

電子佛典中處理中文版本的方法

謝清俊* 陳昭珍** 莊德明*** 周亞民****

* 中央研究院資訊科學研究所研究員,臺灣大學圖書館學研究所兼任教授

** 輔仁大學圖書資訊學系講師,臺灣大學圖書館學研究所博士班學生

*** 中央研究院資訊科學研究所特約研究助理

**** 元智工學院資訊科技研究中心系統工程師

摘要

本文以佛教經典為例，探討如何在計算機中處理古籍多版本的問題，其中包括：多版本的表達方式，各版本間相互比對和轉換的方法等等。選擇佛典的理由之一是它通常有科文，科文是一個樹狀結構的經文提要，比目錄還詳細，以往主要用於方便解釋或檢索經文。在計算機裡，科文正可用做經文的內容、主題、或文句等的檢索和比對。若為一般古籍編一科文（或編一較詳細的目錄），則此古籍即可適用於本研究提出的系統。在漢學研究的領域裡，無論是考據之學或是義理之學，版本問題都居首要，是故本研究應該對我國古籍之電子化具有積極的意義。

本研究曾在個人電腦上作了系統之雛型，本研究所用的文件表達和標誌方式亦曾轉換為 ISO 8879 SGML文件標準形式，以利文件流通和共享，這些工作的例子，均在本文中有所說明。

中華民國八十三年四月六日於文獻處理實驗室
中央研究院資訊科學研究所

壹、佛學與佛典

根本之佛典，唯印度梵文或巴利文寫成之經、律、論等三藏。然而，就古代佛學內容之深之廣，和理論之嚴謹而言，其精華恐怕集中在印度和中國【註一】。近代佛學之興起，肇自十九世紀初歐洲學者，他們以語文學和文獻學之研究方法，對治梵文和巴利文之佛典。這種研究方法，有類於我國對治古籍的考據之學。由於佛典涉及多種語文，如：梵文、巴利文、藏文、日文、漢文等等，並跨越千年以上之時空，此法遂成為近代佛學研究的基礎功夫。

之後，藏文資料漸經發掘使用，對佛典義理之學的研究亦逐漸開展，在研究方法上也因而頗有增益變化。近年來，歐日美澳各地區均有相當的研究規模，研究成果漸碩。可是，對於中國漢文的佛典，則一直沒有得到應有的重視【註二】。有鑑於此，有識之士遂於五、六年前倡議建立計算機中多語文大藏經（即電子佛典）。目前，巴利文、藏文和泰文之電子藏經已有成績，而漢文的電子藏經一則由於數量龐大，二則因計算機處理漢語有諸多問題，而遲遲未決。是故發展漢文電子佛典的處理技術，仍為目前佛學研究的重要命題。【註三】

我國自漢桓建和元年（西元147年）傳譯佛典以來【註四】，歷代皆有譯述。然而，六朝以前之譯品（稱為古譯）良莠不齊，至唐玄奘（西元600 - 664年）之後，始有較完備之譯經制度，並產生較優良的譯品【註五】。然而，除語文有隔的原因外，佛典之翻譯制度之良窳，譯者領悟之深淺，和時空之差異等原因，常致使一部經有數種譯本。例如：《華嚴經》有40卷，60卷，80卷等不同之譯品，又如《首楞嚴經》有九種譯本等等，不勝枚舉。當然，我國擁有之佛典並非全由譯得，國內撰寫之論著並無上述的譯本問題，但是由翻譯而產生的版本問題卻是我國佛典的特色。

至於佛典之流傳，在宋以前多靠手抄本及零星之彫本，刻於石上更是一特色。自從宋太祖發起刻印《蜀本大藏經》後，宋元明清各朝均有數起官方或民間起刻《大藏經》之事，故各種《大藏經》之版本即不下十餘起【註六】，若將海外典藏計入，如日本和高麗之藏等等，則為數更多。雖然佛典有如一般古籍，常經兵災之厄而保存不易，又經輾轉傳誦抄刻，難免有失真偽誤之情，然而佛典較一般古籍有利者，即在於海內外各大藏版本可提供蒐輯遺佚、底本互勘、和群籍鈞稽之機會。可是，要做到些事，只靠人力是不太可能的，如果計算機能夠協助處理多版本問題，將大有助於佛學之研究。

【註一】吳汝鈞，《佛學研究方法論》（台灣學生書局 1982）第107頁

【註二】同註一，第198頁

【註三】美國加州大學柏克萊分校 Lewis Lancaster 教授從事多語文電子藏經計劃已有多年；台灣佛光山亦有〈中文電子佛經〉之計劃；日本花園大學 Urs App 教授亦從事漢文佛經之研究，並出版 The Electronic Bodhidharma 可資參考；台北中央研究院資訊科學研究所之文獻處理實驗室也在1988年起從事電子佛典相關之研究，並製作佛經全文資料庫。1994年1月，TEI/SGML 工作組特別假香港科技大學邀集美、日、韓、馬、港、台各地學者研究漢文電子佛典之事。

【註四】有關開始譯經之年代，其實有多家說法。可參考：1)梁起超《佛學研究十八篇》(台北，中華書局) 2)高聞天，《般若波羅蜜多心經釋要》(台北，大學蓮友社，1987) 3)呂徵《佛學泛論》(台北，三人行出版社，1974)

【註五】早期譯經情形可參考宋 贊寧《高僧傳》卷三

【註六】呂徵，《佛學泛論》(台北，三人行出版，1974)第28~31頁；倭虛大師述，《影塵回憶錄》(高雄淨宗學會，1992)第421-449頁

貳、古籍之校讎

對治不同版本之古籍，古有校讎，今日校書。校讎方法大致包涵了：逸書蒐輯，真偽辨別，底本互勘，群籍鈞稽，篇第審定，和目錄論次等工作【註七】。這些工作，原來都是人去做的，是故目前所有的方法也都是因人而設，因人而寫。如果要計算機來幫忙，首要之事便是要分辨和規劃：那些事適合給計算機做，那些事還是該由人做，以及人和機器如何相互配合等。

在逸書蒐集方面，主要還是靠人。待人蒐集到書之後，計算機可協助編立書目，並可將該書原文製成機器可閱讀之電子形式，除了準備給自己用以外尚可供同行參考。在本研究中假設所輯之書均已製成機讀形式。至於真偽辨別方面，還是靠人；計算機或可提供快速之比對與相關文件之互參功能，來協助人做真偽之辨，然而這些功能卻已包涵在底本互勘之中。

