時間；四、免費公開，使用者不必再花錢購買或到圖書館借閱，取得資訊不但更容易，還增加文獻的曝光度；五、提供新的典藏媒介，讓知識的累積與

保存多一分保障。

本文以《客英》及《客法》二大辭典為例，說明客語文獻數位典藏的實質工作內容，《客英》原名「A Chinese-English dictionary: Hakka-Dialect as spoken in Kwang-tung province」。1847—1854 年間（Basel）傳教士韓山明（Rev. Th. Hamberg, 1819-1854）與黎力基（Rev. R. Lechler, 1824-1908）等人就已經到達大陸進行傳教，本辭典是根據巴色會的手稿所編著，在序文中所介紹巴色會傳教士的手寫本，內容為客語通行區域西南邊的方言，經過修訂、縮寫後的辭典屬於東北邊區域的梅縣、興寧、平遠、蕉嶺、五華、大埔一帶的方言，初版於 1905 年出版，由麥愛華（D. MacIver, 1852-1910）編撰，1926 年，經麥肯錫（M. C. Mackenzie）重新修訂校正後發行二版，本計畫所使用的版本為二版。全書辭典部分共計 1,142 頁。

《客法》則由法國籍汕頭傳教士雷卻利（Charles Rey）牧師於 1901 年在廣東所出版，歷經多次修訂，於 1926 年有了現在通行的版本。該辭典的內容不但收字豐富，並收錄有許多教會用語、諺語及成語詞條，對早期客語研究來說是相當珍貴的資料。前言附有語音、聲調、語法及構詞介紹，附錄為索引及中國特有的時制、度量衡、親屬稱謂等圖表，亦是特色之一。辭典部分共 1,412 頁。

該計畫除典藏這兩本辭典的影像外，亦將其輸入為全文電子檔以建構客語資料庫，未來並將提供線上檢索系統。數位典藏只是數位發展的開端，如何推廣及應用這些計畫的成果，將是我們需要更積極努力與成長的目標。藉由整理這些早期客語文獻資料的漢字及詞語，能夠更完善地保存這些珍貴的客家方言，透過原件資料的建構並結合網際程式的技術，將賦予數位典藏計畫更深的意義。

貳、客英、客法大辭典的音韻及用字說明

本節首先說明《客英》、《客法》兩大辭典的音韻系統，以及辭典中對於異讀、異（俗）體字、缺字、特殊用語等辭典體例相關問題的處理方式。

一、聲母與韻母

在《客英》及《客法》兩大辭典中皆為 23 個聲母，加上零聲母（無符號表示）。《客英》辭典計有 58 個韻母，其中包含兩個成音節的鼻音 m、ng；而《客法》辭典則有 69 個韻母，亦含兩個成音節鼻音 mg、ng。以下分別說明此

兩本辭典中使用的符號，並以國際音標解釋其所反映的音韻系統。

(一) 聲母系統

1. 《客英》

《客英》計有 24 個聲母（含零聲母）。塞音計有 4 組共 8 個不送氣與送氣相對的塞音聲母以及一個零聲母（不送氣），為雙唇塞音 p、ph，舌尖塞 t、th，舌根塞音 k、kh，圓唇舌根塞音 kw、khw。原文中的送氣音以 h 表示，所以 ph、th、kh 的對應國際音標為 [pʰ]、[tʰ]、[kʰ]；在《客法》中無 kw、khw 這兩個圓唇舌根塞音，k 與 kh 後面的 w，同法文發音的 ou，黃詩惠（2003）指出音值為 u 介音，由於原文已有 u 韻母，故將 w 納入聲母系統中。若後面所接的韻母中已有 u，則單以 o 表示，例字：快「khwai3」，等同於《客法》中的「k'ouai2」、骨「kwut」等於《客法》的「kout」。

鼻音計有 4 個聲母，為雙唇鼻音 m、舌尖鼻音 n、舌根鼻音 ng、舌面前鼻音 ny，其中 ng 依國際音標表示為 [ŋ]，ny 則以 [ɲ] 表示，為 [ŋ] 的有定分音，多與介音及 i 元音結合。

擦音計有 6 個聲母，為唇齒擦音 f、v，舌尖前擦音 s，舌面前擦音 sh、y，喉音擦音 h。其中 sh 的實際音值同於國際音標中的 [ʃ]，林英津（1994）指出 y 同於《客法》中 y 的音值，應用以表示介音 i 或舌面擦音 j，而用作舌面前擦音 j 更接近於語言的事實。

塞擦音計有 4 個聲母，為舌尖前塞擦音 ts、tsh，其中的送氣音 tsh 與國際音標中的 [tsʰ] 相同；舌尖面塞擦音 ch、chh，依實際音值可以國際音標 [tʃ]、[tʃʰ] 表示。

邊音計有一個聲母，為舌尖邊音 l。

2. 《客法》

《客法》計有 24 個聲母（含零聲母）；塞音計有 3 組，共 6 個不送氣與送氣的相對塞音聲母以及一個零聲母（不送氣），為雙唇塞音 p、p'，舌尖塞音 t、t'，舌根塞音 k、k'。

鼻音計有 4 個聲母，為雙唇鼻音 m、舌尖鼻音 n、舌根鼻音 ng、舌面前鼻音 gn；林英津（1994）指出，其中的 ng 依國際音標表示為 [ŋ]，與後元音相接，為一個獨音音位，而 gn 則以 [ɲ] 表示，非獨立音位，可作為 [ŋ] 的有定分音及變體音位處理，只與前、高元音結合，此兩聲母在分配上形成互補狀態。

擦音計有 6 個，為唇齒擦音 f、v，舌尖前擦音 s，舌面前擦音 ch、y，喉音擦音 h，其中的 ch、y 實際音值以國際音標表示為 [ʃ]、[j]，而 y 同於前述《客英》裡的 y，可以舌面擦音 j 表示。

塞擦音計有 6 個聲母，爲舌尖前塞擦音 ts、ts'，舌尖面塞擦音 tch、tch'，以及 tj 和 tz。其中的 tch、tch' 音值相當於國際音標[tʃ]、[tʃ']，亦同於《客英》中的 ch、chh；當 tch 後接 i 介音時，tchi 讀音相當於 tje，而 tj 只和 e 結合，可視爲 tch 的有定分音；tz 後接 e 時，發音相當於《客英》中的 tsy。

邊音計有 1 個聲母，爲舌尖邊音 l。

(二) 韻母系統

1. 《客英》

《客英》計有 58 個韻母，其中包含 2 個成音節鼻韻 m、ng，並將 ng 以國際音際[ŋ]表示。

在開尾韻母（陰聲韻母）中，開口韻母計有 8 個，爲 a、ai、au、e、eu、y、o、oi，其中 y 的實際音值同[i]，但是在原件資料輸入時，以 ii 表示。齊齒韻母有 8 個，爲 i、ia、iau、ie、ieu、io、ioi、iu。合口韻母計有 2 個，爲 u、ui。

