

評《中文圖書標題表》 —兼談標題表的敘詞化改進 On the Chinese Subject Headings

侯 漢 清
Hang-qing Hou

南京農業大學信息管理系教授

Professor

Department of Library & Information Science
Nanjing Agricultural University

【摘要 Abstract】

《中文圖書標題表》是規模最大的一部中文標題表，具有結構完整、選詞精當、宜於擴充、實用性強等優點。本文選用等同率、關聯比、參照度等指標對這部詞表的性能進行了定量測評，指出此表存在著設置入口詞及參照偏少、詞間關係顯示不充分等缺點。論文提出：為了滿足文獻標引和檢索的要求，適應檢索語言的發展趨勢，應當在大幅度增加詞量的基礎上，將這部標題表改造為一部能夠適應計算機檢索的敘詞表或分類主題表。本文還提出了實行敘詞化改造的具體步驟和方法。

The Chinese Subject Headings (CSH) is one of the largest lists of subject headings developed in China and is recommended as the standard for library & Information work. This article evaluates the quality of CSH based on equivalence ratio, connectedness ratio, accessibility measure. Quantitative methods are used. Suggestion is given for the reformation from a list of subject headings to a thesaurus.

關鍵詞 Keyword

標題表 質量評價 詞彙控制 敘詞表

Subject headings ; Quality evaluaton ; Vocabulary control ; Thesaurus



1984年台灣中央圖書館編目組編製出版了中文圖書標題總目初稿，經過10年的使用，1995年出版其修訂版，並更名爲中文圖書標題表。這是中國圖書館事業發展史及資訊檢索語言發展史上的一件大事。早在半個世紀以前，我國著名學者何多源出版了第一部中文標題表—標題表（1933年）。此後呂紹虞、沈祖榮、程長源等學者相繼編製、出版了多部中文圖書標題表^①，但是它們基本上都未能付諸實際應用。因此可以說，台灣這部標題表是規模最大、用戶最多的一部綜合性中文標題表。現在已經成爲圖書館及資訊工作的規範標準之一，對中文主題目錄編製、書目資料庫建設以及圖書館自動化產生了重要的影響。

這部詞表主要採用美國國會圖書館標題表（LCSH）的詞表技術，以中國圖書分類法爲基礎而編成，因此具有結構完整、選詞精當、宜於擴充、擁有文獻保證、便於實現分類標引與主題標引數據轉換等多項優點，是一部實用性強的標題表。但是，無庸諱言，它也存在著一些缺點和急待解決的問題。現對這部詞表的詞彙性能作一些定量的測評，在此基礎上擬提出一些不成熟的改進詞表的建議，以就正於海峽兩岸的同行。

一、《中文圖書標題表》詞彙性能的測評

中文圖書標題表除了序論（包括編制、使用說明）以外，還包括三個組成部分：1.字順標題表（即標題主表）；2.分類標題表（即分類索引）；3.通用複分標題表（即一般複分標題）。其中第2部分是一般標題表，包括LCSH、西爾氏標題表（Sear's List of Subject Headings）等所沒有的。它將表中的全部標題詞按照賴永祥修訂的中國圖書分類法（第七版）的體系予以排列，相當於敘詞表中的範疇索引，從而提供

分類和字順兩種互補的顯示方式。用戶可以從分類的途徑查詞選詞，也便於用戶把分類標引的結果轉換成主題標引的結果。分類標題表的配置是中文圖書標題表宏觀結構的一大特點，大大增強了這部詞表的兼容性和實用性。

標題表的性能不僅取決於它的宏觀結構，而且取決於它的微觀結構，即標題詞的組織及詞間關係的顯示。現以美國學者蘭開斯特（Lancaster, F. W.）在資訊檢索詞彙控制^②一書中總結的三項測試指標，對中文圖書標題表的詞彙性能進行初步的測試和評價。