底本之互勘，據張舜徽引陳垣《元典章校補釋例·第四十三〈校法四例〉》所用者四端：【註八】

- 一、校對法：以同書之祖本或別本對讀
- 二、本校法：以本書前後互證
- 三、他校法：以他書校本書
- 四、理校法：無古本可據、或數本互異無所適從時用之。

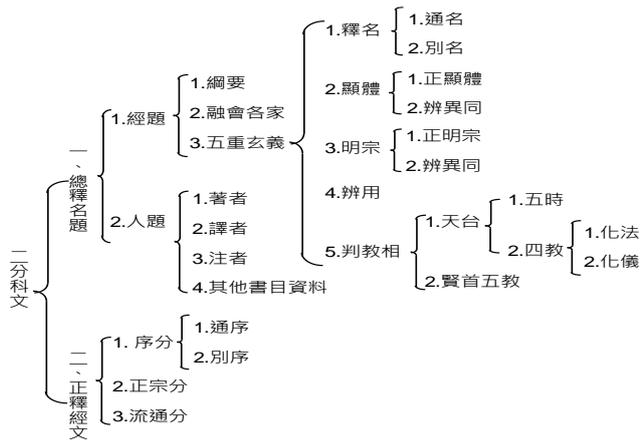
以上四法，均涉及同書或不同之書（或不同之版本）相關資料間之鈞稽比對，如果設計得當，計算機於此應可發揮極大之作用。至於群籍鈞稽之條，事實上已含在理校法之中。至於篇第審定和目錄論次之主要工作，主要還是靠人，計算機可協助作目錄或篇第之整理工作。

在佛典中有目錄作用者為該經典之科文，亦稱科判。科文是一個樹狀結構之經文提要，比一般目錄詳細。由於佛典有一致之編輯體例，故一般之佛典可共用一個科文，請參照圖一之說明。

表一、張舜徽所舉之十種常用的注記體例

- 一、凡文字有不同者，可注云：『某，一本作某。』
(或具體寫明版本名稱)
- 二、凡脫一字者，可注云：『某本某下有某字。』
- 三、凡脫二字以上者，可注云：『某本某下有某某幾字。』
- 四、凡文字明知已誤者，可注云：『某當作某。』
- 五、凡文字不能即定其誤者，可注云：『某疑當作某。』
- 六、凡衍一字者，可注云：『某本無某字。』
- 七、凡衍二字以上者，可注云：『某本某字下無某某幾字。』
- 八、字倒而可通者，可注云：『某本某某二字互乙。』

- 九、字倒而不可通者，可注云：『某本作某某。』
- 十、文句前後倒置者，可注云：『某本某句在某句下。』



圖一、佛經之科文

【註七】胡樸安、胡道靜，《校讎學》（台灣商務印書館，1990台二版，原書成於1931） 54~84頁

【註八】張舜徽，《中國古代史籍校讀法》（上海古籍出版社 1986二刷） 181~182頁

在圖一之科文中，經題和人題部份即含有該經典之部份目錄資料。每部經典均可整理出總釋名題中之經題和人題、以及正釋經文中之序分，正宗分，以及流通分等部份。各經典重要的正文，全在正宗分之內。【註九】（至於正宗分之結構請參閱下文中之說明）。科文之主要用途在方便解釋或檢索經文，其結構式極適合計算機之處理，它不僅提供可相互比對之處所及範圍，更可作內容或主題之檢索及參考，甚至可協助作詞彙的整理以及文句中訛、衍、缺、脫等之校勘。

參、體例之整理

關於各版本間互異之實際情況，有衍文、偽體、倒置、脫落、誤改、誤解、誤增、誤刪，以及簡策錯亂、篇章顛倒等多種現象，俞樾在《古書疑義舉例》中將他校書經驗歸納為37例，可稱通例【註十】。依此37例觀之，其分劃頗細，分劃之原則多涉及導致錯誤的原因。這些原因涉及文章的內容，以〈衍例〉為例，即劃分為「兩字義同者」、「兩字形似者」、「涉上下文者」、及「涉注文者」等四例。這些原因均非目前之計算機技術所易於處理者。

然而，若從注記之常例而觀之，張舜徽僅舉出十種作為常見之情況（請參照表一【註十一】）。例如其第一條：「凡文字有不同者，可注云：『某，一本作某』。」這種注記實是側重文字差異之形式，而非如俞樾之注重其內容。計算機是長於處理「形式」者，是故計算機可以協助做注記常例之比對。至於如何將常例之形式詮釋為內容的關係，則目前恐仍須靠人力而為之了。

如用計算機來處理注記之常例，則處理之規律（即其形式之變化）仍可再減。對於計算機而言，一個版本之文章，其機讀形式就是一字串（如下節說明）。若比較兩版本之異同，實即比照兩字串之同異。因此，以計算機處理字串的指令來看，只要有：增（insert）、刪（delete）、取代（replace）及互換（swap）等四者，就足夠能表示兩字串在其形式上不同之處了。若再能用移（move）或抄（copy）指令將相同之字串自一處移至他處，那麼，對張舜徽所舉之注記常例，除無法區分八與九兩例之外，均可由計算機自動處理了。

肆、不同版本在計算機中之表達

一、文件之字串形式

原始文件 (以D表示，下文仿此) 經打字後即變成該文件之一種字串形式，此字串形式之文件，其文字部份和原始文件相同，然而其文字間或夾有標誌符號，以載明原

【註九】關於科文，在許多流通佛典內均有詮釋，本文仍根據：1)周止菴《般若波羅蜜多心經註》(台北,華藏佛教圖書館印，1992) 2)江味農《金剛經講義》(同前，1983) 3)弘一大師，《心經大意》(台灣，精進念佛會印)；4)寶靜法師輯，《諦閑大師語錄》(台北，新文豐出版公司，1993)第251-269頁，等四書所歸納者。