鼻音尾韻母（陽聲韻母）以 m 為韻尾的有 5 個，爲 am、iam、em、iem、im；以 n 為韻尾的有 9 個，爲 an、ian、en、ien、in、on、ion、un、iun，其中的 an 與 en、ian 與 ien 可各自合併爲一組韻母；以 ng 為韻尾的有 6 個，爲 ang、iang、ong、iong、ung、iung。

塞音尾韻母（入聲韻母）以 p 為韻尾的有 4 個，爲 ap、iap、ep、ip；以 t 為韻尾的有 8 個，爲 at、iat、et、iet、it、ot、ut、iut；以 k 為韻尾的有 6 個，爲 ak、iak、ok、iok、uk、iuk。

2. 《客法》

《客法》計有 69 個韻母，含 2 個成音節鼻音 mg、ng，mg 只有收「唔」1 個漢字，可以國際音標[m]表示，ng 則同於[ŋ]。

開尾韻母（陰聲韻母）中，開口韻母計有 10 個，爲 a、ai、ao、e、e.、eou、o、oi（oe.）、ou、oui。《客法》在前言的發言規則中說明：

當 e 在字尾時的讀音同法文中「je、me」的發音，若在 e 的右上角加上符號「·」的 e.，讀音同「aimé」，因此當作[e]。（客法大辭典，1988：前言）

所以 e 的實際音值應爲[i]，湯培蘭（1994）指出，在共時音韻的角度來看，oe. 應可視爲 oi 的變體。齊齒韻母有 7 個，爲 i、ia、iai、iao、iou、io、ioi。合口韻母計有 4 個，爲 oua、ouai、oue.、ouo，ou 讀音同法文中的 ou，音值爲[u]，亦等同於 w。

鼻音尾韻母（陽聲韻母）以 m 為韻尾的有 4 個，為 am、iam、em、im；以 n 為韻尾的有 11 個，為 an、ian、ouan、ien、ouen、in、on、ion、ouon、oun、ioun；以 ng 為韻尾的有 10 個，為 ang、iang、ouang、eang、ong、iong、ouong、eong、oung、ioung。林英津（1994）指出，當 e 作韻頭、韻尾，其音韻地位可逕視為 i 的變體，所以 eang、eong 只和邊音聲母結合，相當於介音，例如，領

表 1 聲母對照表

《客英》	《客法》	國際音標	客語通用拼音
p	p	p	b
ph	p'	p ^h	p
m	m	m	m
f	f	f	f
v	v	v	v
t	t	t	d
th	t'	t ^h	t
n	n	n	n
l	l	l	l
ny	gn	jn	
k	k	k	g
kh	k'	k ^h	k
ng	ng	ŋ	ng
h	h	h	h
ts	ts	ts	z
tsh	ts'	ts ^h	c
s	s	s	s
ch	tch	tʃ	zh
chh	tch'	tʃ ^h	ch
sh	ch	ʃ	sh
y	y	j	
kw			
khw			
	tz		
	tj		
零聲母	零聲母	ø	ø

leang7、良 leong5。

塞音尾韻母（入聲韻母）以[p]為韻尾的有4個，為 ap、iap、ep、ip；以 t 為韻尾的有9個，為 at、ouat、et、iet、ouet、it、ot、out、iout；以 k 為韻尾的有7個，為 ac、iac、oc、ioc、ouoc、ouc、iouc，其中字尾 c 的音值同於國際音標中的[k]。

聲母依《客英》、《客法》、國際音標、通用拼音的記音方式，整理如表1所示。

二、聲調、異讀

（一）聲調

《客英》依原文的前言一共有6個聲調，所舉的例子如表2所示。

上平與下入兩調無符號標記，依照原件數位典藏的資料建構規則，輸入方式為：夫 fu、乎 fu5、府 fu2、富 fu3、服 fuk8、福 fuk4，共6組。而《客法》則同於《客英》一樣有6個聲調，依原文所舉的例子如表3所示。

表2 《客英》聲調表

調類	例字	輸入標記
上平（陰平）	夫 Fu	無
下平（陽平）	乎 Fū	5
上聲	府 Fú	2
去聲	富 Fù	3
上入（陽入）	服 Fúk	8
下入（陰入）	福 Fuk	4

資料來源：客英大辭典（2007）。

表3 《客法》聲調表

調類	例字	輸入標記
上平（陰平）	夫 Foū	7
下平（陽平）	湖 Foū	5
上聲	虎 Foù	3
去聲	父 Foú	2
上入（陽入）	復 Foúc	8
下入（陰入）	福 Foúc	4

資料來源：客法大辭典（1988：XII）。

為統一做數位典藏的建構方式，採用同符號同標記數字，因此《客英》的上聲記為「2」、去聲記為「3」，而《客法》的上聲記為「3」、去聲記為「2」，二者正好相反，此 6 個聲調依序標記為：夫 fou7、湖 fou5、虎 fou3、父 fou2、復 fouc8、福 fouc4。

（二）異讀

在「一字多音」的異讀現象中，一般有兩種情況，一是「四聲（聲調）別義」，為同一漢字而音義卻相異；二為「文白異讀」，「文」指在文言文、詩歌及成語中的讀音，而「白」即日常用字的口語發音，有些漢字文、白兩種語體使用不同的讀音，故有文白異讀之分。

《客英》與《客法》漢字記音中蒐集很多豐富的異讀音。以下將依原件資料的異讀音，以原件資料舉例說明其異讀音：

「難」的讀音為「Nan5」（如圖 1），指「很難、困難」之意，詞條列舉「難做」，說明某件事情很難做到。「難」的讀音若為「Nan2（Nan3）」（如圖 2），屬四聲別義的讀音，則是「受難」之意，詞條列舉「難民」，指受難的人民，由於《客英》與《客法》的聲調標記不同的關係，事實上「Nan2（Nan3）」為同一聲調，皆為入聲字，與前例所舉的「難」字為異音異義的異讀音。另舉原文資料的文白異讀例子如下：

難^{172,11} **难**^{172,2} **難**^{172,11} Nān.
(V. Nán.)
Difficile, ardu, pénible, fatigant, ennuyeux.
| 做 nán tsó difficile à faire.

圖 1 異讀音舉例之一
資料來源：客法大辭典（1988：614）。

難 Nān. Difficulty. Distress. See under Nān.
| 民, n. mīn, people in distress from ravages of war.