1.詞彙等同率（Equivalence ratio）

等同率是指一部標題表中非標題詞（表中△號後的詞）與標題詞（表中“見”後的詞）的比率。前者是不能用於標引的詞，只能充當檢索的入口。後者是標引中正式採用的詞，詞彙等同率是衡量詞表中入口詞豐富程度的一項重要指標。中文標題表的標題主表中共有標題詞7620個（其中包括一級標題詞6550個，二級標題詞646個，三級標題詞424個）各級標題詞均設有非標題詞（即同義詞或入口詞）1035個，這樣，

$$\text{等同率} = 1035 \div 7620 = 0.136$$

據筆者對205種英文詞表的統計，平均等同率均爲0.406。大陸29部詞表的平均等同率爲0.167。與其他詞表相比，中文圖書標題表的等同率是比較低的。

2.關聯比（Connectedness ratio）

關聯比是指一部標題表中設有交互參照標題款目數（即至少具有一個以上的參見參照或反參見參照的標題款目）與標題表中標題款目總數的比率；或者是指標題表中標題款目總數



減去無關聯詞款目（即不與其他標題發生任何聯繫的詞，既無參見參照，也無反參見參照）的總數與標題款目總數的比率。也就是說，詞表中關聯詞數目越多，關聯比就越大。中文圖書標題表共有 6550 個標題款目（複分標題未計入），其中無關聯詞款目多達 5717 個，因此，

$$\text{關聯比} = (6550 - 5717) / 6550 = 0.127$$

也就是說，全表中僅有八分之一的標題款目帶有參見或反參見參照，其餘八分之七的標題款目與任何其他標題詞都不發生聯繫，可見此表中標題之間關聯性之差。

3. 參照度 (Accessibility measure)

參照度是指一部詞表中每個標題詞所配備參見及反參見參照的平均數。這和關聯比一樣，也是測試詞表中設置交互參照數量，顯示詞表中標題詞之間聯繫密切程度的指標。參照度數值越高，意味著標題詞之間的聯繫越密切，用戶查詞選詞就越方便。中文圖書標題表標題主表中共有 492 個參見參照，561 個反參見參照（另有 60 個說明參照未計入），因此，

$$\text{參照度} = (492 + 561) / 7620 = 0.138$$

也就是說，平均每七個標題詞才配備一條參見或反參見參照，可見此表中交互參照設置之少。

表一：《中文圖書標題表》與 LCSH 等綜合性詞表的比較

詞表名稱	標題詞 (敘詞)	標題詞 (非敘詞)	等同率	關聯比	參照度
LCSH (11 版)	162750	133200	0.818	0.799	1.032
MeSH (1967 年版)	--	--	0.50	0.57	0.381
漢語主題詞表 (1 版)	91158	17410	0.165	0.80	> 2
中文圖書標題表	7620	1035	0.136	0.127	0.138

表一是把中文圖書標題表與國內外著名的綜合性標題表或敘詞表作一個比較。從中可以看出，中文圖書標題表的前述三項指標都是比較低的，說明此表入口詞少，設置參照偏少，詞間關係的顯示很不充分，這一切都嚴重影響這部詞表性能和使用。詞彙性能較差的根本原因不在於編制上的疏漏，而是由於傳統的標題表對詞彙的控制（包括詞形控制、詞義控制、詞間關係控制）和顯示還不如後來居上的敘詞表。

二、標題表或敘詞表的選擇

進入電腦時代，圖書館及資訊機構在選擇

資訊檢索語言時面臨著第一個重大選擇是採用受控制語言（標題表、敘詞表等），還是採用自然語言（如關鍵詞、自由詞）。在圖書館及資訊界中，一種看法認為：「自然語言將變成資訊檢索的規範，普通受控敘詞表的使用將會衰退」^③，認為用對題名、摘要（或內容提要）、目次、甚至正文等部分中的關鍵詞的檢索可以代替主題詞的標引；^④另一種看法則認為：「自然語言要全面勝過人工語言是不可能的」，「不管今後自然語言系統如何發展，對詞彙的控制將永遠不會取消」。因此，自然語言與控制的人工語言二者的結合將是最好的。^⑤台灣圖書館界同行傾向於後者，在規劃圖書館自動化的



