

漢語分類主題一體化詞表的發展和技術特色

The Integration of Chinese Classification and Thesaurus : Its Progress and Technical Features

侯漢清，陳樹年

Hanqing Hou, Shunian Chen

南京農業大學信息管理系

Dept. of Library and Information Science
Nanjing Agricultural University, Nanjing

【摘要 Abstract】

自80年代以來，我國學者開始漢語分類主題一體化詞表的研制，現已編纂、出版了10餘部分類主題一體化詞表。這種由分類法和叙詞表結合而成的新型詞表，已經成為大陸情報檢索語言研究的熱點之一。本文回顧、總結這種一體化詞表的發展，闡述它的類型、結構和功能，分析其中分面叙詞表和分類表—叙詞表雙向對照索引的技術特色和不足之處，提出改進其編製，及將其應用於自動標引和自動分類的建議。

Ever since the 1980s the Chinese scholars have started to work on the integration of Chinese classification and thesaurus and have so far compiled and published more than ten thesauri. Such new reference tools featuring the integration of the Chinese classification and thesaurus have become one of the highlights of information retrieval language studies on mainland China.

The paper aims to make a retrospective and a summing up of the progress of the said integrated thesaurus by stating its category, construction and functions, analyzing the technical features and drawbacks of the indexing contrasts between the separated thesauri and the classified ones, and making relevant proposals for the improvement of the system and the compatibility of the automated indexing and automated classification in application.

關鍵詞 Keyword

分類主題一體化 分類主題一體化詞表 檢索語言編製 檢索語言兼容

Integration of classification and thesaurus, Classification/thesaurus system

Construction of indexing language, Compatibility of indexing language



壹、漢語分類主題一體化詞表的發展

一、漢語一體化詞表編概況①

80年代初，大陸學者把國外分類主題一體化詞表（國外稱為分類法／敘詞表系統，簡稱一體化詞表）介紹到我國②，引起了圖書館同行的興趣。據統計，在10多年的時間內，先後發表了百餘篇評介、研究一體化詞表，探討分類法與主題法結合的理論和技術的論文，一體化詞表已經成為國內情報檢索語言研究的一個熱點。在理論研究的沃土上，終於結出了豐碩的果實。1983年和1987年國內學者以英國艾奇遜(J. Aitchison)的分面敘詞表為楷模，用手工和計算機先後編成常規武器工業分面敘詞表和教育分面敘詞表。這是我國圖書情報界編製分面敘詞表型一體化詞表的最早嘗試。此後，我國陸續編製、出版了19部一體化詞表，其中包括了3部大型詞表：中國分類主題詞表（共6冊，1500萬字）、農業科學敘詞表（共4冊，1140萬字）和社會科學檢索詞表（200萬字）和7部中型詞表。其中中國分類主題詞表和社會科學檢索詞表的編製和出版，對漢語一體化詞表的發展起了推波助瀾的作用，產生了廣泛的影響，國內相繼出現了一批專業一體化詞表。一體化詞表已經無可爭辯地成為一種新穎的、受到廣大用戶歡迎的情報檢索語言。這些一體化詞表的出現，象徵著我國情報檢索語言的研究和編製進入了一個新的階段。

二、漢語一體詞表的類型、結構和性能

漢語一體化詞表大致可以分為兩類，一類是分面敘詞表，如社會科學檢索詞表、農業科技敘詞表等（見表一）。這是以英國艾奇遜主編的世界上第一部一體化詞表Thesaurofacet的名稱命名的，可以說是最典型的一體化詞表。它通常由一部分面或半分面的分類表和一部字順敘詞表組

成，有的還附有輪排索引及英漢對照索引。這兩部分的類目和敘詞一一對應，兩部分用分類號或詞號相連繫。字順敘詞表起著分類表字順索引的作用，分類表則不僅起著傳統敘詞表範疇索引和詞族索引的作用，而且還可以直接用於主題標引。這是因為分面分類表中的上位類、下位類、同位類，加上表中設置的各種揭示等同關係和相關關係的參照，可以全面、系統、詳盡地顯示主題概念之間的各種關係，其顯示的深度和廣度均已超過一般的敘詞表。從主題標引查詞和選詞的便利性來看，分類表實際上也優於字順敘詞表。這類一體化詞表結構新穎，性能優異、適應性廣。

另一類漢語一體化詞表是分類表—敘詞表雙向對照索引，如中國分類主題詞表、中國法教育專業分類表等（見表二）。我國最早的分類號與敘詞雙向式對應工具文化教育主題分類表和分類主題詞表和〈中國法〉藝術及生物科學大類與〈漢表〉雙向對照索引是為調查中國法與漢語主題詞表（簡稱漢表）之間的對應關係，尋求實現分類語言和主題語言之間的兼容及互換而編製的。這些雙向對照索引通常由分類號與主題詞對應表、主題詞與分類號對應表兩部分組成。前者為每個類目列出其對應的一個或多個主題詞或詞串（即主題詞的組配形式），後者為每個主題詞列出其對應的一個或多個分類號。這種對照索引好似一部體系分類表與一部字順敘詞表的合璧，但加上互相對應的部分以後，它的功能就遠遠超過了這兩部分功能的總和。就拿中國分類主題詞表來看，它的第一卷（分類號與主題詞對應表）把漢表的全部主題詞納入了中國法的分類體系，換句話說，等於生成了一部按照中國法分類體系組織的漢表範疇索引，等於增加了一個用主題詞注釋的新版中國法分類表。它的第二卷為漢表的全部主題詞及表達中國法類目概念的主題詞串增添了對應的分類號，就等於生成了一部中國法的字順