圖 2 異讀音舉例之二
資料來源：客英大辭典（2007：501）。

「放」的讀音為「Fong2（Fong3）」（如圖3）時，視為文讀音，而這些文讀音大多用於文言之中。「放」的讀音若為「Pióng2（Pióng3）」（如圖4）時，指白讀音，則用於口語中，所以除四聲別義外，異讀還分有文白異讀兩種讀音。

放 66,4 *Fóng.* (V. *Pióng*). Placer, déposer.—Lâcher, laisser aller, donner liberté.— Abandonner, rejeter. Reléguer dans un lieu déterminé.

担子 | 下來 *tam tzè fóng hā loî* dépose ton fardeau.

圖3 異讀音舉例之三

資料來源：客法大辭典（1988：117）。

放 66,4 *Pióng.* (V. *Fóng*.) Placer, déposer.—Laisser aller, laisser faire.

| 在 桌 面 *pióng tsái tsôc mién* mettre sur la table.

圖4 異讀音舉例之四

資料來源：客法大辭典（1988：756）。

三、漢字、特殊用語

(一) 多字一音

在原件資料中的漢字除一字多音（異讀）外，也出現「多字一音」的異（俗）體字現象，大多為二字或三字一音。由於異（俗）體字部分出現較多罕見字，並無收錄在一般電腦的中文輸入字中，故在輸入資料時，採用「漢字構形資料庫」來造字來解決電腦缺字的問題：

中央研究院資訊所文獻處理實驗室自1993年起，即先由字形著手，建置漢字構形資料庫。漢字構形資料庫早期收錄的字形是以楷書的現代印刷字體為主，其後陸續增加小篆、金文、甲骨文、楚系文字等古漢字。收錄這些不同歷史時期的漢字，除了要作字形、字義銜接外，還要依據不同的形體來作構形分析。字形的銜接是依據字形的演變，在電腦中使用相同的編碼位置，編入不同的字型。字義的銜接是參考字義的隸屬，在電腦中使用不同的編碼位置，編入異體字表。（莊德明、謝清俊，2005：摘要）

原件資料舉例如圖 5、圖 6。以上各舉二字及三字異（俗）體字為例，但若遇到「同字異義」的情況發生，在輸入時拆成兩組同字異義，若漢字本身無釋義，則與其漢字後面所舉的詞條結合成一組。有關漢字原件資料建構的輸入方式在下一節將詳細說明。

鬧鬧 Nàu. Noise; bustle.

| 房, n. fōng, the bride's reception room after marriage ceremony.

圖 5 多音一字舉例之一

資料來源：客英大辭典（2007：503）。

帕^{50,5} **吧**^{50,4} **帕**^{145,5} P'a. Mou-

手 | chou p'a mouchoir, essuye-main.

圖 6 多音一字舉例之二

資料來源：客法大辭典（1988：686）。

（二）有音無字

1. 記音不記字

除漢字輸入的特殊現象外，尚出現有「有音無字」及「借字」的情況發生，「有音無字」在《客英》中尤為常見，在記音採集時，如不確定漢字為何，則僅以讀音為標記。原件資料舉例如圖 7。原件中的音標記為「Mang5」，為尙未、還沒之意，現今依「台灣客家語常用詞辭典」的建議用字為「旨」。但是在《客法》裡出現「有音無字」的情況較少，一般是在不雅及粗俗用語的詞條中較常出現，原件資料舉例如圖 8。

Mâng. Not yet. Used for **m**
tshien.

m. chhut mìi, not yet finished.

圖 7 有音無字舉例之一

資料來源：客英大辭典（2007：453）。

圖 8 中的詞條皆為較不雅的用語，漢字「屌」的意思指男女交媾，屬不雅用字，其後所舉的詞條「Tiao3 nia7 me>7」意謂辱罵他人，故在原文辭典中除了只顯示該發音的漢字外，其後所舉的詞條皆只以音標標記，不採記詞條的對應漢字。

屌 44,6 *Tiaò*. Violer, copulam facere. *Tiaò kiā tsì tjē* (ou simplement *tiaò kiā*), je viole sa sœur (ou simplement je la viole). Parole d'impatience dite à tout propos contre les êtres soit animés, soit inanimés; elle répond aux mots grossiers b.. f., cr.. etc.
Tiaò niā mē je viole ta mère. Parole d'impatience comme la précédente.

圖 8 有音無字舉例之二
資料來源：客法大辭典（1988：1052）。

2. 借字

由於客語並非規範的語言，有許多發音一時難以找到相對漢字，在不確定用字下，便採取借用形、音、義相近者的「借字」方式。在早時的客語文獻資料中，因為漢字無系統整理，較難統一歸納，故「借字」情形常見，但近幾年來，教育部在《台灣客語常用詞辭典》公布一套建議用字，可作為參考之用。原件資料舉例如圖 9、圖 10。圖 9、圖 10 的三字皆為借字，客語發音為「hen5」，意思為「緊」，指物品被綁的很緊或拴的很緊之意，依教育部台灣客語常用詞辭典的建議用字為「緜」。「緜」有繃急琴絃之意，《楚辭·九歌·東君》：「緜瑟兮交鼓，簫鍾兮瑤虞。」王逸注：「緜，急張絃也。」由「急張絃」可引申為「牢固、不鬆動」意。另在《客法》出現有一罕見用字「𢂔」，原件資料如圖 11。

𢂔 Hēn. Stretched tight, taut.
h. pāng pāng, tight.

圖 9 借字舉例之一
資料來源：客英大辭典（2007：156）。

狠^{94,6} **很**^{60,6} *Hén* (pour *hèn*.) Très fort, qui résiste; solide, violent, sévère.— Désobéir, résister, intraitable. Quereller, contester. Cœur dur. Très, le plus.

毛辮編倒忒 | *maō piēn p'ìn tào tiêt hén* mes cheveux sont tressés trop serrés.

圖 10 借字舉例之二

資料來源：客法大辭典（1988：193）。

𢚣 ^{30,1} *Mâng*. Pas encore (contraction de 吻 還 *mô hân*).

| 愛 *mâng oí* pas encore, je n'en veux pas encore.

圖 11 借字舉例之三

資料來源：客法大辭典（1988：562）。

客語發音為「Mang5」，此音在《客英》中為有音無字，只採記音標不記其字，但在此處則以「𢚣」表示，屬於客家方言俗字，依教育部台灣客語常用詞辭典的建議用字為「𠵼」。

(三) 特殊用語

1. 教會用語

《客英》與《客法》二部辭典皆為傳道會及傳教士所編著。《客英》是麥愛華根據 Basel 傳道會（巴色會）的手稿所編，原文雖是由傳道會所著，但內容多採集客家方言字及成語、諺語，富涵客家文化，較少涉獵到教會用語。

反觀《客法》所收錄的詞條則頗有傳教意義，此部辭典是由法國籍汕頭傳教士雷卻利所編，內容除大量收錄諺語、成語外，其中亦參雜許多教會用語，可以看作一部研究客語早期文獻資料的工具書。原件資料舉例如圖 12—圖 14。

聖母麼 | 原罪 *chín móu mó niám nián ts'ouí la* St. vierge n'a pas été infectée par le péché originel; Immaculée Conception.