藍圖時，決定編製一部屬於人工語言的控制詞表，顯然是符合當前資訊檢索理論研究及實踐的進展。

面臨的第二個重大選擇是採用標題表，還是採用敘詞表。中文圖書標題表的使用說明介紹台灣同行選擇先組式的標題表，主要是考慮到標題表編製比敘詞表簡單，而且不會像敘詞表那樣容易出現誤檢。應該指出，這兩點考慮是不全面、不準確的。

首先，就實際情況來看，標題表似乎比敘詞表更難以編製。因為在收詞、選詞、審詞、編配參照、排序、試標引這些步驟上，二者的工作難度是不相上下的。但是標題表的編製還多了一個步驟，即必須事先考慮好複分標題的選用、搭配及次序等，必須擬定由主標題、副標題、次副標題等組成的標題詞串。敘詞表編製時則無需考慮這些問題，可以留待將來由標引人員自行確定。70年代以來，敘詞表的編製有了眾多的專門軟體和工具，可以實現計算機輔助編製，編製分類主題一體化詞表時甚至可以自動生成字順敘詞表、分類敘詞表及各種索引，大大降低了編製難度，提高了編製速度，這些都是標題表無法相比的。

其次，標題表確實「不易出現誤檢現象」，這是因為它屬於先組式語言，「標題詞的直觀性和含義的明確性都勝過敘詞的組配標識」。^⑥但是，標題法為此也付出了重大的代價，諸如，它不能像敘詞法那樣實行自由組配，因而無法實現按任何主題特徵自由檢索，大大降低標引能力，其標題詞的數量卻會一年一年地增長（LCSH現在每年出一新版，每版都要增加八、九千個新詞），給標題表的使用帶來不便。標題表不易誤檢雖然會提高準確率，但是卻會出現漏檢現象，降低檢全率。至於敘詞組配易出現誤檢缺點，現在已經找出了一些可以有效地

減少誤檢的對策，諸如：適當增加先組詞和引入自然語言（專有名詞等）；對於多主題的文獻實行分組或使用關聯符號，以避免虛假聯繫；制訂防止字面組配及二義組配的標引規則等。

第三，在目前的網路環境中，究竟選擇那種檢索語言，不應該只考慮人工條件下編製圖書主題目錄需要，應該全面考慮資訊檢索系統的建立以及網路信息資源的開發。也就是說，要同時考慮人工檢索和電腦檢索、人工檢索系統與電腦檢索系統的接口、書目資料庫的建立、圖書和報刊資料的索引、索引資料在不同資料庫之間或不同網站之間的兼容和互換。這樣，最好的選擇應該是採用敘詞語言。

在國外，尤其是英語世界的一些國家，長期以來一直使用 LCSH，所以他們在廣泛採用敘詞表標引文獻的同時，還繼續保留了用 LCSH 標示圖書的傳統。近年來，國外學者發現，使用公用目錄（OPAC），最常見的問題是難以用 LCSH 擬定檢索策略，難以選擇確切的標題。用 LCSH 檢索要求所選標題詞（包括主、副標題）在數量、詞序、句法及拼寫等方面都要一點不差，否則會造成檢索失敗。

我國圖書館及資訊機構的情況不同於國外，以前沒有主題索引的傳統，因此，應該從我國的國情出發，採用敘詞語言來統一標引各類文獻，包括圖書、期刊、報紙等。這樣就不會像西方一些國家那樣，圖書及其他文獻的標引出現「兩張皮」的現象，從而可以節省大量的人力和物力。

敘詞表既可當作後組式語言使用，也可當作先組式語言，即當作標題表的代用品使用。目前建立書目資料庫的通常做法是：為一篇文獻標引若干敘詞（可以多到十餘個），輸入電腦後即可建成後組式檢索系統，便於用戶進行