表一 漢語分面叙詞表一覽表

編號	編製年代	詞表名	主編	備註
1	1983	常規武器工業分面叙詞表	熊興成	以中國法TJ武器工業類為基礎編成
2	1987	教育分面叙詞表	侯漢清、戴維民	用英政機編軟件編成
3	1987	航空航天技術主題詞 分面表	鐘慧	以中國法V航空航天大類為基礎編成，經張琪玉教授指導
4	1987-1992	農業科學叙詞表	方陸明等	1994年中國農業文獻出版社出版，共4冊
5	1987-1993	社會科學檢索詞表	朱鐵生、呂其蘇、 侯漢清	1995年社會科學文獻出版社出版
6	1987-1988	標準文獻主題詞表	中國標準化綜合研究所	ISO的Root Thesaurus的中譯本，共3冊
7	1987-1990	教育主題詞表	曹青陽、趙燕群 侯漢清	1993年教育科學出版社出版
8	1990-1991	人口主題詞表	孟健民、王榮授	1993年賀玲勇等製成聯機多功能一體化詞表
9	1992-1994	音像資料叙詞表	趙錦、戴維民	1996年計量出版社出版
10	1992	中國語文教育分面叙詞表	胡慶燮	
11	1992-1994	服裝叙詞表	洪漪等	



表二 漢語分類法—敘詞表雙向對照索引一覽表

編號	編製年代	詞表名	主編	備註
1	1984	文化教育主題詞分類表和 分類主題詞表	楊丹	以 <u>中國法</u> 文化教育類為基礎 編成，由趙燕群教授指導
2	1986	<中國法>藝術及生物科學大類 與<漢表>雙向對照索引	鄧嘯國	由張琪玉教授指導
3	1987	<中國法>U6水路運輸與<漢表> 雙向對照索引	北京圖書館 圖書館學研究部	
4	1987—1993	中國分類主題詞表	劉湘生、張琪玉、 侯漢清等	1994年華藝出版社出版， 共6冊
5	1991	<中國法>與MeSH、<中醫藥學主 題詞表>對照表	林美蘭等	1992年中國科學技術出版社 出版
6	1992	航空航天醫學主題詞表	吳國興	1993年由中國科學技術出版 社出版，其中分類主題詞表 為雙向對照索引
7	1992	<中國法>教育專業分類表與<教 育主題詞表>對照索引	趙燕群、柳芳等	1993年由教育科學出版社出 版
8	1992	郵電通信技術主題詞表	郵電表編委會	分類範疇表為中國法與郵 電表雙向對照索引

主題索引。這種分類表—敘詞表對照索引在編製上雖然借鑒了英國圖書館界編製的分類的國會圖書館標題表、主題標準檔等雙向對照索引，但是它「青出於藍勝於藍」不僅可以用於選類和選詞，而且可以同時用於分類及主題的標引和檢索，從分類標引及主題標引查詞（查類）和選詞（選類）的方便性來看，分類號—主題詞對應表實際

上優於一般分類表和主題詞表的字順表。這種對照索引已成為圖書館文獻標引和不同檢索語言之間兼容互換的重要工具。從兼容水平來看，對照索引顯然不如分面敘詞表。但是，從其實際的作用和影響來看，前者已遠遠超過了後者，更受到圖書館及情報界的歡迎。

實際上表二中的中國醫學科學院編製的中國



圖書館圖書分類法(R類)與醫學主題詞表,中醫藥學主題詞表對照表也可以列入第三類,即集成詞表(Integrated thesaurus)。它是「將某些特定主題領域的若干叙詞表和分類法匯編成一種集成詞表」,用於聯合分類標引和主題標引,用於實現分類語言與叙詞語言之間的兼容及互換。這種雙向對應式工具就是以一部分類表(中圖法的醫學大類)為主幹與兩部叙詞表(國際醫學界通用的醫學主題詞表—MeSH)和中醫藥學主題詞表進行對應標引。用戶在用中圖法分類標引之後,通過該對照表可以很快地找到相應的主題詞。這種集成詞表是前述對照索引的發展,規模和功能也超過前者。目前山西省圖書館與北京圖書館、中國科學院文獻情報中心,中國人民大學圖書館等正在合作研製一種「對照標引系統」,實際上就是編製一部包括中圖法、漢表、科圖法、人大法等在大內的大型集成詞表。這部機讀化的集成詞表的研製對於我國中文文獻的自動分類、自動標引以及檢索語言的兼容將會產生較大的影響。

從上述分面叙詞表和分類表—叙詞表雙向對照索引的結構可以看出,在一體化詞表中,通過對分類表和叙詞表的術語、參照、標識和索引等四部分的控制,分類法和叙詞表已經渾然一體,互相配合、互相輔助,從而發揮了最佳的整體效應。

實踐表明,這種分類主題一體化詞表具有多種功能和優點,主要是③:

- (一)可以通過標引數據的轉換同時完成文獻的分類標引和主題標引,降低標引難度,提高標引質量,大大節省人力和物力。
- (二)用戶可以在一系統中進行分類檢索、主題檢索或者由分類號、主題詞、關鍵詞三者組成的混合檢索,實現分類檢索和主題檢索的主題檢索的互補,提高檢索效率。一些只建立分類目錄、未建立主題目錄的圖

書館,通過中國分類主題詞表也可以進行主題檢索。

- (三)可以用於自然語言檢索系統,經過改造可以用作後控詞表,並用作自動分類和自動標引的工具。
- (四)可以成為計算機網路中不同檢索語言之間兼容互換的工具,為在數據庫檢索時從一種詞表(或分類表)轉向另一種詞表(或分類表)時提供方便。
- (五)可以由一個機構統一進行分類表及叙詞表等的編製和管理,提高詞表和類表編製的質量和速度。
- (六)為術語學研究和術語庫的建立提供參考工具。