圖 12 教會用字舉例之一

資料來源：客法大辭典（1988：649）。

告解愛 | 心 *kaó kiāi oí siaò s̄im la*
confession doit être prudente.

圖 13 教會用字舉例之二

資料來源：客法大辭典（1988：841）。

耶穌只 | 十二宗徒 *yá sē tchít t'iaō chǐp
gní tsuāng t'oú Jésus n'a pris que douze
apôtres. (V. S.)*

圖 14 教會用字舉例之三

資料來源：客法大辭典（1988：1054）。

以上在《客法》中所收錄的詞條依序為「聖母麼染原罪」、「告解愛小心」、「耶穌只挑十二宗徒」，其中「聖母」、「告解」、「耶穌」等具有濃厚宗教色彩的用字與用語出現得十分頻繁。

2. 成語、諺語

《客英》與《客法》二部辭典的編輯體例中大多會列舉多組相關詞條做為釋義，其中常出現大量的成語及諺語，這些成語、諺語經口耳相傳及文字記載由來已久，可以藉由研究這些早期客語文獻資料的同時瞭解傳統客家文化及先

驕惰 *Kiau. Proud, arrogant,
boastful.*

| 兵必敗, *k. pin pit phài, the
boastful soldier will suffer,
defeat.*

圖 15 成語舉例之一

資料來源：客英大辭典（2007：275）。

鷗蚌相爭 | 人得利 *yoūt p'ōng siōng
tsāng, nḡ gñin tēt lí quand le martin-chas-
seur et l'huître sont aux prises, c'est le
pêcheur qui en profite (en se saisissant des
deux).*

圖 16 成語舉例之二

資料來源：客法大辭典（1988：622）。

民生活，具有智慧及教化功能，並有助於方言的學習。原件資料舉例如圖 15、圖 16。「驕兵必敗」的意思指為人驕傲必招失敗，若謙虛待人則可無往不利；「鶴蚌相爭，漁人得利」，意思為雙方各不讓步，結果兩敗俱傷，反讓第三者得利。其他還收錄有許多具客家文化的諺語，舉例如圖 17、圖 18。「飯好亂食，話唔好亂講」，意指米飯可以隨意吃，但話卻不可以亂說，若口不擇言恐易招惹是非；「拳頭唔打笑面人」，意為笑臉迎人讓人看了也難起爭執，往往能使事情更順遂，何必惡臉相向。故無論是成語抑或諺語的詞條收集，大多充分顯示出教化人心的功能，弘揚傳統美德。

- 飯** Fán Cooked rice, food & meal.
 (In Chhao-chiu pronounced
 Phón and Fén.)
- | 簍, *f. sau*, a basket for holding
 rice when steaming.
 - | 好亂食, 話唔好亂講 *f. hau*
lòn shit, và m hau lòn kóng, eat
 as you please, but be careful
 how you speak.

圖 17 諺語舉例之一
 資料來源：客英大辭典（2007：88）。

拳頭唔打 | 面人 *k'ian t'eoū mḡ tà siaó*
mién gnín on ne bat pas celui qui vous
montre de l'affabilité; la bonté désarme.

圖 18 諺語舉例之二
 資料來源：客法大辭典（1988：842）。

參、客語文獻數位典藏的資料建構

客語文獻數位典藏的資料建構可以分為「原件數位化」及「建立資料庫」兩大部分。原件數位化較接近目前大多數數位典藏的概念，主要目的在於「保

存」與「收藏」，以數位化影像取代實體原件珍貴易損壞，或存取、流通不易的缺失；透過基本的編碼和管理，未來可以成立數位博物館，一方面以數位化的方式典藏原件，另一方面置於網頁上供民眾瀏覽和參觀，利用網際網路無遠弗屆的特性，達到流通與社會教育的目的。

以《客英》及《客法》為例，將原件資料逐頁掃描成圖片檔案，以進行原件數位化。掃描的解析度均為 300dpi,¹ 但依用途的不同，分為「典藏級」與「瀏覽級」兩種類型。「典藏級」的圖片檔案設定為「.tif」格式，此等級的圖檔主要作為保存原件文獻資料所使用，其典藏方式可分為製成光碟片及異地備份，即在不同建築物提供備份硬碟或磁帶存放。「瀏覽級」的圖片檔案設定為「.jpeg」格式，用於存放在瀏覽網頁上的資料庫圖片，目的是提供給一般民眾線上檢索之用，可以直接在線上連結點選，以利同時比對辭典的原件資料。文獻數位典藏資料建構中的「建立資料庫」則是一項比較大的工程，其最終目的是希望建立一個較為完整且可供學術研究的客語語言資料庫。建立資料庫的基本工作是先用 word 檔建立《客英》、《客法》兩本辭典的全文電子檔，再將各項資料依其欄位性質轉換成數位資料庫。

由於《客英》、《客法》屬辭典性質，每一詞條中包含漢字、拼音、解釋及例句等不同的訊息，為便於利用電腦程式分辨各種不同的訊息，以利全文電子檔轉換成資料庫，輸入時使用特定的符號加以分隔。此外，在數位化時經常會遇到有關造字、漢字、拼音及編排、校對等方面的問題，因此，於建立全文電子檔時，也須先制訂出一套資料建構的方法，以人工逐一檢驗和校對。以下說明《客英》、《客法》轉成數位化全文資料時所作的特殊處理。

一、欄位順序

將每一漢字及詞條分為四個欄位，依序為漢字（或詞條）欄、拼音欄、解釋欄及頁碼欄，各欄使用「*」分隔，依序以「漢字（詞條）欄*拼音欄*解釋欄*頁碼欄」建構資料，但其中的頁碼欄須輸入四碼數字，即第一頁以「0001」表示之。

在原件詞條中，遇到標記「|」的符號時，直接輸入與原漢字相對應的漢字。原件舉例如圖 19。在內文中的「| 公」應該輸入為「阿公」，以下列方式表示：「阿公*a-kung*grandfather.」。

¹dpi 的原文為 Dot Per Inch，意指掃描器在掃描時每英吋的取樣點數。

I 公, a *kung*, grandfather.