檢索。然後再從中選用幾個敘詞組配拼接成一、二個先組式標題，用於編製人工用的主題目錄或主題索引。

從檢索語言的發展歷史來看，敘詞語言將標題語言、元詞語言及關鍵詞語言等的多種優點熔於一爐，因而能夠「青出於藍勝於藍」，這已經成為人們的共識。二十世紀60年代以來，敘詞語言取代標題語言及元詞語言已經成爲一種發展趨勢。世界上最爲著名的標題表 LCSH 從 1986 年底起採用了敘詞表的款目格式和參照符號。有人批評它是「偽裝成敘詞表」^⑦，不管怎樣，LCSH 已經邁開了步伐，朝著敘詞表的方向走去。又如，Mesh（醫學標題表）和 SHE（工程標題表）爲了適應電腦檢索，早已經改變其傳統的標題表結構，現已經演變成兩部真正的敘詞表。

總而言之，無論從資訊檢索語言發展趨勢，還是從中國的國情及建立網路資訊檢索系統的需要來看，明智的選擇應該是採用敘詞語言（當然還應該與自然語言結合使用）。

三、標題表的敘詞化改造

標題表的敘詞化改造是一切傳統標題表極待完成的一項重要任務，現有 LCSH 及 ASTIA（美國武裝部隊技術資訊局敘詞表）兩種不同的改造模式。

（一）LCSH 模式

爲了更好地滿足電腦編目的需要，LCSH 從 1988 年出版的第 11 版（實際爲 1986 年 12 月）開始，正式改用敘詞表的款目格式，將原來含義不甚明確、直觀性較差的參照符號 See、X、See also、XX 改換成 USE（用）、UF（代）、BT（屬）、NT（分）、RT（參）、SA（說明參照）。這是一種「換湯不換藥」的改造，遭到

了圖書館界學者的批評。此後美國國會圖書館於 1991 年 5 月專門召開了一次研討會，專家們爲 LCSH 適應線上公用目錄提出了一系列改進建議，其中包括審訂重建標題、簡化複分、引入地理區域代碼及年代範圍代碼等，勾畫了 LCSH 敘詞化改造的具體步驟和方法。近年來這些提案已經逐漸見諸行動，但是大規模的改革尚需時日。^⑧

LCSH 模式實際上是一種治標不治本的改造模式，是對歷史悠久的 LCSH 的一次全面清理和修訂，通過引入敘詞表的技術和方法，使它更加符合電腦檢索的需要，但是改造的結果 LCSH 仍然是一部標題表，至多披上一件敘詞表的外衣。

LCSH 模式的形成源於 LCSH 這部大型標題表的實際情況。LCSH 自 1898 年問世以來，現已經積累了 30 多萬個標題詞和非標題詞。據統計，LCSH 已被 12 個國家的線上公用目錄採用，占全球 OPAC 的 43.52%，現在僅美國就已經積累了數百萬條用 LCSH 標引的記錄。^⑨百年的歷史和十多萬用戶使 LCSH 背上了沉重的包袱，積重難返，因而無法全面進行敘詞化的改造，只能改頭換面、修修補補。

（二）ASTIA 模式

早期的敘詞表大多是通過對標題表的改造而形成的，ASTIA 敘詞表就是其中典型的範例。ASTIA 標題表原有 10 萬個標題詞，1960 年被壓縮、轉換成 7,000 個敘詞，但是詞量減少了十分之九的 ASTIA 敘詞表的標引能力卻遠遠超過以往的 ASTIA 標題表。這是一種脫胎換骨的改造，是由表示主題的標題詞向表示基本概念的敘詞的一種轉換。改造的主要方法是砍斷由主標題—副標題—次副標題等組成的鎖鏈，拆整爲零，將每個主標題或複分標題都



變成可以自由組配的敘詞。另外還包括對原有標題詞的審核、篩選、歸併，和將那些繁冗的標題詞串（包括主標題和複分標題）改造為詞組性標題或短語式標題等。這樣改造的結果才是能夠真正適應計算機檢索需要的敘詞表。

中文圖書標題表的敘詞化改造應該採用那種模式呢？如效仿 LSCH，則可將原表中的參照符號——見、△、△△直接對換成 USE(用)、UF(代)、BT(屬)，至於「參見」項，則需逐一審查，凡屬下位概念一律轉換成 NT(分)，凡屬相關概念一律轉換成 RT(參)。除此而外，還應採取一些類似於 LCSH 的措施。