三、漢語一體化詞表的編製模式

按照編製技術,一體化詞表的編製模式可以分為下列四種④:

- (一)改造分類表:即對一部現存的列舉式分類法進行分面改造,使之成為一部半分面分類表(或稱體系一分面分類法),同時要對分類表的詞形、詞義等實施嚴格的詞彙控制,然後在此基礎上用機編詞表軟件,自動生成一部對應的字順叙詞表。像武器工業分面叙詞表就是在中圖法TJ武器工業大類的基礎上用手工改編成的。這種方式尤其適合於分面分類法或半分面分類法,可以利用它們提供的數據,經過調整和加工,完成分面叙詞表的編製。由於目前國內圖書館大多採用中圖法,而中圖法不允許對其類表進行較大幅度的變動,這樣採用中圖法的圖書館藏不大適宜採用這種模式,但可選用其他模式編製專業分類表—叙詞表對照索引。有些情報機構擬對資料法進行分面改造,並在此基礎上編製專業分類表和叙詞表,正可採用此種模式。



(二)改造敘詞表：通過對敘詞表的分面改造，將敘詞表的範疇索引和詞族索引改造為一部分面或半分面的分類表，或合二為一組成一個新的範疇索引，從而改善詞表的分類顯示功能，並提高詞表的易用性。航空航天技術主題詞分面表就是將漢表範疇索引的「74航空航天技術」的全部主題詞和詞族索引的相關部分進行分面改造而編成的。農業科學敘詞表就是採用這種模式，將詞族索引與範疇索引合併，建成一個半分面的分類表。這種模式適合於採用漢表或專業敘詞表的專業圖書館和情報機構。採用這種模式，改動的重點在範疇索引和詞族索引，對字順表影響小，因而較為可行，易於收效。

(三)對現有的分類表和敘詞表進行對應標引，編製雙向對照索引：像中國分類主題詞表就是採用這種模式，在兩種現存的檢索語言基礎上建成了一種新的一體化詞表。這種模式適合於國內現有的大多數敘詞表和分類表，因為雙向對照索引的編製幾乎無需改動現存的詞表和類表。國內一些新編的敘詞表，如郵電通信技術主題詞表、教育主題詞表，往往採用這種模式編製中國法與該詞表的對照索引，用作詞表的範疇索引，以提高詞表的兼容性和實用性。

(四)新編分面分類表，自動生成字順敘詞表：像社會科學檢索詞表、音像資料敘詞表都是採用這種模式。這適合於計劃新編的敘詞表。採用這種模式難度高，工作量大，但編出的詞表性能優於第一和第二種模式中。當然新編一部分面敘詞表同樣需要參考現存各種詞表和類表，必要時也可借鑒前兩種模式，利用現有的分類表和詞表提供的詞彙和詞間關係資料。另外還應充分考

慮新編詞表與相關檢索語言之間的兼容和互換。

另外還可採用翻譯的方式，將國外的一體化詞表直接移植到我國來，必要時可作局部刪節和修改。這種方式當然只限於那些國際通用的詞表。如國際標準化組織（ISO）主持編製的基礎敘詞表（Root Thesaurus）第二版，於1988年被標準文獻情報中心譯成中文正式出版，改名為標準文獻主題詞表，現已廣泛運用於中文標準文獻的標引。

四、「一體化詞表熱」的原因分析

據統計，自1980年漢表出版以來，國內已經編製及正在編製的敘詞表大約有100多部。但是自1980年第一篇介紹國外一體化詞表的論文在國內專業期刊上發表以來，15年內總共編製了19部漢語一體化詞表，還有多部也正在編製之中，諸如職業教育主題詞表、熱帶農業主題詞表、中國法、農業專業分類表等。可以說一體化詞表已經佔到國內同時期編製漢語敘詞表總數的近20%，而且無論從詞表數量上或規模上均已超過國外編製的一體化詞表。

在我國出現這種「一體化詞表熱」，固然是因為受到國外的影響及國內出現了「建（數據）庫熱」和「詞表熱」，或者是因為「從某種意義上說，分面敘詞表具有全面的優點」，但是，其根本原因是一體化詞表適合於我國的國情。正如張琪玉所指出的，「分類法主題法一體化已成為在傳統使用分類檢索語言的中國，現在轉向兼用主題檢索語言時的一種強烈呼聲。在一些傳統使用主題檢索語言的國家，這種呼聲可能不會那麼強烈」^⑤。這也正是一體化詞表最早在國外出現，但是國外的發展卻不如中國的真正原因。

貳、漢語分類主題一體化詞表的技術特色



一、詳細分類表的引入——詞表宏觀結構的革新

80年代以來，我國漢語敘詞表的編製以漢表為代表，逐漸成了一種固定的結構模式，即以字順表為主，附加範疇索引、詞族索引和英漢對照索引。有些詞表還單獨設置了附表（或稱專有敘詞表）。據對80年代29部漢語詞表的統計^⑥，同時設有字順表、範疇索引、詞族索引和英漢索引的詞表共16部，佔55.17%。同時設置字順表和範疇索引的詞表23部，佔79.31%（見表三）。

表三 漢語敘詞表的宏觀結構

	詞表數	比率
設有字順表	29	100%
設有範疇索引	23	79.31%
設有詞族索引	16	55.17%
設有英詞索引	16	55.17%
設有附表	13	44.83%

