



從回溯建檔談日文圖書編目

鄭玉玲 國家圖書館編目組編輯

陳慧文 國家圖書館編目組書記

一、前言

回溯建檔，又稱回溯資料轉換，或回溯轉換（Retrospective Conversion），亦即在圖書館自動化過程中，根據各圖書館政策與作業標準將一館已有之人工書目紀錄轉換成機器可讀形式。一般圖書館回溯書目資料轉換主要針對三類書目資料：書目紀錄、權威紀錄、館藏與館藏地紀錄。（註1）本館中西文館藏早在自動化初期即已完成回溯建檔工作。

日文資料部分，本館曾在民國74年針對中國機讀編目格式對於日韓文資料建檔之適用性加以測試（註2），並於民國78年至79年間，採用傳技電腦公司所發展之中華壹號III型中英日韓文終端機，進行日韓文卡片之回溯建檔。然而，這些資料於本館系統更換之際，日韓文資料從傳技系統轉至URICA系統時，由於URICA系統轉檔程式係以登錄號作為比對條件，而日文資料和中西文資料登錄號重複情形嚴重，造成大量資料遭覆蓋。當時本館緊急委請鼎盛公司設法將原已轉入之中西文資料還原，另將日韓文資料刪除後，以不轉入登錄號之方式，全數重轉，此仍導致部分資料消失。時隔多年，其間本館閱覽組日韓文室繼續以卡片方式編目，並進行館藏清點、登錄號重新給號及重貼條碼。編目組於民國91年1月起，全面接手負責日韓文圖書資料之編目業務，首先面臨的任務便是既存電腦書目資料的整理和卡片回溯建檔。

二、計畫過程

為配合書目資料之特性，本計畫將日文回溯建檔分為兩階段進行。作業時間自民國92年6月至93年12月止，約一年半左右，共補登及回溯建檔37,500多筆日文書目資料（約40,560張卡片）。

（一）電腦紀錄之補全

早期本館日韓文書目在URICA系統中，僅存書目而未載登錄號。有鑒於此，日文回溯建檔必須先針對既存於系統中之書目資料進行查核、修改、及登錄號補鍵。作業程序大致如下：

1. 以國際標準書號、書名、著者或索書號查詢系統中是否已有此筆資料。
2. 若系統中已有此筆資料，則核校系統內之資料是否與卡片內容符合，是否有資料漏失、亂碼、空字、舛誤之情形，若有則予以修正。書目記錄有欄位540者，表示曾有資料覆蓋之情形，將欄位540刪除，並增補其他欄位資料。
3. 檢查欄位681與984之類號是否相同，若不同，則根據中國圖書分類法判斷後，予以修改。
4. 若缺欄位984館藏地及索書號，則需自行新增。
5. 確認書目資料無誤後，於欄位985加入登錄號及年代部冊號。
6. 若有「登錄號重號」或「有卡無書」之情形，先集中登記，交給日韓文室。
7. 若確認系統內無此筆資料，則用螢光筆將書名劃起來，留待第二階段處理。

（二）卡片目錄之建檔

回溯建檔途徑包括：轉錄自書目資訊中心、委託廠商轉錄或鍵入、整批自行鍵入或轉錄、隨借隨鍵等方式。本館於作業之初亦曾考慮從國際圖書館電腦中心（Online Computer Library Center，簡稱OCLC）、日本國立情報學研究所（National Institute of Informatics，簡稱NII）的NACSIS-CAT，或日本國會圖書館光碟系統進行抄錄編目，然而這些系統所採用之機讀編目格式與內碼均與目前本館使用之系統不同，資料轉入前，需先請廠商



開發機讀編目格式及內碼轉換之程式。且即使轉入本館系統之後，由於使用之分類系統不同，本館仍須重新再進行分類及權威控制。在時間、費用、及效率的權衡之下，決定採用委外鍵入之方式來進行回溯建檔。

本館於合約生效日送交廠商待回溯建檔A3大小之卡片影印紙3,381張（約40,560張卡片），合約中約定完全依照中國編目規則及其他本館作業規範（註3），將卡片資料轉換成中國機讀編目格式對應欄位，直接鍵入本館電腦系統編目模組中（如圖一、圖二）。館藏地點配合日韓文室提供之典藏分配。類號861之圖書指定為「日韓文庫」，參考書指定為「日韓閱覽」，其他圖書及參考書複本指定為「日韓書庫」，若需由日韓文室辦公室處理或一時無法如上分配之資料指定為「日韓文室」。

圖一 日文機讀格式驗證單（一）

圖二 日文機讀格式驗證單（二）

三、問題探討

回溯建檔對於編目人員向來便是一大挑戰。過程雖不似新書編目從無到有，而是將已編好之目錄卡片轉換成機讀資料，然而由於語文之特殊性及館藏之複雜性，建檔過程中，館藏資料的潛藏問題均會在此時一一浮現。茲將遭遇的問題彙列如下：