如果要將中文圖書標題表真正改造為一部名符其實的敘詞表，則需採用 ASTIA 模式，統計目前字順主表中共有副標題詞和次副標題詞 1,070 個，除去 347 個各國歷史下的複分標題，只剩下 723 個複分標目，分佈在 294 個標題款目下，也就是說，全表僅有 4.5% 的標題款目帶有副標題或次副標題，平均每個款目僅帶 2.5 個複分標題。從這一點看，此表與敘詞表十分相近。由於全表中複分標題為數不多，又都集中分佈在少量的標題款目中，將它們全部從主標題下游離出來，與主標題一起改造為敘詞，波及面不會太大，加之中文圖書標題表使用的時間不長，積累的標引記錄尚不多，所以採用 ASTIA 模式是完全可行的。

建議中文圖書標題表的改造採用下列步驟和方法：

1. 表中所有的標題，不管是主標題、副標題、還是次副標題，全部「平起平坐」凡屬反映基本概念的標題詞經篩選後轉換為正式敘詞。一些標引頻率較高的、不屬基本概念的標題詞，也可予以保留。
2. 凡非標題詞一律轉換為非敘詞，鑒於原

表等同率較低，在轉換時應大量增加非敘詞，即詞表的 UF(代)項，可以將正式敘詞的同義詞、準同義詞、同義詞組、部分反義詞以及一些過於專指的下位詞等均作為非敘詞收入，以增加這部詞表與自然語言的兼容互換能力。

3. 取消通用複分標題，取消仿分，允許自由組配。
4. 採用分類主題一體化詞表的編製技術。先將原分類表予以擴充和加細，在此基礎上對每一個類目進行嚴格的詞彙控制，規範其詞形和詞義，識別其同義詞、上位詞、下位詞和相關詞，然後用計算機將這些數據對應生成為敘詞表的代項、屬項、分項和參項，以提高詞表的關聯比及參照度。除字順敘詞表以外，還可以用計算機生成一部 KWIC 型的輪排索引。至於範疇索引及詞族索引，則可用下述分類法—敘詞表對應表代替，不必重覆設置。
5. 利用中文機讀目錄中的分類標引及主題標引數據，在計算機的輔助下，用統計轉換法生成一部分類法—敘詞表對應表（即分類主題詞表），代替原有的分類索引，以便更好地發揮其輔助標引、兼容互換等方面的作用，並為日後實現基於分類主題詞表的自動標引和自動分類創造條件。
6. 鑒於原表詞量偏少，應該大幅度地增加詞量，擴充詞表的規模，以適應各學科門類的發展，滿足文獻標引和資訊檢索的需要。

(收稿日期：1998 年 11 月 6 日)



註釋：

- 註①：呂紹虞編，中文標題總錄初稿，1935年；沈祖榮編，標題總錄，1937年；程長源，中文圖書標題法，第二版，（北京：商務印書館，1951年）。
- 註②：蘭開斯特（F. W. Lancaster）著；侯漢清等譯，資訊檢索詞彙控制，第二版，（上海：同濟大學出版社，1992年）。
- 註③：同上註。
- 註④：特引自 M. L Zeng, "Access to multilingual databases through universal standardized subject headings: pros and cons of implementation," 62nd IFLA General Conference (25-31, Aug., 1996, Beijing)。
- 註⑤：侯漢清主編，索引技術和索引標準（北京：北京圖書館出版社，1997年）。
- 註⑥：張琪玉著，資訊語言學基礎，增訂二版，（武漢：武漢大學出版社，1997年）。
- 註⑦：Mary Dykstra, "LC Subject Headings disguised as a thesaurus," Library Journal 113:43 (1998)。
- 註⑧：沈正華，「主題編目及國會圖書館標題表」，北京大學圖書館同人論文集（北京大學圖書館，1992年）。
- 註⑨：侯漢清、陳樹年，「漢語分類主題一體化詞表技術的發展和技術特色」，圖書與資訊學刊 21（民國86年）。

（本文曾發表於廣州中山大學在1998年3月29至4月2日所舉辦之海峽兩岸第四屆圖書資訊學術研討會）