在這些敘詞表中，詞彙的字順顯示主要是由字順表完成的，詞彙的分類顯示則是由範疇索引、詞族索引及字順表中的參照系統完成的。就分類顯示而言，前述詞表模式存在著下列缺點：

- (一) 範疇索引的分類粗略，通常只分為兩、三級，一個範疇小類下集中的詞彙少則幾個，多則幾百，甚至一、二千（漢表有的小類最多收詞2250個，最少則為2個詞），加之同級詞彙往往按照字順排列，無法顯

示詞彙之間的關係，形成了在範疇索引中查詞難的狀況。

(二) 詞族索引只顯示詞彙的等級關係，不反映詞彙之間的等同關係和相關關係，也不反映詞族與詞族之間的關係。

(三) 範疇索引和詞族索引雖然同為分類顯示的手段，但二者分設，不建立任何聯繫，影響詞表分類顯示功能的發揮。其實在漢語敘詞表中二者之間重覆交叉甚多。據統計，漢語主題詞表的社會科學部分，92.7%的詞族全屬於範疇索引的同一個大類，78.2%則同屬於一個二級類^⑧，其他漢語敘詞表的情況與此類似。

四 範疇索引和詞族索引功能單一，只能用於查詞和選詞，不能用於分類標引和主題標引。80年代末陸續問世的漢語一體化詞表針對傳統敘詞表存在的上述缺點，先是進行範疇索引和詞族索引的合併，後來大膽地在詞表中直接引入一個詳細的分類表（見表四）。詞表宏觀結構的這種變革，宣告一種新型的漢語分面敘詞表——分類法與敘詞表一體化詞表的問世。

從表四可以看出，這些漢語一體化詞表的宏觀結構突破了傳統敘詞表舊的模式，呈現了下列不同於傳統敘詞表的特點：

- (一) 用一部詳細的分類表（或是分面分類法或是體系——組配分類法）替代以往的範疇索引和詞族索引，簡化了詞表的結構。
- (二) 由於在分類表中實施了嚴格的詞彙控制（詳見段二），包含了敘詞的全部信息，分類表已成為一體化詞表的主體，字順表成了分類表的派生物，重複顯示分類表中包含的敘詞信息，提供字順檢索的途徑。
- (三) 分類表中的類目與字順表中的敘詞一一對應，用分類號和詞號相聯繫，分類表和字



表四 漢語分面叙詞表的結構

詞表名	詞表結構	備 註
教育分面叙詞表	分面分類表+字順叙詞表	
航空航天技術主題 詞表分面表	主題詞分面表	將漢表範疇索引及詞族索引的 有關部分合二為一
農業科學叙詞表	字順表+分類表+簡表+逆序輪排索 引+英拉漢索引	分類表系將詞族索引和範疇索 引合二為一
社會科學檢索詞表	主題字順表+分面分類表+英漢對照 索引+輪排索引	輪排索引有機讀版，無印刷版
教育主題詞表	字順叙詞表+分面分類表+附表+輪 排索引+英漢對照索引	
人口主題詞表	分類表+字順表+輪排索引+英漢索 引	
音像資料叙詞表	分類叙詞表部+字順叙詞表部+輪排 索引+英漢索引+附表	
服裝叙詞表	分類表+字順表	

順表二者融為一體，相互輔助。

四分類表的上下位類不僅顯示等級屬分關係，而且還顯示了非等級關係，如整體／部分關係，行動／施動者關係、事物／方面關係或其他相關關係。

(五)改以往一組詞設一範疇號的辦法為每一類目（叙詞）配備一分類號，使分類標記個別化，便於檢索。

把分類表引入漢語叙詞表，加之對分類表的嚴格詞彙控制，可以全面、系統地顯示詞彙之間的複雜關係。一體化詞表的分類表部展示的不僅僅是一條條類鏈(chain of a class)、一個個詞族(hierarchy)或局部的學科或專業，而且提供了廣闊的主題領域的全貌，克服了範疇索引、詞族索引在分類顯示功能上的不足。引入分類表的好處也是很顯然的，諸如結構簡明，以簡取繁，整體



優化，功能多樣，便於檢索，便於標引（包括分類標引和主題標引），提高兼容性，便於標引數據的互換，及便於詞表的計算機輔助編製和管理等。

二、嚴格的詞彙控制——詞表微觀結構的改造

一部分類表與一部敘詞表之和不等於一部分類主題一體化詞表，這是因為只有機械的相加，而沒有內部要素的協調，不會產生新的整體效應。這裏所說的「內部協調」就是施加統一的、嚴格的詞彙控制，也就是說必須通過對詞表微觀結構的改造及其相互配合，真正實現分類表和敘詞表的一體化。

漢語分面敘詞表的詞彙控制主要包括下列方法：

(一)詞表構件的改造

分面敘詞表由分類表和敘詞表兩部分組成。這兩部分的構造已不同於一般的分類表和敘詞表。一般敘詞表款目由款目詞、標記項、注釋項，及參照項組成。漢語分面敘詞表基本保留了一般敘詞表的款目形式，作以下幾項改造（實例見表五）：

- 1.增設了詞號。按敘詞表的漢語拼音字母順序給表中每個敘詞和非敘詞逐一配備流水序號，以便能提高查詞的速度。
- 2.給每個敘詞添加詳細的分類號，以代替以往粗略的範疇號。它和詞號均為聯繫字順表與分類表的手段。
- 3.為注釋增設了專用符號「J」，以往敘詞款目中的注釋直接寫在敘詞下，不使用專用符號，不便於識別。
- 4.在屬項、分項、參項參照之後逐一配備了詞號，並為非本類的、即分散在其他類的上位詞、下位詞和相關詞逐一配備相應的分類號，這樣既提高了查詞及查類的速度