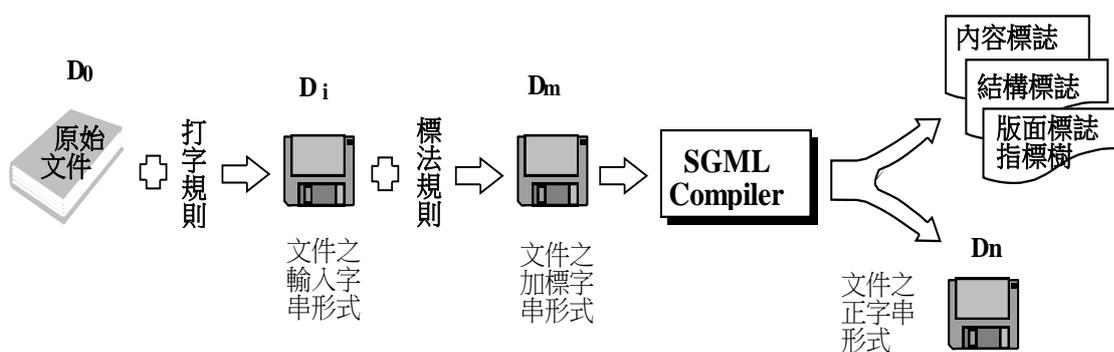
【註十】同【註八】152~154頁

【註十一】同【註八】180~181頁

始文件之一些版面特徵，諸如頁次、夾注等等。換言之，在輸入時，為保留原書之版面訊息，吾人常加入表示版面結構之標誌。此形式之字串稱為原始文件的「輸入字串」(D_i)。

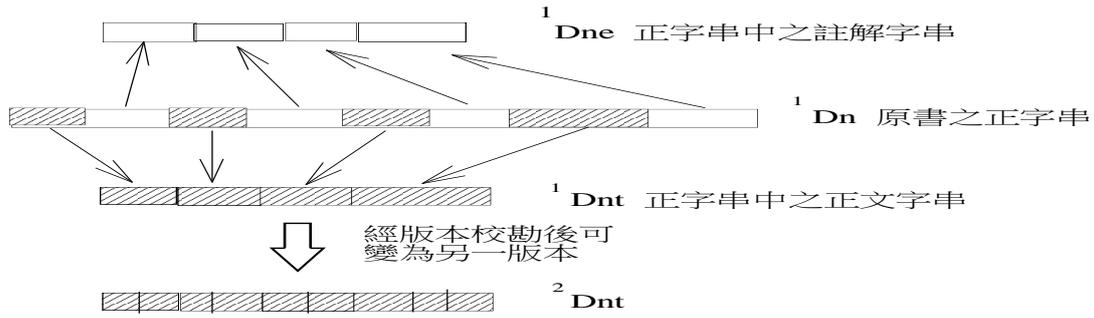
為了使計算機能明了原始文件的內容結構，如目錄或科文等，或是欲使計算機能識別一些與文件內容有關的文句，如檢索詞或句讀等等，便須在 D_i 中加入一些特殊的標誌符號，即加入結構標誌與內容標誌，以資識別【註十二】。加過這些標誌後的文件字串稱為「加標字串」(D_m)，而此 D_m 即可讀入計算機中處理。

設若計算機中有一編譯器 (Compiler) 可以解讀上述之各種標誌，則在編譯器處理過 D_m 後，可產生三種樹狀結構的標誌檔案。這些檔案除保有原標誌的信息外，還有該標誌指向它在 D_n 中位置的指標 (pointer)。而 D_n 則是將 D_m 中所有標誌均剝除後，所得到的字串，稱之為「正字串」。這些關係如圖二所示。



圖二、文件之各種字串形式和文件在計算機中的表達示意圖

在本文所討論之文件，都假設已是 D_n 的形式，而科文的結構則是上述的那種樹狀結構檔案。【註十三】如果 D_n 只是一部佛經的原文，那麼問題就很單純了，我們可以直接探討多版本的表達與彼此間對映的方式。如果 D_n 中夾有注疏、校勘、評論、句讀等等不是本文的註解，則我們可以將該經之原文和註解分開，即分 D_n 為兩個字串 D_{nt} 和 D_{ne} ；其中 D_{nt} 表示 D_n 中之正文部份，即「正文字串」，而 D_{ne} 表示 D_n 中註解部份，即「註解字串」，如圖三所示。由於 D_n ， D_{nt} 或 D_{ne} 的內容或彼此的關係都是不允許更改的，我們仍然可以得到原經文之字串 D_{nt} 去做多版本之表達。故為方便計，下文中我們直接用 D_n 來討論多版本表達的方法。



圖三、經書中正文字串和註解字串之映射關係，斜線部份表示正文，空白部份表詳註解文字

二、版本間的對映

為方便起見，讓我們用《般若波羅蜜多心經》（簡稱《心經》）的版本作例子來說明各版本的表達與彼此間對映的方式。依周止菴的蒐集，《心經》現存共有十五個不同的版本；其中有八譯本由梵譯漢而得，有二譯本為梵文之音譯，有五譯本則是由梵譯為藏，再由藏譯為漢。【註十四】今取鳩摩羅什及玄奘兩譯本為例，原文如表二中所示。

表二：《心經》兩版本原文

鳩摩羅什版

觀世音菩薩。行深般若波羅蜜時。照見五陰空。度一切苦厄。舍利弗。色空故。無惱壞相。受空故。無受相。想空故。無知相。行空故。無作相。識空故。無覺相。何以故。舍利弗。非色異空。非空異色。色即是空。空即是色。受·想·行·識亦復如是。舍利弗。是諸法空相。不生。不滅。不垢。不淨。不增。不減。是空法非過去。非未來。非現在。是故空中無色。無受·想·行·識。無眼·耳·鼻·舌·身·意。無色·聲·香·味·觸·法。無眼界。乃至無意識界。無無明。亦無無明盡。乃至無老死。亦無老死盡。無苦·集·滅·道。無智。亦無得。以無所得故。菩薩依般若波羅蜜故。心無罣礙。無罣礙故。無有恐怖。遠離一切顛倒夢想苦惱。究竟涅槃。三世諸佛。依般若波羅蜜故。得阿耨多羅三藐三菩提。故知般若波羅蜜。是大明咒。是無上明咒。是無等等明咒。能除一切苦。真實不虛。故說般若波羅蜜咒。即說咒曰。揭帝。揭帝。波羅揭帝。波羅僧揭帝。菩提僧莎訶。

玄奘版

觀自在菩薩。行深般若波羅蜜多時。照見五蘊皆空。度一切苦厄。舍利子。色不異空。空不異色。色即是空。空即是色。受·想·行·識·亦復如是。舍利子。是諸法空相。不生。不滅。不垢。不淨。不增。不減。是故空中無色·無受·想·行·識。無眼·耳·鼻·舌·身·意。無色·聲·香·味·觸·法。無眼界。乃至無意識界。無無明。亦無無明盡。乃至無老死。亦無老死盡。無苦·集·滅·道。無智。亦無得。以無所得故。菩提薩埵。依般若波羅蜜多故。心無罣礙。無罣礙故。無有恐怖。遠離顛倒夢想。究竟涅槃。三世諸佛。依般若波羅蜜多故。得阿耨多羅三藐三菩提。故知般若波羅蜜多。是大明咒。是無上明咒。是無等等明咒。能除一切苦。真實不虛。故說般若波羅蜜咒。即說咒曰。揭帝。揭帝。波羅揭帝。波羅僧揭帝。菩提僧莎訶。