圖 19 欄位順序舉例
資料來源：客英大辭典（2007：1）。

其他有關「斷詞」方面，若原件依中央研究院的分詞原則被認定為「詞」，須以符號「-」做為詞與詞間的連結，例如，阿公 a-kung。

二、字母上標的符號

字母上標的符號轉換成 2、3、4、5、7、8、ii、ü，共計 8 種，置於拼音後面，例如，上 Shóng，輸入為 Shong2；而法文中原本的上標符號則直接置於字母後面，如，répandre，輸入為 re2pandre。以表 4 說明之。

《客法》、《客英》拼音規則的輸入方式如次：《客英》第 1 調等於《客法》第 7 調，《客英》的第 1 調及第 4 調在標示原則上省略，但為方便電腦判別與英文之不同，以利處理全文檢索，拼音在解釋欄出現時仍會補上調號，如圖 20。輸入如：「上*Shong2*To go up. To send up, to put on. See under

表 4 字母上標符號表

客英大辭典	客法大辭典	輸入方式
á	á	a2
à	à	a3
	â	a4
â	â	a5
	ã	a7
ã	ã	a8
ü		ii
	ē	e>

上 Shóng. To go up. To send
up, to put on. See under
Shong. shóng.

圖 20 字母上標符號舉例之一
資料來源：客英大辭典（2007：724）。

Shong1, shong3.」。又如：「催*Kok*A surname. The name of a general (Li2-kok4) mentioned in the Sam Kwet Chi.」。

催 Kok. A surname. The name of a general (*Li k.*) mentioned in the Sam Kwet Chi.

圖 21 字母上標符號舉例之二
資料來源：客英大辭典（2007：320）。

《客法》的拼音在解釋欄中與法文的辨識不易，為方便電腦區辨及處理全文檢索，會在解釋欄中出現的客語拼音前後以「@」包覆表示，如圖 22。輸入如：「灘**Siam*5.*Source intermittente.— Tremper, humecter, répandre des bienfaits.— Employé pour 罷 *siam*5@ de2truire, exterminer.」。

灘^{85,17} *Siam*. Source intermittente.— Tremper, humecter, répandre des bienfaits.— Employé pour 罷 *siam* détruire, exterminer.

圖 22 字母上標符號舉例之三
資料來源：客法大辭典（1988：837）。

三、符號使用

（一）符號「,」

同一詞條當中，有時會出現多個相通漢字（俗體字）對應一讀音或多個讀音對應一漢字，甚至多個漢字對應多種讀音的情形，此時多重表示的各字或各音中間須以半形逗號「,」分開，並用「\$」符號前後包覆，如圖 23。輸入如：「\$𦨇,𦨇,\$*\$Ta7m, Nia7m.\$*Oreille plate et sans rebord.— Mai5tre.」。

𦨇^{128,4} **𦨇**^{128,5} **𦨇**^{158,5} *Tām, Niām.*
Oreille plate et sans rebord.— Maître.

圖 23 符號使用處理舉例之一
資料來源：客法大辭典（1988：902）。

但如果在原件中不同的詞語對應同一解釋時，須將這些詞語個別分開成數組相同解釋的詞條，如圖 24。如：「敬重*kin2-tch'oung2*avoir en haute considération.」，「尊重*tsoun7-tch'oung2*avoir en haute considération.」。

**敬 | kin tch'oung 尊 | tsoun tch'oung avoir
en haute considération.**

圖 24 符號使用處理舉例之二
資料來源：客法大辭典（1988：1021）。

（二）符號「（）」

在漢字、音、詞條中，單方面出現「（）」的括號標示時，直接輸入不做其他變更，但若遇有字、音與法文解釋都括有夾注號時，則將之分為兩組詞條，如圖 25。如：「盡人事聽天命*ts'in2 gniñ5-se2 t'in2 t'ien7-min2*faites tout votre possible, et remettez-vous en au bon plaisir du ciel.」，「盡人事聽主命*ts'in2 gniñ5-se2 t'in2 tchou3-min2*faites tout votre possible, et remettez-vous en au bon plaisir du Dieu.」。

**| 人 事 聽 天 (主) 命 ts'in gniñ sé t'in
t'ien (tchou) min faites tout votre possible,
et remettez-vous en au bon plaisir du ciel
(de Dieu).**

圖 25 符號使用處理舉例之三
資料來源：客法大辭典（1988：1232）。

（三）符號「#」

在詞條與詞條間中有時會插入一段補充說明的文字敘述，但此段文字與上下文中的詞條意義不直接相關，因此須在前後加上「#」做為區別，如圖 26：文中這段文字「Chaque mois a son nom particulier」指每個月都有其特定的名稱，與上下文中的詞條敘述無直接關係，故須在此段文字的前後加上「#」符號。輸入如：「#Chaque mois a son nom particulier.#」；又如圖 27 中的文字「Hiou. Dans certains pays les mots en Hi se prononcent hiou」指在某些地區會將 hi 發音成 hiou，補充說明漢字的「休」。輸入如：「#@Hiou@. Dans certains pays les mots en @ Hi@ se prononcent @hiou@.#」。

| 尾 *gnët mì* la fin du mois.

四 | 尾 *sí gnët mì* à la fin de la quatrième lune.

Chaque mois a son nom particulier.

朔 | *sôc gnët* le premier jour de la lune.

正 | *tcháng gnët* 元 | *niân gnët* la première lune.

二 | *gní gnët* 杏 | *hén gnët* la 2^e lune.

圖 26 符號使用處理舉例之四

資料來源：客法大辭典（1988：151）。

HIOU

Hiou. Dans certains pays les mots en *Hi* se prononcent *hiou*.

休_{9,4} *Hiou*. Se reposer. Cesser. Résigner sa charge. Négation qui s'emploie devant un impératif. Répudier. Bon, bien-fait, avantage; heureux, prospère. Trouver bon, approuver, louer. Pardonner, être indulgent. Modéré, frugal, économique.

圖 27 符號使用處理舉例之五

資料來源：客法大辭典（1988：227）。

四、字音義之特殊處理

(一) 一字多音

出現一字多音的情況時，如前文所述，須在各漢字中間以逗號隔開，如圖 28。文中的「Lin5, Lien5」兩種拼音移至第 2 欄位，在兩音中間以半形逗號「,」隔開後，前後須再加上「\$」以區隔，輸入如：「憐*\$Lin5, Lien5\$*To commiserate, have pity on.」。

(二) 多字一音

辭典中有時會將同一漢字的異體字放在一起，是為多字一音，處理方式同「一字多音」，如圖 29。輸入如：「\$辯, 辟\$*P'ien2.*Distinguer, se2parer, discerner, juger, de2cider. Discuter. Gouverner, diriger. Employe2 pour 辯 p'ien2.」。

憐 Lîn. To commiserate, have pity on. Also pronounced Lièn.