，又密切了字順表與分類表的聯繫，至於出現在本類附近的上位詞、下位詞及相關詞，因款目詞後已添加相應的分類號，就不再重複顯示了。

一般分類表通常由類號、類名和注釋三部分組成。在注釋中，既有類目的同義詞、含義注釋、內容範圍注釋和標引方法注釋，又有各種分類參照統統混在一起。這既妨礙詞間關係的顯示，也不利於分類表與字順表的對應。另外在一般分類表中往往很少設置分類參照，即使設置參照也不區分它們顯示詞間關係的種類。分面敘詞表對一般分類表的構件進行了下列改造和細化（實例見表五）：

- 1.新增類數，用圓點直觀地表示類目在分類體系中的層級。
- 2.在類名下設置D項，專門陳放類目的同義詞、準同義詞或過於專指的下位詞（類目含義注釋入J項）。
- 3.把類目各種注釋納入新設置的J項，原注釋中的項目參照不入J項。
- 4.在類名下設置S項、F項和C項，把出現在類表其他地方的上位詞、下位詞和相關詞分別列入S項、F項和C項參照。
- 5.在類目後附加相應的詞號，並在S項、F項和C項詞後也附加相應的分類號和詞號，以便迅速轉查字順表和分類表的其他類目。
- 6.大量設置分面標頭，說明類目劃分時採用的分類特徵。

改造後分類表增設了D項、S項、F項、C項等參照項，有助於用戶理解類目涵義。詞號和類號的附加，使分類表與字順表之間、分類表類目之間的聯繫大大加強，提高了分類表內主題概念之間的網路性，也為與字順表的對應和轉換創造了條件。



表五 多重等級關係的顯示

教育主題詞表分面分類表	教育主題詞表字順敘詞表
F · 各類教育 00554	師範教育 F A 01754
C 教育事業	Teacher education
F A · · 師範教育 01754	S 各類教育 00554
C 教師培訓 J Q7 00867	F 高等師範教育 00510
進修課程 N T793 01208
(按師範教育特殊性分)	職業師範教育 02667
F A G · · · 職業師範教育 02667	C 教師培訓 00867
D 職業技術師範教育	進修課程 01208
S 職業教育 F B 02663	職業教育 F B 02663
C 職業學校 G P 02668	Vocational education
	D 技術教育
	職業技術教育
	S 各類教育 00554
	F 職業師範教育 02667
	C
	職業師範教育 F A 6 02667
	Vocational teacher education
	D 職業技術師範教育
	S 師範教育 01754
	職業教育 02663

(二) 類目名稱敘詞化

一般分類法比較重視分類體系的構造，不大重視類目名稱的規範，敘詞表則不僅重視詞間關係的控制而且重視詞形（敘詞字面形式）和詞義的控制。為了實現由分類輸入數據向字順表的對應和轉換，必須把類目名稱敘詞化。漢語分面敘詞表通常實行「一個主題一個位置」及「一個主題一個語詞」兩條原則。

20世紀初，英國圖書館學家布朗(Dr.Brown)在編製主題分類法時提出了「一個主題一個位置」的思想。隨著主題分類法衰亡，這一思想也遭

到責難和摒棄。為了實現類名和敘詞的一一對應，一體化詞表必須遵循這一思想，要求同一主題在分類表及敘詞表（指字順表）中只出現一次，只有一個位置，而且使用同一個詞或短語來表示。如何解決一個主題有時可以歸入兩個或兩個以上的學科，即同一主題在分類表出現兩次及兩次以上呢？社會科學檢索詞表、教育主題詞表等一體化詞表採取了下列做法。

如果同一個主題在兩個學科出現，其含義完全相同，一般分類法對這種具有多重等級關係的主題採用設置交替類目的辦法來解決。一體化詞



表則採用設置分類參照的辦法。如教育主題詞表處理「職業師範教育」這一主題，在分類表的「師範教育」類下設置正式類目，在分類表「職業教育」類下則不設置類目。與此同時，在「FA6 職業師範教育」類下設置一個屬項參照（「職業師範教育 S 職業教育」），這樣就可以把「職業師範教育」同時從屬於兩個類的多重等級關係顯示出來（詳見表五）。這種做法與分類表設置交替類目異曲同工。

有些語詞在分類表中出現多次，如「小說」一詞在中國法中出現多達一、二十次，但每次的實際含義是不同的。在「F 2 中國文學」下出現是指中國小說，在「I106（世界文學）作品評論和研究」下出現是指「小說的評論和研究」，在「I210魯迅著作」下出現是指「魯迅小說作品」。社會科學檢索詞表等一體化詞表規定：類名

必須完整，採用全稱，不准使用簡寫，必須採用準確反映其主題含義的語詞。這樣就可以達到一個主題一個位置，一個主題一個語詞的目的。

（三）被組代詞和聚類詞的設置

在分面叙詞表中，幾乎全部類目和叙詞都一一等值對應，但是只有兩種情況——被組代詞和聚類詞例外。一些叙詞表為先組度較高的詞設置了組代參照，但是一般的分類表沒有這種做法。鑒於分類表的先組度往往高於一般的叙詞表，漢語分面叙詞表採取了在分類表和字順表中分別設置被組代詞的做法。通常在分類表中分別將一個先組度高的被組代詞用作類名，在類名下設置一條組代參照，顯示這個被組代詞可以用兩個叙詞組配表示。舉例而言，在標準文獻主題詞表的主題顯示表中，如：