- 周止菴蒐集《心經》減版查。
- 一、無罣礙。亦無得。以無所得故。菩提薩埵。依般若波羅蜜多故。心無罣礙。無罣礙故。無有恐怖。遠離顛倒夢想。究竟涅槃。三世諸佛。依般若波羅蜜多故。得阿耨多羅三藐三菩提。故知般若波羅蜜多。是大明咒。是無上明咒。是無等等明咒。能除一切苦。真實不虛。故說般若波羅蜜咒。即說咒曰。揭帝。揭帝。波羅揭帝。波羅僧揭帝。菩提僧莎訶。
6. 般若波羅蜜多心經 燉煌石室本 唐三藏法師沙門法成譯
7. 佛說聖佛母般若波羅蜜多心經 龍藏薄字函又據日本大藏本校正 宋西天譯經三藏施護譯
8. 般若波羅蜜多心經 唐義淨三藏譯
- 二、直取梵音表以漢字，共二版本
9. 梵語心經一卷 日本大藏經本
10. 唐梵翻對字音般若波羅蜜多心經 燉煌石室本
- 三、藏譯漢，共五版本
11. 般若心經 日僧能海寬由藏文譯日文，茲更由日文轉譯
12. 佛說般若波羅蜜多心經 孫慧風由藏文轉譯
13. 般若波羅蜜多心經 釋楚禪由藏文轉譯
14. 般若波羅蜜多心經 釋慧清由藏文轉譯
15. 般若波羅蜜多心經 釋超一由藏文轉譯

在表二中的經文，其句讀符號參夾在正文中，這是為了讀者易於閱讀的關係。在計算機的經文正字串中，並無句讀符號，是故在計字數或算計字的位置時，勿將句讀符號計入。

如果將此二版本分別存在計算機中，而不表明此二版本間對應的關係，我們認為是無意義的，因為計算機在此情況下實難對它們做內容之比對和做任何進一步的處理。所以，讓我們用前述之字串運作指令，將版一本改變為另一版本來觀察此二版本之間的關係。茲將鳩摩羅什版改變為玄奘版的程序列如表三所示：

表三：將鳩摩羅什版《心經》轉換為玄奘版《心經》之程序

- 1、斷詞取代 [鳩<'觀世音菩薩'>·'觀自在菩薩';
鳩<'般若波羅蜜'>·'般若波羅蜜多';
鳩<'舍利弗'>·'舍利子';
鳩<'五陰'>·'五蘊']
- 2、取代 [鳩<18,1>·'皆空';鳩<64,4>·'色不異空';
鳩<68,4>·'空不異色';鳩<86,2>·'菩提薩埵';
鳩<301,3>·'莎婆訶']
- 3、刪 [鳩<24,37>;鳩<108,12>;鳩<210,2>;鳩<216,2>;
鳩<257,1>;鳩<263,1>;
- 4、增 [鳩<249,->·'是大神咒']

茲將此程序之義意說明如後：「斷詞取代」的指令是要求計算機先在文件的正字串中認出某一字串，如例中之'觀世音菩薩'...'五陰'等，並以另一字串，即如'觀自在菩薩'...'五蘊'去分別替入。在此指令中，「鳩」表示鳩摩羅什本《心經》正字串之檔案名稱，如：鳩<'舍利弗'>即表示在鳩摩羅什本《心經》中任何位置上凡是有「舍利弗」字樣者，餘仿此。

「取代」指令是指某正字串中某一位置之字串，以後列之字串取代之。例如：「鳩<86,2>·'菩提薩埵」表示鳩本《心經》中第86個字開始長度為2之字串（即'菩薩'），用'菩提薩埵'取代。而刪〔鳩<24,37>〕則表示鳩本《心經》中第24個字起刪37個字。增〔鳩<249,->·'是大神咒'〕則表示在鳩本《心經》第249個字之前（'-'表示之前，'+'表示之後）加入'是大神咒'四字。

用這種方法，可以十分簡易地將一個版本轉換為另一個版本。是故計算機中若存有一版本之正字串，以及存有將其轉換為另一版本之程序時，則無必要保存第二個版本之正字串。依此類推，則無論有多少版本，只要保存一個版本，並保留這個版本轉換為各其他版本之程序即可。

由於這些版本的文字都是不容改動的，所以版本間換的程序亦可自動地從一個方向的程序產生反方向轉換的程序。例如，有了鳩本轉換為玄奘本的程序，則可自動地得自玄奘本轉換為鳩本的程序。也正因此，若有多版本的情形，各版本間直接轉換的程序，皆可自動獲得【註十五】。

其次，在上例中，字串間的轉變均以詞為單位。這是因為考量到使用上給使用者方便的關係。如：以「舍利子」取代「舍利弗」而不以「子」取代「弗」，又如：以「皆空」取代「空」而不是在「空」之前加入一「皆」字等。這樣做的好處是可以產生詞彙的對照表如「舍利子」之於「舍利弗」，語意的對照表，如「非色異空」之於「色不異空」...等等。這些性質將大有助於文件內容之比對和檢索。