圖 28 一字多音處理舉例
資料來源：客英大辭典（2007：405）。

辯 ^{160, 14} 辨 ^{160, 7} P'ién. Distinguer, séparer, discerner, juger, décider. Discuter. Gouverner, diriger. Employé pour 辩 p'ién.

圖 29 多字一音處理舉例
資料來源：客法大辭典（1988：750）。

（三）同字異音義

辭典中將同一漢字下不同音不同義的兩詞語置於同一詞條中時，直接增錄成二組詞條。如圖 30。針對「Liau2, Liuk8」這兩種拼音及後接的兩種釋義，輸入為：「蓼*Liau2*Various species of Polygonum,Smartweed.」，「蓼*Liuk8*luxuriant growth.」。

蓼 Liáu. Various species of Polygonum. Smartweed. Also read Liuk, luxuriant growth.

圖 30 同字異音處理舉例之一
資料來源：客英大辭典（2007：417）。

（四）同字音異義

原件詞條與詞條中插入一段文字，而此段為漢字的另一種解釋時，視為同字音不同義，將依兩種解釋分為兩組，與前述參、三（三）節中上符號「#」的處理方式不同，如圖 31。嘎 Sa 的兩種解釋：「Voix casse2e, enroue2e. Montrer les dents. De2rober, voler.」及「Bruit du rat qui grignotte le riz.」，輸入如下：「嘎*Sa2.*Voix casse2e, enroue2e. Montrer les dents. De2rober, voler.」，「狗牙嘎嘎忼人*keou3 nga5 sa2-sa2 hong5 gni5*le chien gronde en montrant les dents.」，「嘎*

Sa2.*Bruit du rat qui grignotte le riz.」，「老鼠食穀嘎嘎响*lao3-tch'ou3 chit8 kou4c sa2-sa2 hiong3*on entend le rat qui grignotte.」。

嘎 30,10 Sá. Voix cassée, enrouée. Montrer les dents. Dérober, voler.

狗牙 || 恍八 keoù ngâ sá sá hông gnîn
le chien gronde en montrant les dents.

Bruit du rat qui grignotte le riz.

老鼠食穀 || 响 laò tch'ou chit koûc sá
sá hiòng on entend le rat qui grignotte.

圖 31 同字異音處理舉例之二

資料來源：客法大辭典（1988：794）。

（五）合音詞

當兩個拼音合成一音時，是為合音詞，如圖 32。輸入如：「一個絲毫都唔愛*yt4-ke>2 se7-hao5 tou7 \$moe>2, mg5-oi2\$*je ne veux absolument rien.」。

一個 | 毫都唔愛 ÿt ké· sē haô tou moé·
(mg̊ oí) je ne veux absolument rien.

圖 32 合音詞處理舉例

資料來源：客法大辭典（1988：810）。

（六）同義詞

當同義詞放在一起時，須拆成兩組詞條，如圖 33。當文中詞條的「冤慘」與「悽慘」為同義時，輸入如：「冤慘*yan3-ts'am3*e5tre dans la peine, la souffrance.」，「悶慘*ts'i7-ts'am3*e5tre dans la peine, la souffrance.」。

**冤 | yan ts'äm 悅 | ts'i7 ts'äm être dans la
peine, la souffrance.**

圖 33 同義詞處理舉例

資料來源：客法大辭典（1988：1153）。

（七）漢字與詞組合的詞條

當詞條由漢字及詞彙組合而成者，須分成兩組詞條輸入。如圖 34。方式如

下：「快*Fai2, (V. k'ouai2.) *」，「快利*fai2-li2*aigu, aiguise2, qui coupe bien.」。

快^{61,4} *Fai.* (V. *k'ouai.*) | 利 *fai li aigu,*
aiguisé, qui coupe bien.

圖 34 漢字與詞組合處理舉例
資料來源：客法大辭典（1988：85）。

(八) 英文解釋

以英文解釋意思之餘，還附上「口語白讀」，如圖 35。當「鍵」這個漢字同時出現口語白讀「Lien3」時，輸入如：「鍵*\$Lien5, Lien3\$*A chain.」。

鍵 *Lièn.* A chain. Coll. *Lièn.*

圖 35 英文解釋處理舉例
資料來源：客英大辭典（2007：401）。

五、原件有誤之情形

(一) 漢字改正

改正錯字時，須在改正字的後面加上原件錯字，並以「[]」括弧，以保留原件內容，如圖 36。原件中的漢字「山」，實際應為「出」，輸入如：「出[山]版*tch'out4-pan3*sortir des presses.」。

山 | *tch'out pàn sortir des presses.*

圖 36 漢字改正處理舉例
資料來源：客法大辭典（1988：697）。

(二) 拼音改正

改正拼音時，須在改正音的後面加上原件拼音，亦以「[]」括弧。如圖 37。原件中的拼音「Yap7」應為「Yap8」，輸入如：「葉*Yap8.[Yap7]*Feuille d'une plante, ce qui ressemble a3 une feuille.」。

(三) 原件漏印

1. 缺字無法判別

當有拼音但原件漏印而導致缺字無法判別的情況時，以●表示，如圖 38。

原件中的「 | 腳」的對應漢字為「人腳」，而在後面由於無法辨識拼音「han5 tshau5 (yun5)」的對應漢字，輸入如：「人脚●●*nyin5 kiok han5 tshau5 (yun5) *the men are so enterprising, capable.」。

葉 140,9 *Yāp.* Feuille d'une plante, ce qui ressemble à une feuille.

圖 37 拼字改正處理舉例

資料來源：客法大辭典（1988：1361）。

| 脚, n. *kiok h&fn tsh&au (yun)*, the men are so enterprising; capable.

圖 38 原件漏印處理舉例之一

資料來源：客英大辭典（2007：554）。

2. 缺字可判別

在缺字可判別的情況下，直接加上所缺之字，並在修正後的漢字後面輸入[0]，以表示原件缺漏。漢字缺字舉例如圖 39。原件中的「擎」為原對應漢字，並查到拼音後面所接的「k'oi7」漢字應為「開」，輸入如：「擎開[0]麵包*pac4-k'oi7 mien2-pao7*rompre un morceau de pain.」。

| 麵包 *pâc k'or mién pâo rompre un morceau de pain.*

圖 39 原件漏印處理舉例之二

資料來源：客法大辭典（1988：688）。

拼音缺字舉例如圖 40。輸入如：「唐字今有愈多新個*t'ong5-se2 kin7[0] you7 e>2-to7 sin7 ke>2*aujourd'hui on trouve beaucoup de caractères chinois nouveaux.」。

唐字今有愈多 | 個 *t'ong sé you é· tō sin kē aujourd'hui on trouve beaucoup de*

圖 40 原件漏印處理舉例之三

資料來源：客法大辭典（1988：896）。

3. 原件印刷難以辨識

當原件的印刷不清晰而導致難以辨識時，亦使用●處理，如圖 41、圖 42。輸入如：「輿*Yi5*A carriage, To carry, T●earth, the people.」，「*Tshoi2*To ●, to guess, to call to mind.」。

 Yi. A carriage. To carry.
T earth; the people.