	(特定溫度測量用組配，例如)
BJD.FF 沸點測定 →溫度測量BJD.F+沸點CIL.GLL
BJD.FJ 混濁點測定 →溫度測量BJD.F+混濁點CIL.RB

其中，「沸點測定」這一組代詞在字順表中有下列對應叙詞款目：

沸點測定BJD.FF →溫度測量 BJD.F+沸點 CIL.GLL	沸點 CIL.GLL +溫度測量 BJD.F = ... 沸點測定 BJD.FF
(其中 ... 為被組代詞，→為用項，+為組代關係)	

這種做法不是一對一，而是一對多，是用一個先組詞與兩個後組詞對應。這實際上是對詞量和先組度的控制，對於壓縮詞表的篇幅、保證標引的一致性及組配檢索的正確性，都是有益的。

聚類詞（或稱虛設類目）是分類表中只用於

聚類、不用於標引的類目詞，在字順表中沒有對應的叙詞或非叙詞。聚類詞常常是一些沒有檢索意義的詞，如同中國法中的「一般性問題」、「按...分」，在音像資料叙詞表中，「BC藝術類型」、「BR44雕塑類型」、「P12傳播類型」等



均為聚類詞，在詞後都注了星號（*），表示該詞不能用於標引。

四、輸入數據表的加工

前述漢語一體化詞表的種種詞彙控制方法都是通過輸入數據表而實施的。分面敘詞表的編製流程包括：(1)詞表總體設計；(2)輸入數據的搜集和加工；(3)計算機自動生成分類表、字順表及輪排索引等；(4)試標引和修改；(5)編輯定稿。從編製流程可以看出，輸入數據表是自動生成詞表的基礎，也是計算機輔助編輯排版、詞表管理的重要工具。

輸入數據表的加工，除了搜詞、選詞外，就是進行詞彙的分面分析和詞彙控制，在此基礎上構造一份呈分類表狀的輸入數據表（亦稱分類輸入數據表）。其中輸入數據包括：分面標頭、類級、類號、類目及其英譯名、類目注釋及參照（包括D項、S項、F項、C項等）。它有一般分類表的差別在於：

1. 在類號前增加類級，日後可以生成顯示類目層級的級點（小圓點，一級類用一個圓點二級類用兩個圓點，三級類用三個圓點…，並用以控制分類表的版面格式。
2. 在類目名稱後添加英文譯名，日後英譯名移至字順表的漢語敘詞及非敘詞之後，並用以生成英漢對照索引。
3. 對類名進行切分，即在類名的詞彙之間添加分詞符號，以便日後生成輪排索引。如果有專業詞彙詞典（或稱部件詞典），也可進行對類名的自動分詞或計算機輔助切分。
4. 設置了上位類性質項。此項數據用於識別本類與上位類之間的關係。段貳之一述及，一體化詞表的分類表為了全面展示每一學科領域主題（即敘詞）之間的關係，在表中既顯示等級關係，也顯示非等級關係

。上下位類之間除了等級屬種關係（不加任何代碼）以外還存在著下列幾種關係：(1)本類與上位類為相關關係，用代碼C表示；(2)本類與上位類既非屬種關係，也非相關關係，用代碼N表示；(3)本類目是聚類詞，不用於標引，用代碼P表示。計算機生成時可以根據代碼C在上下位類之間生成一條相關參照；根據代碼N，不生成任何參照；根據代碼P，下位類可以跳過直接的上位類，與再上一級的類目生成必要的參照（標注P的類目為聚類詞，不生成對應的敘詞）。根據缺損值，可以生成屬分參照。在生成必要的參照後，表示上位類性質的代碼自動清除，不出現在生成的分類表中。這裏所進行的上、下位類關係的識別和標注，實際上就是一個詞間關係控制的過程，為日後生成各類參照作準備。

一般敘詞表的編製是在搜詞、選詞和審詞的基礎上為每個敘詞分別製作「主題詞登錄卡」，為什麼分面敘詞表的輸入數據要採取一種詳細分類表的形式呢？這實際上是用一種系統的編表方法來取代以往孤立的，零敲碎打的構造敘詞款目的方法。構造一個完整的分類輸入數據，實際上是建立一個完整、詳細的分類體系。有了這樣一個分面或半分面的分類體系，有助於發現搜詞選詞和等級關係顯示中的遺漏，發現詞彙的多重等級關係，還有助於辨別同義詞和建立有用的相關關係等。構造分類數據表的過程是一個分面分析和詞彙控制的過程，雖然要花費較多的力氣，但還是值得的。

至於漢語一體化詞表的分面分析技術和計算機輔助編製技術，限於篇幅，當另文論述。

三、對應標引技術

雙方對照索引式一體化詞表，是將一部分類



法和一部叙詞表通過「叙詞—類目」的多重對應關係，形成一個有機的整體，實現兩種檢索語言兼容及互換。我國建國後編製的分類法絕大多數是列舉式分類法，屬先組型，如何與後組型的叙詞表實現對應，是關鍵性的技術。在編製雙向對照索引式一體化詞表時，主要是解決類目與叙詞的對應，至於叙詞間與類目（類號）對應，則可通過詞表編製軟件自動生成。現以中國分類主題

TL24	乏燃料後處理	輻照過的核燃料—核燃料後處理 核燃料後處理；輻照過的核燃料儲存；Butex 法；Eurex 法；Nepetex法；……裂變產物—分離法
------	--------	--