【註十二】張翠玲等，〈文件結構標誌手冊〉，中央研究院資訊科學研究所文件處理研究室第83001號技術報告，1994

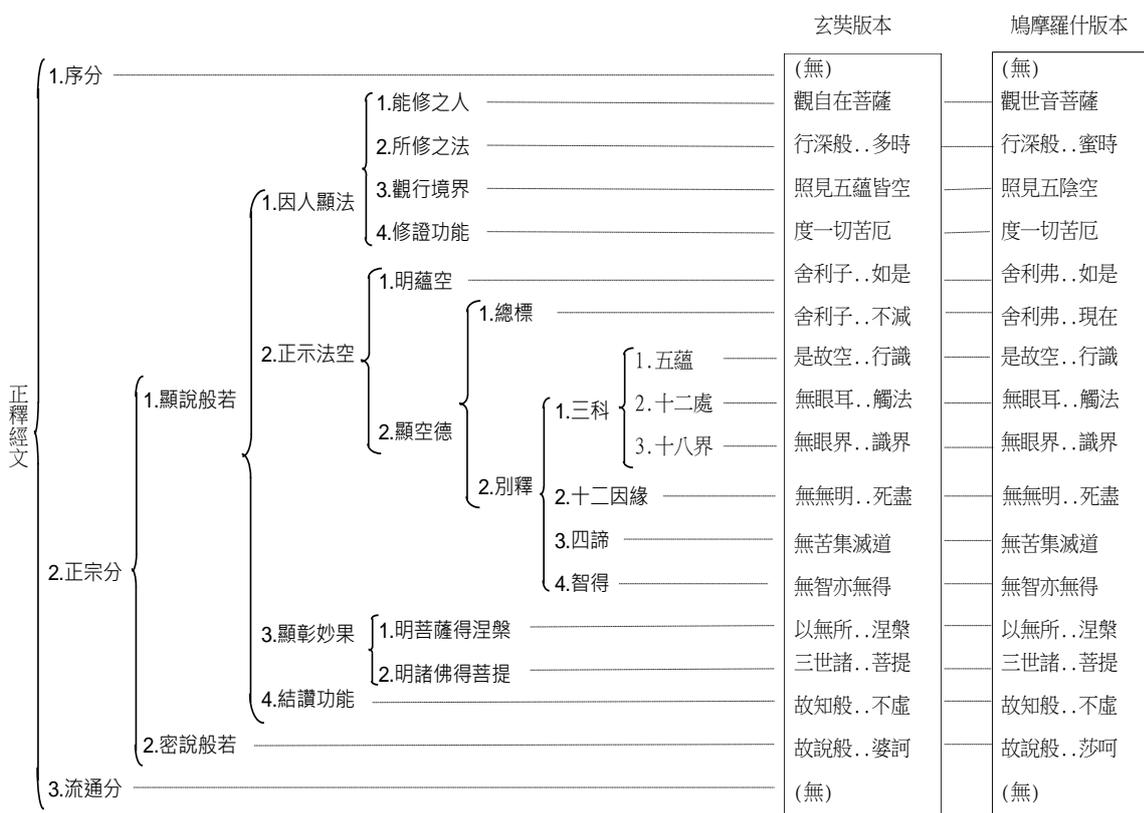
【註十三】關於標誌（markup）部份，可參閱ISO 8879 SGML（1986）。關於文獻之表達方面，可參照謝清俊等〈談古籍之電子版本〉（海峽兩岸中國古籍整理研究現代化技術研討會論文集，北京，中文信息學會，1993）

【註十四】周止菴，〈般若波羅蜜多心經註〉（台北，華藏佛教圖書館印，1992），3-22頁。

【註十五】此證明甚簡單，但因省篇幅故從略。

伍、科文之利用

前文已談到，所有的佛經內容結構可共用圖一中的科文來解析，而不同的佛經其正文不同之處，則主要呈現在「正宗分」之內；然而，若是異版的佛經，則可共用同一個「正宗分」，讓我們用例子來說明。



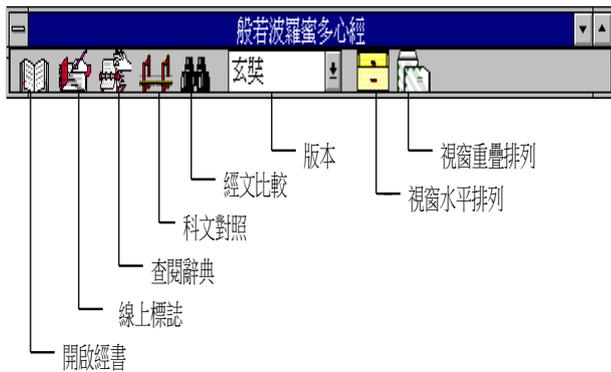
圖四：《心經》的「正宗分」科文

《心經》的「正宗分」科文如圖四所示。在圖四中，正宗分的樹狀結構提要已將鳩本《心經》和玄奘本《心經》剖析得一覽無遺。有了這樣的結構關係，我們立刻可以用個人電腦中 Ms Window 所提供的工具，很方便地做了一個多版本文獻系統的雛型。以鳩本《心經》和玄奘本《心經》為例，其操作情形如圖五中所示。圖五中之各圖例自己均可為自己說明，為省篇幅在此不贅述。然而，在圖例三和四中所顯示的查閱辭典功能，亦可用之於連繫各書之注疏、校勘、評述、句讀或其他參照之文件。這些關係將構成各文獻間一個內容關係的網（即構成 Hypertext，並可提供上述各書中註解文字之相互參照）。由於此部份結構已超越了本文之範圍，故亦從略【十六】。

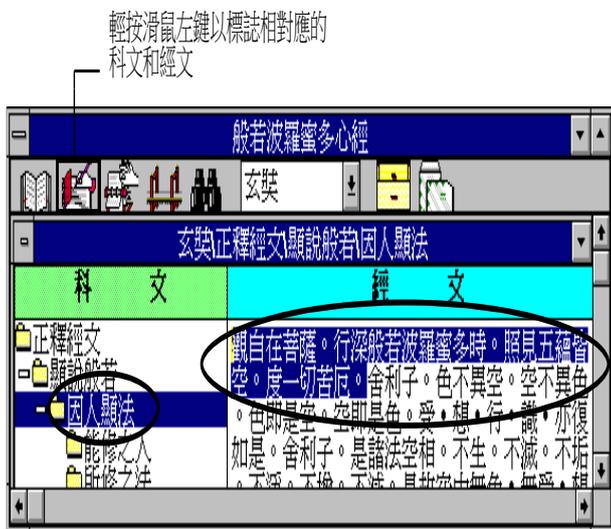
【註十六】此部份可參考：周亞民等《使用SGML與物件導向資料庫轉換古書為 Hypertext》（海峽兩岸中國古籍整理現代技術研討會論文集，北京 中文信息學會，1993）