（一）系統轉換過程中整筆或部分資料消失

如前言所述，日文書目資料自傳技系統轉至URICA系統時發生資料錯蓋及漏失之現象，因此在回溯建檔過程中無可避免必須先進行書目記錄查核的程序，而無法以卡片內容是否為電腦產生，來判斷檔內是否已有此筆記錄。

此外，由於字碼對應上的問題，書目資料自傳技系統轉至URICA系統時，有些字碼無法順利轉入，形成書目資料中缺字而出現方框之情形。例如：出版社的「社」，及日文假名延長音「-」。這些字在書目記錄中出現次數頻繁，因此，在第一階段書目記錄查核及登錄號補鍵的同時，需針對每筆記錄中缺字及誤轉之情形逐筆修改。資料轉換是自動化系統更換是否成功的重要指標之一（註4），因此，在系統更換之前，電腦資訊單位及編目組應互相配合，審慎地針對新舊系統功能（尤其是輸入輸出的內容和比對條件）、檔案格式、及字碼等進行詳細研究和評估，不同的格式和不同的字碼間，應及早編製對應關係表。即使是相同的機讀格式和字碼，也可能因為各家廠商作法不同而造成部分差異。在格式或字碼上的部分差異，經過資料轉檔後，便可能造成上千上萬筆資料的誤失。有些誤失可以批次補救，有些錯誤則必須人工逐筆查核修改，耗費時日。本次計畫的過程中，我們面臨及解決了上一次系統更換時所造成的問題，也再次確認了系統更換前相關評估及準備工作的重要性。

（二）日文漢字字形選取的抉擇

日文漢字字形與我們平常習用的中國文字，原本就有相當的出入。因此，在字形選擇上需特別留意。有些字從表面上看起來類似，但是在點撇之間便有不同。例如：「富」與「冨」，「鄉」與「郷」，「国」與「國」……等，類似情形不勝枚



舉。由於在人工的卡片影印紙上各種字形都有，回溯建檔的過程中未能見到原書，僅能暫時以卡片影印紙上的字形為準，牽涉人名及團體權威記錄部分，才全部改為日文漢字字體。日後若發現書上字形與卡片不符時，再予以更改。

（三）日文漢字同字異碼選取的抉擇

在中文資訊交換碼中，同字異碼的字有1,325組（註5），亦即在螢幕上出現的相同字體實為不同的字碼。尤其日文漢字多位於異體字區，日文書目資料輸入的同時，更容易發生此種困擾，不但編目人員無所適從，同時也會影響複本查核及讀者查詢的結果。例如：「沢」的CCCII碼為4B4921及22473A；「鉄」的CCCII碼為335E3D及234322。為配合本館WEBPAC簡異體字係以還原前2位碼全部轉為繁體字顯示之方式，本組於面臨日文同字異碼情形時，仍以選取異體字為主，以便於目前WEB BIG-5環境中全部轉為繁體顯示。簡繁體對照查詢問題未解決前，日文漢字在本館館藏WEBPAC中，仍只可顯示而無法查詢。全國圖書資訊網中之日文漢字，則因INNOPACQ系統已開發簡繁體對照查詢功能，而得以被查尋。未來系統轉換至Unicode的環境時，由於Unicode編碼原則並未如CCCII具備層屬架構，因此圖書館更必須留意委請廠商開發簡繁體對照查詢功能，以解決日文及大陸圖書書目資料之查詢問題。

（四）日文著者標目權威控制之考量

權威控制的目的在維護標目的一致性，並建立其間的參照關係，以確保讀者查詢的完整性。日本個人著者的筆名、本名、改名、別號，及機關團體名稱的改名、合併、分衍、裁撤等情形，除了根據書上所提供之著者資料外，大部分必須查核相關的工具書及NII或日本國會圖書館的網站資源。尤其是日本政府於平成13年左右經過組織改造後，政府機關名稱多所變動，有些甚至必須至該機構網站上查詢，方能分辨其間的關係。（註6）在本次回溯建檔過程中，由於計畫進度的關係，僅能先完全依照卡片上的標目鍵入，日後再繼續進行查核修改。另西洋人名在卡片中經常僅著錄假名而未載西

文原名，回溯建檔時僅能至NII或日本國會圖書館網站查詢原名，再透過ITSMARC網站確認標目形式，並與館內權威檔對應。若無法於上述途徑查得資料者，西洋人名僅能以假名形式著錄。