由於一個事物或問題可以從多學科角度進行研究，因此同一個叙詞與多個類目對應。在對應時根據叙詞與類目的聯繫程度採用不同的對應方法。如果依分類法類目內容劃分規則能判定某叙詞的主要學科屬性，則將該詞對應於主要隸屬的類目，而在其它類下以〔×××〕的形式對應，

TP701	理論（遠感技術）	遠感技術—理論 地面實相；地物光譜特性；航空磁學；遠感光譜段；遠感數據 注：以下主題詞均應組配標引，如：遠感技術—大地輻射—理論 大地輻射；大氣傳輸特性；地面反射；〔電磁輻射〕
-------	----------	---

（二）特殊類目與叙詞的對應。類目與叙詞一個很大不同點是，它往往不僅表達一個獨主的主題，還可以表達幾個並列的主題，或只起構造類目體系作用而沒有實質的主題意義，對此種類目均需採用特殊的對應方法。例如，類組性類目，用叙詞分別與其包含的主題概念加以對應；「其他」類目，用其上位類對應的叙詞和其本身主題對

應的叙詞一併加以對應；「一般性問題」本身不予對應，但其下位類則對應相應的叙詞；「各種××」、「按××分」之類的類目，使用叙詞加自然語言修飾的方法加以對應，如：「各種體制的雷達」，用「雷達，各種體制的」對應；帶起止符號的類目，一般採用叙詞加自然語言修飾的方法對應，如「K281/288各民族史志」，用「中

國表第一卷分類號—主題詞對應表為例，作如下分析：

（一）一般類目與叙詞的對應。在分析類目及其注釋涵義和外延範圍的基礎上，根據叙詞標引規則，使用叙詞和叙詞詞串（叙詞的組配形式）加以描述，並將屬於該類號範疇內的全部叙詞對應在該類號之下。例如：

例如「過電壓保護」主要對應在TM86類下，在TM07等類下則對應為〔過電壓保護〕。

由於相當多類目的主題內容只能採用叙詞組配的方式描述，這樣無疑會大大增加對應表的篇幅，因此凡具有相同組配形式的叙詞詞串，可用注釋的方法加以說明，例如：



華民族—民族志，各民族」對應；屬某一學科或事物在其它領域應用的類目，根據不同的情況分別對應為「××—應用」，或「××—應用—××」，或「××應用—××」；屬某學科與其它學科關係的類目，根據不同的情況分別對應為「××—關係」或「××—關係××」，或「××—影響—××」，或「××—作用—××」，等等。

④複分表類目與敘詞的對應。複分表（包括總論複分表和專類複分表中的子目）只起對主類號細分的作用，不能單獨使用，與敘詞的對應只需

對應其自身的主題概念，而不必與其所處的位置相聯繫，這種對應法有助於相關敘詞組配時選詞參考。

⑤類目注釋與敘詞的對應。類目注釋中屬補充說明該類目內容範圍的，用敘詞加以對應，屬說明分類方法的，在敘詞對應欄轉換為敘詞組配標引方法注釋；屬說明參見關係的，對應時轉換為參見相應的敘詞；屬說明類目內容劃分範圍的，對應時轉換為指示用某類下對應的敘詞標引；屬書次號編製方法的注釋不予對應。例如：

049-09 物理學史 依世界地區表分	物理學史 注：可用該主題詞與世界地區表對應的主題詞組配標引。如：049-093.13可對應為：物理學史—日本。
R517.3 破傷風 參見633	破傷風 注：參見R633對應的主題詞。

⑥交替類目與敘詞的對應。凡交替類目對應時，均將對應的敘詞用方括號圍起來，同時注明「見×××類對應的敘詞」。

⑦通用詞的對應。敘詞中有不少是用於組配的通用概念；如設計、製造、作用、進行等，它們不能獨立表達一個文獻主題，因此在分類中也無相應的類目與之對應。解決的方法是在總論複分表中增設「通用概念」一類，對應類無所屬的通用敘詞。

⑧專用敘詞的對應。機構名稱、人名、設備及產品名稱等專用敘詞，可採用編製附表的方法加以對應，以減少主表的篇幅。

通過上述對應技術，就可以把分類法與敘詞表有機地聯結成一個整體，達到兩種檢索語言兼容互換的目的。

參、漢語分類主題一體化詞表的未來

漢語一體化詞表的出現和發展回答了以往圖書館和情報界爭論不休的分類法與主題法可否實現一體化、分類與主題可否對應等理論問題，標誌著我國敘詞表和分類法的研製，已經開始實現從傳統結構模式向新的結構模式的轉變；從手工編製向計算機輔助編製轉變，從分散編表向兼容互換方向的轉變。它對於我國計算機輔助標引和自動標引、文獻數據庫建設及計算機檢索系統的發展，起到了一定的促進作用。

從目前已問世的十幾部漢語一體化詞表的編製和使用情況來看，仍存在著下列問題：

1. 分面敘詞表追求分類體系的完整性，因而



出現詞表中部分詞彙在分類體系中是必不可缺的，但在實際標引和檢索中却是冗餘的。

2. 分類輸入數據表的構造，雖可以參考現存的分類體系和叙詞詞族，但仍頗費功夫。分類法／叙詞表雙向對照索引的對應標引全部由人工完成，興師動眾，曠日持久。
3. 分面叙詞表的分類表和字順表兩部分數據重複交叉嚴重，因而詞表篇幅大，不便使用，出現「翻半天詞表，標幾個叙詞」的情況。
4. 目前漢語分面叙詞表詞彙顯示方式單一，字順表部參照只有等級關係單級顯示，無多級顯示或全顯示方式。
5. 缺少自然語言接口，用戶無法用關鍵詞和自由詞進入檢索系統，難於發展機助標引和自動標引。
6. 漢語分面叙詞表分類表部的分類體系與國內目前通行的中圖法、資料法的體系不相兼容，用戶不願同時熟悉和使用兩種分類體系。
7. 在參照設置、英譯名編配、主題詞對應標引等方面存在著一些質量問題。