圖五、科文之利用舉例

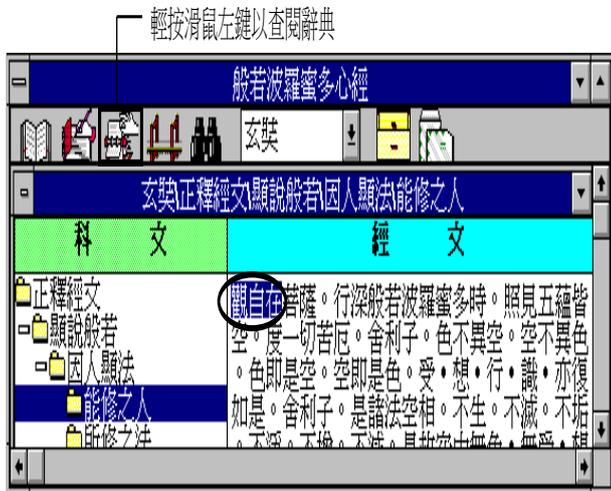
圖例一 功能簡介



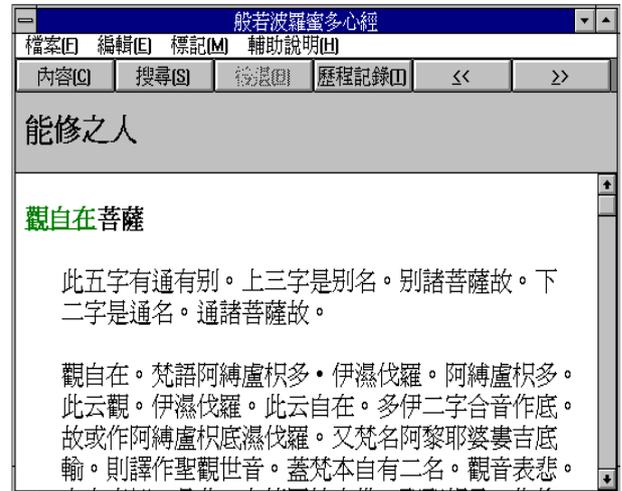
圖例二 線上標誌



圖例三 查閱辭典

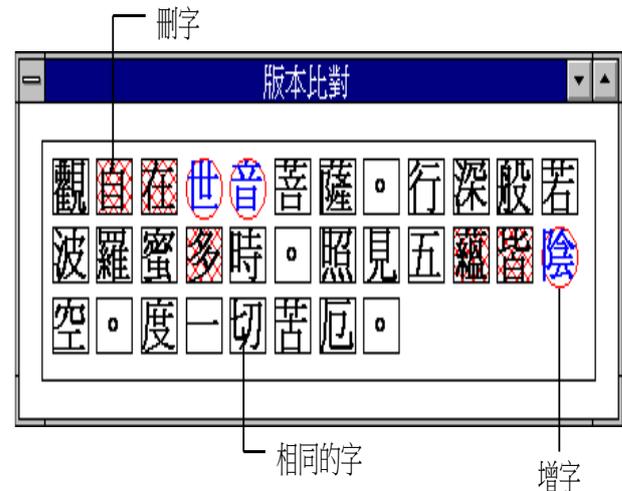


圖例四 辭典內容

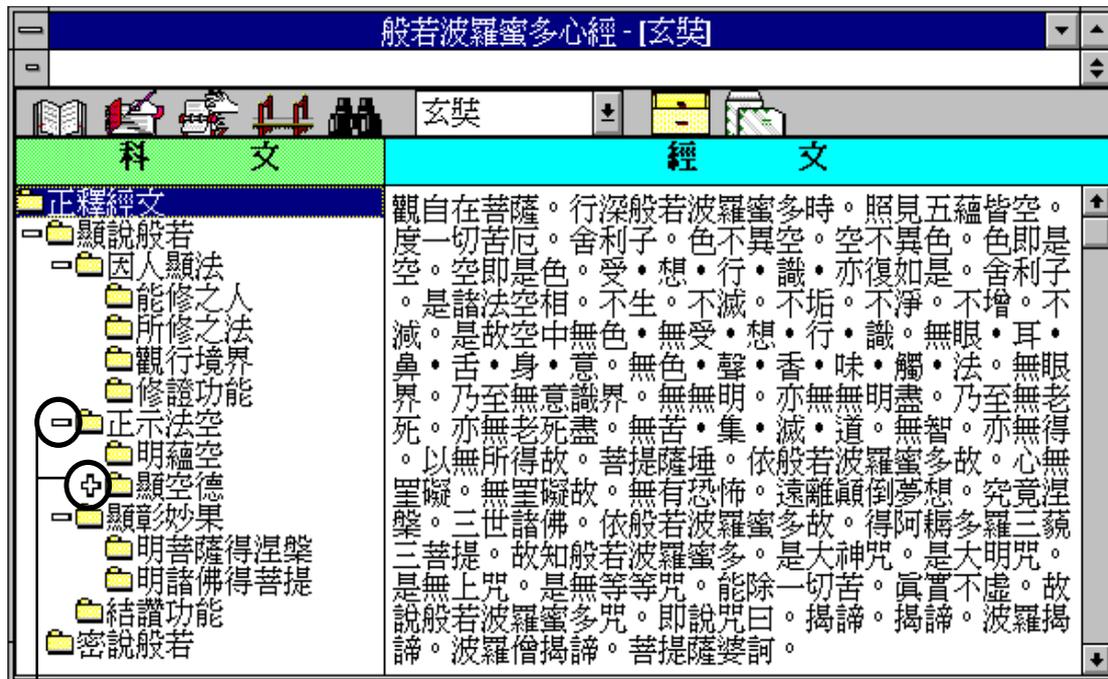


圖例五 經文比較的結果

下圖表示欲將『觀自在菩薩。行深般若波羅蜜多時。照見五蘊皆空。度一切苦厄。』改為『觀世音菩薩。行深般若波羅蜜時。照見五陰皆空。度一五苦厄。』所須刪減及增加的字。



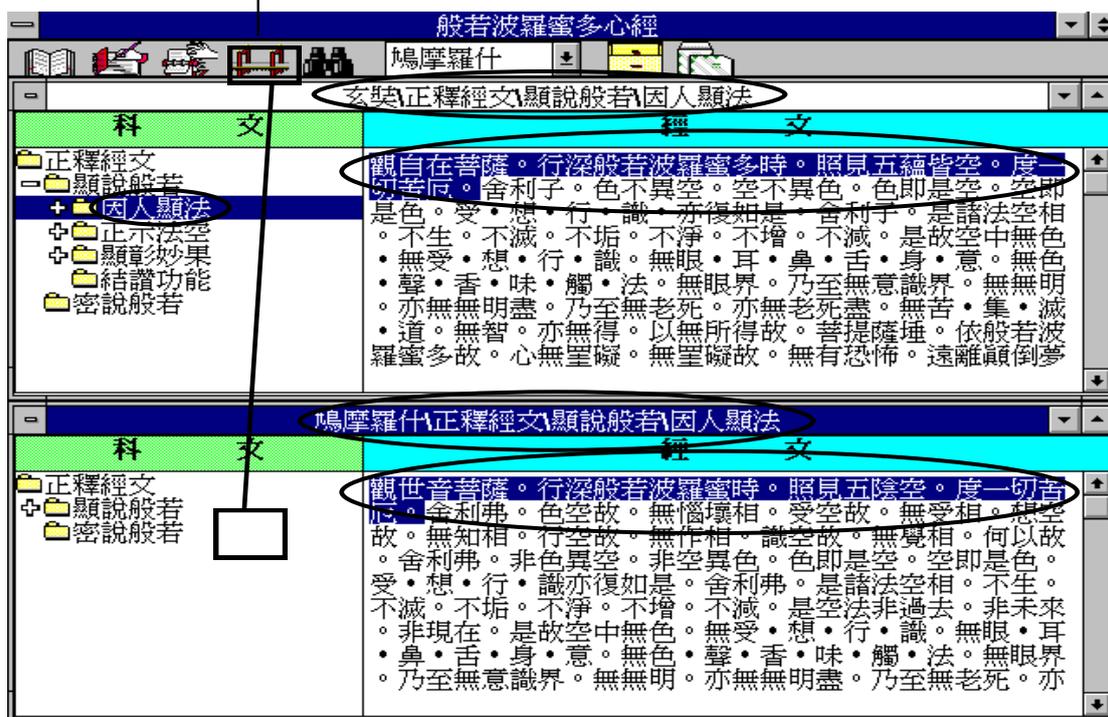
圖例六 顯示與隱藏科文



輕按滑鼠左鍵以顯示(+)或隱藏(-)科文目錄

圖例七 不同版本科文的對照

以滑鼠拖曳圖像到科文處即可作科文對照



陸、文件共享的標準形式

文獻電子化的最大好處，就是可以讓大家都充份共享。若把文獻比作資源或是財富，這種財富是不會因共享而減少的，正是所謂：「己所予人已愈有」。如果大家各自擁有的電子文獻可以都拿出來共同享用，那麼，每個人都會變得極為「富有」，國家整體力量也為之大增。反之，若是每個人死抱著自己有的來圖利，就大大削弱了文獻財富的流通，每個人都因此而「窮困」。所以，做電子文獻，必須要注意到共享的問題。

文獻若要共享，必須要克服一些問題。首先，它要能夠在電子通信網路上流通；其次，他的內容要能夠突破時空的限制，也就是說無論是對認識或不認識的人，無論他在什麼地方，無論他是處於現在或是未來，都要能夠解讀文獻的內容。由於這些緣故，我們必須遵從一些文獻表達的標準來呈現我們的文獻，而這些標準就是 ISO 8879 SGML (Standard Generalized Markup Language , 1986) 和它相關文件所定訂的事。以下，就讓我們用例子來說明如何用 SGML來表達多版本文件及其科文。