圖 41 原件漏印處理舉例之四
資料來源：客英大辭典（2007：1091）。

Tshoi. To track; to guess; to call
to mind.

圖 42 原件漏印處理舉例之五
資料來源：客英大辭典（2007：988）。

圖 43 原件中的「| 牙」的對應漢字應為「乳牙」，在拼音部分中的「n.」為「nen3」的縮寫，但最後面的詞條因印刷無法辨識，亦以「●」處理，輸入如：「*nen3 t●u*udder of a cow.」。又圖 44 原件中的「| 監」與「| 口」的對應漢字應為「入監」與「入口稅」，此「稅」字雖印刷不清，但仍可依其拼音「shoi3」來辨識，但是在「入監」這組詞條後面的拼音因印刷無法辨識，輸入如：「*nyip8 tsa●2*to fill closely.」。

| 牙, n. ngā, 1st teeth.
n. ku tun, full breasts.
n. t'u, udder of a cow.

圖 43 原件漏印處理舉例之六
資料來源：客英大辭典（2007：505）。

| 監: n. kam, to go to prison.
n. ted', to fill closely.
| 口 稅 n. khēu shōi, import tax.

圖 44 原件漏印處理舉例之七
資料來源：客英大辭典（2007：559）。

六、多字

若遇到多字的情況發生，則直接刪去，如圖 45。原件中「|」的對應漢字應為「不」，後面的拼音「p.」為「put」的縮寫，而「sin3」所對應的漢字應為「信」，輸入如：「不入官信*put nyip8 \$kwan, kwon\$-sin3*the magistrate does not give it credence.」。

| 入 官 性 信, p. nyip kwan
(kwon) sin, the magistrate does
not give it credence.

圖 45 多字處理舉例
資料來源：客英大辭典（2007：659）。

七、字序顛倒

當漢字的順序顛倒時，直接改為正確字序，原則上以拼音為主，如圖 46。原件的拼音為「on2 chou3」，且在詞條釋義上應為「案首」，輸入如：「案首* on2-chou3*celui qui est le premier sur la liste des 秀才@siou7-ts'oi5@ bacheliers.」。

| 案 ón chouù celui qui est le premier sur
la liste des 秀才 siou ts'oi bacheliers.

圖 46 字序顛倒處理舉例
資料來源：客法大辭典（1988：63）。

肆、結論——數位典藏之預期成果

隨著科技日益發達，落實數位典藏資訊的知識化與社會化，與朝向知識社會之建構及發展的要旨相輔相成。語言之可貴不只在於做為溝通的工具，它還包含許多思維、價值觀等文化，在講求相互尊重和寬容的多元社會，不同語言的交流更能夠豐富彼此的生活，激起更多的創意。由於地理環境與歷史因素影響，造就台灣豐富的語言生態，然而時空移轉，社會型態變遷快速，語言流失日趨嚴重。過去許多平埔族語言與原住民語言或因異族統治，或因耆老過世而消聲匿跡，近代許多地區的客語、閩南語受到外來強勢文化衝擊，也都面臨語

言消失的危機。中華世界遺產協會（2001）之分類，除自然遺產與文化遺產外，新增「口述與無形人類遺產」一類，「語言」亦被列入其中。聯合國教育科學及文化組織的（2003）《保護非物質文化遺產公約》亦決議，通過國際共同合作，保護以「語言多樣性」為優先之人類文化多樣性。可見，無論在觀念上或實質上，語言之必須且值得「典藏」乃世界趨勢。閩客語典藏計畫便希望藉由客語文獻資料的數位典藏化，為保存母語貢獻一己之力。

然而，語言數位典藏若僅止於「收藏」資料，未能充分利用數位化資料存取快速、整合容易，以及便於各種「後設」加工，則辜負了現代科技發展所帶來的便利。隨著現代社會透過網路的無遠弗屆，以及科技人性化的設計與互動式學習，語言的數位典藏可以讓語言典藏的效益不僅在於「保存」，還可以「活化」語言資訊，透過典藏資料庫與使用者之間的互動，發揮母語教學、社會推廣或學術研究等綜效。客語文獻數位典藏於「原件數位化」外，在「建立資料庫」上所作的嘗試便是基於這樣的目標。除了建立文獻的全文電子檔外，在校對的過程中，遇拼音不符、漢字錯字，或原件缺字、漏印、印刷不清晰或多字、字序顛倒等原件內容有誤的情形時，一方面忠於原件，保留其原有樣貌，一方面也藉由校對逐一修正，並使用特殊的符號註記修改的資料，以利數位典藏化後之相互比對。

將文獻電子全文檔以後設標記處理過後，經過轉換，再上傳至網站伺服器的資料庫，便可以使用網頁設計的程式語言製作線上查詢。目前設定的查詢功能提供「查詢詞條」或「全文檢索」；關鍵詞方面，可輸入「閩南語白話字」、「閩南語漢字」、「客語白話字」、「客語漢字」、「日語」、「英語」、「法語」等；亦可以分別選擇所需查詢的辭典，或直接蒐尋全部辭典；查詢方式有「精確查詢」、「詞首相同」、「詞尾相同」、「不區別聲調」、「韻母相同」等；最後可以選擇顯示方式，逐一設定以上的查詢選項後再按下執行鍵，這時可能會出現一筆或多筆網路資料，每筆資料都可供瀏覽者再連結至原件的圖片資料庫，以便與原件資料作比對。

閩客語典藏計畫網站的線上辭典版面，目前以閩語為主，待《客英》及《客法》校對的資料建構完成後，亦會放入此網站中，目前網頁的界面如圖 47 所示。

因為線上檢索是以漢字與詞條為基準，故辭典原件最前頁的序言及後面的附錄，在線上辭典中皆不提供查詢，亦不顯示圖檔。此外，由於《客英》與《客法》兩大辭典的拼音系統不一致，對於不瞭解這兩套拼音系統的人而言，在查詢上會有許多不便，為此，特地在網頁上提供另外的通用拼音介面；使用者只

要使用通用拼音查詢，亦可得到這兩本辭典所對應的音標資料。語言是珍貴的文化資產，數位典藏計畫乃因應網際網路的普及，達成一種資源共享的數位教育成果，以提升學習的效率。以原件資料來建構數位教學平台，透過推廣與應用，不但可以讓數位學習解除遠距限制，並營造出良好的數位學習環境，更可讓這些語言文化完善地保存與傳承。