日本著者權威控制除如上所述需確實查核其名稱內容外，最需留意的問題之一仍是因其漢字字體所衍生的分歧情形。同一個著者名稱在日文圖書文獻中，有可能使用日文漢字，也可能使用繁體，在中譯日文書中亦然。因此，在名稱權威字體的選擇上，不得不謹慎為之。日文新書編目時，除非在編文獻與權威檔內均為繁體，才使用繁體的權威記錄。否則權威記錄中一律使用日文漢字，書目部分則照錄，並視需要，建立繁體名稱標目的參照關係卡片，本組將陸續於系統中鍵入這些參照關係，以便在系統提供簡繁體對照查詢功能前，讀者和館員得以著者之中文繁體查到該資料。回溯建檔時，由於未能見到原編文獻，卡片上的字體亦不一致，所以日文著者標目只能先完全將其以日文漢字鍵入，本組日後仍須陸續進行其間參照關係的建立。

（五）「登錄號重覆」情形之處理

URICA系統中採取「登錄號唯一」之原則，遇有登錄號重覆時便無法鍵入。本館於數年前系統更換後，雖曾更換重覆之登錄號及重貼條碼，然此次回溯建檔過程中，仍遭遇一些登錄號重號而無法鍵入之情形，目前書目先行鍵入，登錄號部分則先加以集中註記，俟重新給號後再行鍵入。

（六）「一書多編」情形之處理

回溯建檔過程中，以書名查詢，發現同一本書分入不同類號之情形。此種狀況雖不致影響讀者查詢結果，但以「書目集中」原則來看，似乎應合併處理為宜。目前均先行鍵入，另以書面集中註記，俟本組繼續進行改編處理。

（七）「卡片資料不完整」情形之處理

回溯建檔過程中，查核卡片與書目檔是否相同時，館員必須詳細比對卡片中的每一處資訊是否符合，否則容易因卡片資料的錯誤、不全或版本資料的疏忽，將索書號相同而事實上不同的兩筆資料



併在一起。另早期日文手寫卡片，遇有連續性出版品或持續出版之套書時，起迄出版年及卷次常以鉛筆書寫或是空白，必須先加以判斷查核。且連續性出版品常因冊數眾多而分佈於多張卡片中，若卡片因年代久遠而不全時，常無法完整鍵入其卷期冊數，需進一步查核之後，方能鍵入。

四、結 論

自1983年美國研究圖書館組織（Research Libraries Group）宣布以電腦處理及檢索中日韓三種語文資料後（註6），北美及全球各地即陸續進行CJK資料的處理。隨著Unicode的出現，多種語文並存的系統更成為圖書資訊系統發展的主流。OCLC亦預定於2005年7月1日將其原先以亞東字集交換碼（East Asian Character Code；簡稱EACC）為基礎的CJK系統，轉換合併至融合多國語文的Connexion系統，並積極開發符合MARC21及Unicode 8-bit的轉出功能。（註7）本次回溯建檔的完成，無疑地為本館的多種語文資料查詢，植入更穩固的基礎！

註 釋

1. 王梅玲，〈回溯資料轉換〉，胡述兆編，《圖書資訊學大辭典》，（臺北市：漢美，1995），頁621。
2. 胡歐蘭，〈由中國機讀編目格式探討東亞國家機讀編目系統建立之可行性〉，《中國圖書館學會會報》，37期（1985），頁25-35。
3. 包括：〈國家圖書館編目組日文文獻編目要點〉、〈國家圖書館編目組「日文書」人名標目著錄原則、日本作者人名之著錄、權威紀錄與人名檢索款目關係一覽表〉、〈中日文編目輸入簡繁體注意事項〉及本館編目組其他作業規範等。
4. Qin Zhu, "Two critical stages for a successful ILS migration: system profiling and data conversion," Computers in Libraries 24:3 (Mar 2004), pp.26.
5. 許靜芬，〈Unicode 工作小組93年度工作成

果〉，《國家圖書館編目園地電子報》，第47期（2005.2.1）。

6. 日本政府組織再造，日本中央省廳部分進行改革如下：

郵政省、自治省、總務廳（合併為）：總務省
文部省、科學技術廳（合併為）：文部科學省
厚生省、勞動省（合併為）：厚生勞動省
運輸省、建設省、北海道開發廳、國土廳（合併為）：國土交通省
總理府下之金融再生委員會、經濟企劃廳、沖繩開發廳（被併入）：內閣府
環境廳（升格為）：環境省
大藏省（改名為）：財務省
通商產業省（改名為）：經濟產業省

7. 同註2，頁28。

8. "OCLC CJK Software", Retrieved May 20, 2003, from <http://www.oclc.org/oclc/menu/cjk.htm>.
"Cataloging migration to Connexion", Retrieved Feb. 2, 2005, from <http://www.oclc.org/connexion/migrating/default.htm>.
"Will Connexion support Unicode export?", Retrieved Feb. 2, 2005, from <http://www.oclc.org/support/questions/connexion/question19.htm>.

後記：感謝計畫過程中本館閱覽組日韓文室的配合，以及編目組同人的協助。