SGML是用文件類別定義 (Document Type Definition , 簡稱 DTD) 來描述文件的結構和語意信息。以佛經之 DTD 而言，即前述之科文。至於設計用來描述版本轉換關係的 DTD，我們認為必需考慮接收端可處理性、交換能力和目視的可讀性。版本轉換的關係，可以用 SGML 中所訂定的 Process Instruction、Link Process 加以處理，但是如果接收端不能處理這兩種特別的 SGML 功能，就無法交換和共享。再者，對於版本差異的信息，原可以直接放入正字串的文句之間，但是如此一來正字串變得極不易閱讀。因此我們用取代、增、刪、互換四種標誌描述之，並專為使用人設計如表三的轉換程序，與經文分開，這麼做不僅增加可讀性，更方便版本的管理。標誌後的佛經，經由 SGML 編譯器驗證後即可經由序化 (SDIF Packer) 進行網路上的傳遞和共享。

《心經》的整個科文，包括「總釋名題」和「正釋經文」，寫成 DTD 形式後如圖六所示。圖中英文詞彙和符號全沿用 ISO 8879 之規定，請參照該標準之說明。圖七中 (A) 顯示的是經過 SGML 標誌後之鳩摩羅什本《心經》，而 (B) 是將鳩本轉換為玄奘本的標誌表達方式。圖六和圖七中呈現的 SGML 形式，可從圖五之運作中自動產生。

```
<!DOCTYPE 心經 [  
<!ENTITY % 位置 "起始位置 NUMBER #REQUIRED">  
<!ELEMENT 心經 -- (總釋名題?, 正釋經文, 版本轉換*)>
```

```
<!ATTLIST 心經 版本 CDATA #REQUIRED>
<!ELEMENT 總釋名題 -- (經題, 人題)>
<!ELEMENT 經題 -- (綱要, 融會各家, 五重玄義)
<!ELEMENT (綱要, 融會各家) -- (#PCDATA)>
<!ELEMENT 五重玄義 -- (釋名, 顯體, 明宗, 辨用, 判教)>
<!ELEMENT 釋名 -- (通名, 別名)>
<!ELEMENT (通名, 別名) -- (#PCDATA)>
<!ELEMENT 顯體 -- (正顯體, 辨異同)>
<!ELEMENT (正顯體, 辨異同) -- (#PCDATA)>
<!ELEMENT 明宗 -- (正明宗, 辨異同)>
<!ELEMENT 正明宗 -- (#PCDATA)>
```