近年來漢語一體化詞表在使用中不斷得到改進，正在朝著標準化、兼容化、計算機化和使用自然語言等方向發展。建議今後採取下列措施改進漢語一體化詞表的編製和使用。

一、為分類法—叙詞表雙向對照索引及漢語分面叙詞表編製入口詞表

目前在計算機檢索中越來越多地採用自然語言，包括關鍵詞和自由詞，目前在一體化詞表中所收非叙詞的比率較低，檢索入口少，影響其作用的發揮，因此需要為它們補編入口詞表。

入口詞是指詞表中收錄，但只能用作查詢入口，不能用於標引和檢索的詞，入口詞表就是這

種非叙詞、自由詞及關鍵詞的彙集。可以把編製一體化詞表過程中落選的詞彙、以及標引記錄、用戶檢索提問中的關鍵詞統統累積起來，單獨建成一個入口詞表，以加快標引和檢索的速度，減輕用戶選擇主題詞的智力負擔。這種入口詞表可以將自然語言轉換為主題詞或分類號，為自動標引和自動分類提供重要的工具。

入口詞表的編製應當注意：(1)充分利用現有的自動切分軟件，把文獻數據庫中的關鍵詞抽取出來。(2)入口詞的形式可以是一對一，也可以是一對二或一對多。(3)入口詞表可以採取輪排表的形式（不必再另編字順總索引），為用戶提供盡可能多的查詞途徑。(4)入口詞表應當是動態的，不斷補充和調整。

二、為漢語一體化詞表編製聯機詞表

入口主題詞表在印刷版的基礎上開發研製出「聯機多功能一體化詞表系統」其主要功能是：(1)提供機讀詞表和印刷版詞表的批式輔助生成系統；(2)提供靈活完善的查找概念和揭示語義關係的手段；(3)提供全面的詞表動態維護功能；(4)提供標引、檢索用詞的直接複寫功能（不必另填工作單），實現標引、建庫、出刊的一體化；(5)具有標引、檢索頻率統計；(6)中外文對照顯示；(7)自動編配參照等功能。

有了這樣的聯機詞表，不僅便於一體化詞表的生成及管理，而且便於查詞換詞，即使詞表篇幅再大，也無影響。如果能把入口詞表也加入聯機詞表，一體化詞表就能如虎添翼，發揮更大的功能，甚至在自然語言檢索中發揮後控制詞表的作用。

三、提高詞表編製自動化的程度

改變以往分類法—叙詞表雙向對照索引完全用手工進行對應標引的狀況，充分利用集中編目和計算機編目成果，將印刷卡片、國家書目或機讀目錄上的分類標引和主題標引的數據，進行排



序整理，加上必要的人工瀏覽和評價，即可用計算機輔助完成對應標引數據。中國分類主題詞表可以用這種方法進行修訂和補充。另外中國分類主題詞表還應當擴大兼容的範圍把國內和國外通用的其他分類法，如科圖法、人大法、中國圖書分類法（賴永祥編）、杜威十進分類法、國際十進分類法等收錄進來，使之成為國內外分類表和詞表的兼容互換中心。

至於漢語分面叙詞表的編製，一是要注意與中國法、漢表的兼容，二是要盡量向中國法的分類體系靠攏，並作必要的調整，三是要充分利用現在的國家詞庫和機續詞表的機續數據，以降低編製分類輸入數據表的難度，增強詞表的兼容性和易用性。

四、詞表顯示及款目形式的多樣化

漢語分面叙詞表的款目的形式，可以根據用

戶的需要選定。分類表和字順表兩大部分可以相互重複或部分重複交叉，也可以不相重複而是兩部分互相補充。

至於分類法—叙詞表雙向對照索引，也可採取國外兼容矩陣的形式，直接用類號對主題詞，不必印出全部分類表和詞表。另外似應為之編製輪排索引，以增加檢索途徑。

隨著近年來中國分類主題詞表、社會科學檢索詞表、農業科學叙詞表等多部大型詞表的陸續問世，今後將會有更多的專業詞表仿效它們，原已出版的詞表中將會有一批通過叙詞表的分面改造，加入到漢語一體化詞表的行列。漢語一體化詞表將會以新的面貌、新的功能來迎接聯機檢索和自然語言檢索的挑戰。

（收稿日期：1996年8月1日）

註釋

- 註①：戴維民，「我國分類主題一體化研究進展」，大學圖書館通訊，6期（1987年）。
- 註②：侯漢清，「分類法與主題法結合的成功嘗試——分面叙詞表」，圖書情報工作，6期（1981年）
- 註③：張琪玉、劉湘生主編，中國分類主題詞表教程（北京：華藝出版社，1984年）
- 註④：侯漢清、馬振華主編，主題法導論（北京：北京大學出版社，1991年）
- 註⑤：張琪玉，走向21世紀的中國情報檢索語言和情報語言學，1993年，手稿
- 註⑥：楊南征，「國內外叙詞表構成情況調查分析報告」，全國漢語叙詞表發展方向研討會，1988年12月
- 註⑦：侯漢清，「論漢語叙詞表編製的新模式和新方法」，中文信息處理標準化國際研討會論文集，1989年3月
- 註⑧：曾蕾，「對《漢語主題詞表》的質量評份及改進意見」，情報學報，3期（1985年）