目前這項數位典藏計畫尚在進行中，且仍會不斷創新與改進，期提供社會大眾最正確與最便捷的文獻典藏資訊。網際網路的無遠弗屆可以打破時間與地域的限制，藉由文獻的校對與整理，建立一套早期客語文獻資料的數位化學習，增進學習資源的開放與分享。將這些文獻資料典藏數位化，不但可以提升全民的數位學習素養，更可推動學術研究的數位學習風氣。與客語語言典藏相關之研究計畫尚有 Kawai Chui & Huei-ling Lai (2008). The NCCU corpus of spoken Chinese: Mandarin, Hakka, and Southern Min、羅鳳珠（2007）發表之〈台灣客語語言典藏的現況與資源整合——以網路之台灣客語資源為範圍〉、謝杰雄（2008）發表之〈說明與演示一個線上客語語料庫〉等。各種不同類型、不同內容的客語語料庫不斷建置和新增，正顯示透過資料庫與網路資源存取和傳遞資料是保存與推廣客語方興未艾的趨勢。

除了《客英》、《客法》外，該計畫持續完成清末民初之傳教士文獻，如：啓蒙淺學、聖經書節擇要、醫界客話讀本等之典藏工作，並繼續蒐集新的文獻，且持續培養客語研究人才，將在 2013 年開始規劃辦理語料庫使用說明之相關座

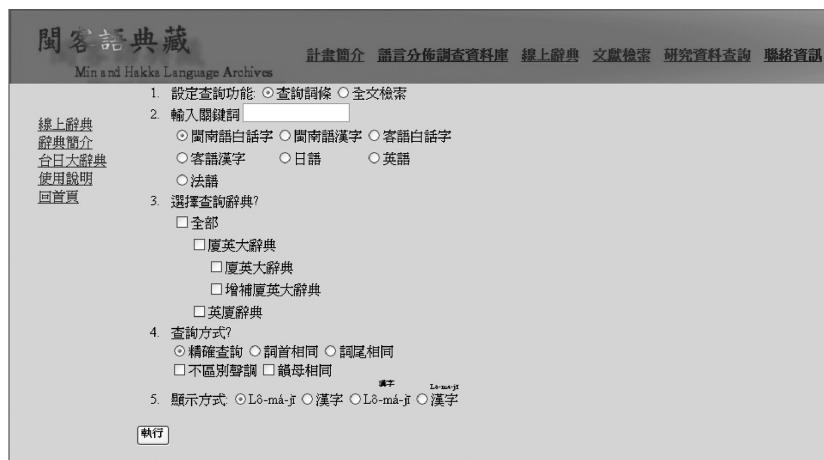


圖 47 閩客語典藏計畫線上辭典的網頁界面圖
資料來源：中央研究院語言學研究閩客語典藏計畫網（無日期）。

談會與教師研習，並與教育部合作研擬鄉土語言相關教材。教育乃百年大計，透過相關規劃，本土語言文化將可深入扎根。此外，亦應進一步結合國內資源，鼓勵相關領域之碩、博士生使用語料庫進行客語語法、詞彙、文本等研究；為培育國家未來的語言研究人才，預計 2013 年起每年舉辦 1 場說明會及研習營，並視成效集結研究成果出版專書；自 2014 年起估計每年至少 1 位碩、博士生運用語料庫撰寫客語相關論文。

閩客語數位典藏的資料豐富且龐大，預計資料庫完成後將成為目前最完整之閩客語料庫，不但可提供學術研究使用，並具教學及社會教育功能。因此閩客語典藏的永續經營將充實典藏內容、維護、管理及推廣現有典藏成果使典藏內容發揮最大功用。

參考文獻

- 中央研究院文獻處理實驗室漢字構形資料庫網站（無日期）**。2009 年 8 月 1 日，取自 <http://cdp.sinica.edu.tw/cdphanzi/>
- 中央研究院語言學研究閩客語典藏計畫網（無日期）**。2009 年 8 月 1 日，取自 <http://minhakka.ling.sinica.edu.tw/bkg/index.php>
- 中華世界遺產協會（2001）**。人類口述與無形遺產——非物質文化遺產。2009 年 8 月 1 日，取自 <http://www.what.org.tw/heritage/#top>
- 聯合國教育科學及文化組織（2003，10 月 17 日）**。保護非物質文化遺產公約。2009 年 8 月 1 日，取自 <http://unesdoc.unesco.org/images/0013/001325/132540c.pdf>
- 教育部台灣客語常用詞辭典（2008）**。2009 年 8 月 1 日，取自 <http://hakka.dict.edu.tw/>
- 數位典藏與數位學習國家型科技計畫網（無日期）**。2009 年 8 月 1 日，取自 <http://www.teldap.tw/>
- 林英津（1994）**。論《客法大辭典》之客語音系。載於中華民國聲韻學學會、台灣師範大學國文系所、高雄師範大學國文系所（主編），**聲韻論叢第二輯**（383-422 頁）。台北市：台灣學生書局。
- 湯培蘭（1994）**。《客法大辭典》音韻研究。國立暨南國際大學中國語文學系碩士論文，未出版，南投縣。

黃詩惠（1993）。《客英大辭典》音韻研究。國立彰化師範大學國文研究所碩士論文，未出版，彰化市。

謝清俊、莊德明（2005）。漢字構形資料庫的建置與應用。載於郭玉茹（主編），**漢字與全球化國際學術研討會論文集**（頁 119-133）。台北市：台北市文化局。

王逸（注）、王熙元（導讀）（1997）。**楚辭（上）、（下）**。台北市：金楓。
謝杰雄（2008）。〈說明與演示一個線上客語語料庫〉。載於羅肇錦、陳秀琪（主編），**2008 第八屆國際客方言學術研討會會前論文集**（437-447 頁）。
中壢市：中央大學客家學院語文研究所、台灣客家語文學會。

羅鳳珠（2007）。〈台灣客語語言典藏的現況與資源整合——以網路之台灣客語資源為範圍〉。載於鄭錦全、何大安、蕭素英、江敏華、張永利（主編），**語言、社會與文化系列叢書之二**（379-385 頁）。台北市：中央研究院語言學研究所。

Rey, Ch. (1988)。客法大辭典 (Dictionnaire Chinois-Francais Dialecte Hac-Ka)。台北市：南天。

MacIver, D. (主編)、Mackenzie, M. C. (修訂) (2007)。客英大辭典 (A Chinese-English Dictionary Hakka-Dialect)。台北市：南天。

Kawai Chui & Huei-ling Lai (2008). The NCCU corpus of spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, 6(2), 119-144.