圖六 (A) 、《心經》之DTD (待續)

```

<!ELEMENT 辨用 -- (#PCDATA)>
<!ELEMENT 判教相 -- (天台, 賢首五教)>
<!ELEMENT 天台 -- (五時, 四教)>
<!ELEMENT 五時 -- (#PCDATA)>
<!ELEMENT 四教 -- (化法, 化儀)>
<!ELEMENT (化法, 化儀) -- (#PCDATA)>
<!ELEMENT 賢首五教 -- (#PCDATA)>
<!ELEMENT 正釋經文 -- (序分?, 正宗分, 流通分?)>
<!ELEMENT 序分 -- (通序, 別序)>
<!ELEMENT (通序, 別序) -- (#PCDATA)>
<!ELEMENT 正宗分 -- (顯說般若, 密說般若)>
<!ELEMENT 流通分 -- (#PCDATA)>
<!ELEMENT 顯說般若 -- (因人顯法, 正示法空, 顯彰妙果, 結讚功能)>
<!ELEMENT 因人顯法 -- (能修之人, 所修之法, 觀行境界, 修證功能)>
<!ELEMENT (能修之人, 所修之法, 觀行境界, 修證功能) -- (#PCDATA)>
<!ATTLIST (能修之人, 所修之法, 觀行境界, 修證功能) %位置;>
<!ELEMENT 正示法空 -- (明蘊空, 顯空德)>
<!ELEMENT 明蘊空 -- (#PCDATA)>
<!ATTLIST 明蘊空 %位置;>
<!ELEMENT 顯空德 -- (總標, 別釋)>
<!ELEMENT 總標 -- (#PCDATA)>
<!ATTLIST 總標 %位置;>
<!ELEMENT 別釋 -- (三科, 十二因緣, 四諦, 智得)>
<!ELEMENT 三科 -- (五蘊, 十二處, 十八界)>
<!ELEMENT (五蘊, 十二處, 十八界) -- (#PCDATA)>
<!ATTLIST (五蘊, 十二處, 十八界) %位置;>
<!ELEMENT (十二因緣, 四諦, 智得) -- (#PCDATA)>
<!ATTLIST (十二因緣, 四諦, 智得) %位置;>
<!ELEMENT 顯彰妙果 -- (明菩薩得涅槃, 明諸佛得菩提)>
<!ELEMENT 明菩薩得涅槃 -- (#PCDATA)>
<!ELEMENT 明諸佛得菩提 -- (#PCDATA)>
<!ATTLIST (明菩薩得涅槃, 明諸佛得菩提) %位置;>
<!ELEMENT (結讚功能, 密說般若) -- (#PCDATA)>
<!ATTLIST (結讚功能, 密說般若) %位置;>
<!ELEMENT 版本轉換 -- (取代|增|刪|互換)*>
<!ATTLIST 版本轉換 新版本 CDATA #REQUIRED>
<!ELEMENT (取代, 增, 刪, 互換) - O EMPTY>
<!ATTLIST 取代 原文 CDATA #REQUIRED
           取代為 CDATA #REQUIRED>
<!ATTLIST 增 原文 CDATA #REQUIRED
           增添詞句 CDATA #REQUIRED>
<!ATTLIST 刪 原文 CDATA #REQUIRED>
<!ATTLIST 互換 詞句甲 CDATA #REQUIRED
           詞句乙 CDATA #REQUIRED>
<!-- 此 DTD 已經由 ARC-Parser 驗證無誤 -->

```

圖六 (A) 、《心經》之DTD (續完)

圖六 (B) 、《心經》版本轉換之DTD

<心經 版本="鳩摩羅什"><正釋經文><顯說般若><因人顯法><能修之人 起始位置="1">觀世音菩薩。</能修之人><所修之法 起始位置="6">行深般若波羅蜜時。</所修之法><觀行境界 起始位置="14">照見五陰空。</觀行境界><修證功能 起始位置="19">度一切苦厄。</修證功能></因人顯法><正示法空><明蘊空 起始位置="24">舍利弗。色空故。無惱壞相。受空故。無受相。想空故。無知相。行空故。無作相。識空故。無覺相。何以故。舍利弗。非色異空。非空異色。色即是空。空即是色。受。想。行。識亦復如是。</明蘊空><顯空德><總標 起始位置="88">舍利弗。是諸法空相。不生。不滅。不垢。不淨。不增。不減。是空法非過去。非未來。非現在。</總標><別釋><三科><五蘊 起始位置="120">是故空中無色。無受。想。行。識。</五蘊><十二處 起始位置="131">無眼。耳。鼻。舌。身。意。無色。聲。香。味。觸。法。7</十二處><十八界 起始位置="145">無眼界。乃至無意識界。</十八界></三科><十二因緣 起始位置="154">無無明。亦無無

明盡。乃至無老死。亦無老死盡。</十二因緣><四諦 起始位置="172">無苦· 集· 滅· 道。</四諦><智得 起
始位置="177">無智。亦無得。</智得></別釋></顯空德></正示法空><顯彰妙果><明菩薩得涅槃 起始位置
="182">以無所得故。菩薩依般若波羅蜜故。心無罣礙。無罣礙故。無有恐怖。遠離一切顛倒夢

圖七 (A) 、經 SGML 標誌後之鳩摩羅什版《心經》 (待續)

想苦惱。究竟涅槃。〈明菩薩得涅槃〉明諸佛得菩提 起始位置="222">三世諸佛。依般若波羅蜜故。得阿耨多羅三藐三菩提。〈明諸佛得菩提〉顯彰妙果〈結讚功能 起始位置="243">故知般若波羅蜜。是大明咒。是無上明咒。是無等等明咒。能除一切苦。真實不虛。〈結讚功能〉顯說般若密說般若 起始位置="274">故說般若波羅蜜咒。即說咒曰。揭帝。揭帝。波羅揭帝。波羅僧揭帝。菩提僧莎呵。〈密說般若〉正釋經文

圖七 (A)、經 SGML 標誌後之鳩摩羅什版《心經》(續完)

<版本轉換 新版本="玄奘">取代 原文="觀世音菩薩" 取代為="觀自在菩薩">取代 原文="般若波羅蜜" 取代為="般若波羅蜜多">取代 原文="舍利弗" 取代為="舍利子">取代 原文="五陰" 取代為="五蘊">取代 原文="64,4" 取代為="色不異空">取代 原文="68,4" 取代為="空不異色">取代 原文="86,2" 取代為="菩提薩埵">取代 原文="301,3" 取代為="莎婆訶">增 原文="18-" 增添詞句="皆">增 原文="249-" 增添詞句="是大神咒">刪 原文="24,37">刪 原文="108,12">刪 原文="210,2">刪 原文="257">刪 原文="263">版本轉換</心經>

圖七 (B)、從鳩摩羅什本轉換 (或對照) 到玄奘本之標誌表示

柒、結語

本文呈現的，是一種處理古籍多版本問題的方法。我們用的例子雖然是佛經，但是只要古籍有一較詳盡的目錄，或輟其主題、眉批、評語……等於原目錄之內，即可構成一類似科文的結構，而即適用於本文所提出之系統。

在佛經的科文中，我們預備在總釋名題之下，在經題和人題旁，另加一「書題」。此舉可將現代書目學「機讀書目」之部份納入，以使系統有整合現代圖書資料之功能。【註十七】。

由於本研究計劃還在進行中，計算機多版本文件的資料構尚未完全定案，所以在本文中未作詳細之介紹。這部份留待以後再撰文報告。然而，多版本文件資料結構的設計考量，在本文已有詳細的討論和舉例，可資讀者指正。

本研究用以為測試的文獻，除佛經外，還有《文心雕龍》的各版本【註十八】。《文心雕龍》的注疏和引徵的文件非常多，正好用來做古籍多版 Hypertext 結構之測試，而這一些研究也有待日後補述。

誌 謝

本文承張翠玲小姐打字排版及莊德明先生協助，尤其是在匆忙中趕稿完成，仍保有一定的高品質，十分難能可貴，謹在此致謝。

【註十七】中央圖書館，《中國機讀編目格式》，第三版，台北，1990

【註十八】同【註十三】