

網路版詞典編輯經驗析介

中央警察大學通識教育中心教授 曾榮汾

【摘要】：本文旨在介紹筆者從事網路版詞典編輯的經驗。以《重編國語辭典修訂本》及《異體字字典》為例，介紹了這兩部網路詞典編輯的流程，並說明編輯時應當注意的要點。文中也特別強調了上述網路版詞典公開後所運到的問題。

關鍵詞：詞（辭）典學（lexicography）；詞（辭）典（dictionary）；網路（web）；
網際網路（internet）；編輯（edit）

一、前言

詞典的編輯結合了新科技的運用，已進入網路時代。利用網路來展現詞典的編輯成果，無論使用或維護效益，皆非傳統紙面版詞典所可比擬。網路詞典是指將編輯成果利用網際網路傳播出去的詞典。它和使用者的界面是網頁，它所提供的內容是多媒體，所提供的查詢方式是全文搜尋。它置於伺服器上，快速地向網路世界公開，它也可以不斷地將資料庫修訂補充。甚至於可以連結許多相關資料庫作進階使用。這些特點，是過去在傳統紙面版編輯時無法想像的。紙面版印量有限，修訂困難，查索不易，也無法結合多媒體。這樣看來，似乎紙面版已無法持續發展，但是不可否認的，它卻有著網路版本所無法取代的書香質感和閱讀印象。反過來看，網路版也有它的局限，因為網際網路的特性影響，所以網路上

的資料必須是快速反應，讀者的期盼無形間受到使用網路習慣影響，會對網路版有著更高、更嚴格的期許。這也是紙面版較不易遇到的問題。

因此，網路詞典對編輯者來說是個新挑戰，它並不是將紙面版詞典內容輸進電腦，而後上網即可。它必須接受網路文化的薰陶，進而化身為真正的網路成員。所以累積經驗，並交換心得是編輯者重要的工作。筆者自民國六十年代初期即參與字詞典的編輯，在此階段，所完成的成果當然都是紙本工具書。而使用電腦來編輯工具書，則始於民國七十六年。第一部的成果是教育部的《重編國語辭典修訂本》，民國八十三年九月發表了臺灣學術網路版本。以時間點來看，當是國內較早公布的電子詞典及網路詞典。從《修訂本》後，筆者陸續幫教育部完成了數部工具書，最新的一部則是教育部的《成語典》。一見三十餘年，



多少累積了一點經驗。自此撰文野人獻曝，將心得分享出來。

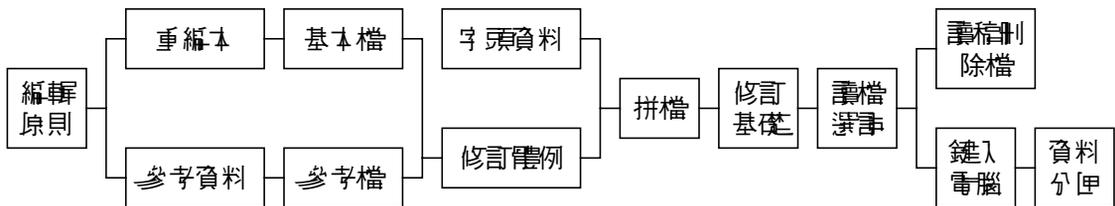
二、《國語辭典》系列的編輯經驗

筆者開始編國語辭典時，國內中文電腦系統剛推出不久。像倚仗的中文系統，仍然不太成熟。資料庫系統最流行的就是 dBase III。一般人對電腦的觀念大概仍停留在「專業的高科技」，只要能用 PE II 進行文書處理就是電腦人才了。在那樣的環境中，筆者對電腦也是了解得很粗淺，只是知道它能快速運算，大量儲存，持續編修。但筆者當時就在考慮，編了十幾年的字詞典，都在卡片

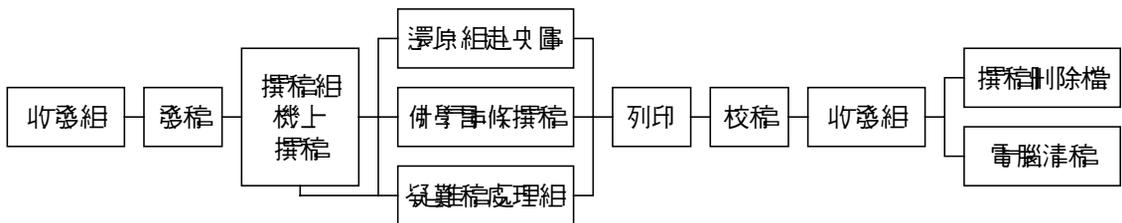
和稿紙上打轉，是否應當提升為電腦編輯。編成電腦檔後，再來進行排版印刷。當時腦海中並無所謂的「電子版本」或「網路版本」的影子。

自此決定嘗試改為電腦編輯後，首先遇到的是硬體的不足和軟體的搭配。我們先由工作人員自行採購兩部 286 個人電腦做為主力。軟體則由筆者根據過去的編輯經驗寫成的編輯系統。於是筆者的編輯歲月從此邁入電腦時代。後來硬體設備逐漸充裕，編輯工作也隨之順暢。對於這部詞典，筆者規劃的編輯流程如下：（註 1）

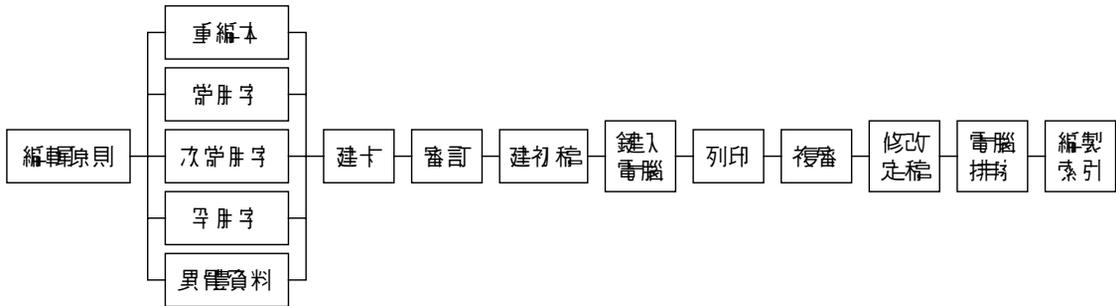
（一）編輯檔案建檔流程



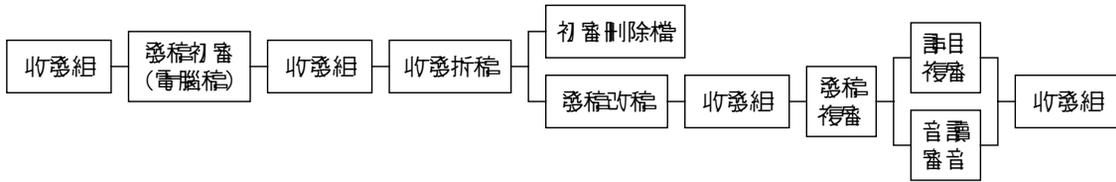
（二）撰稿流程



(三) 字頭撰檔流程



(四) 審稿流程



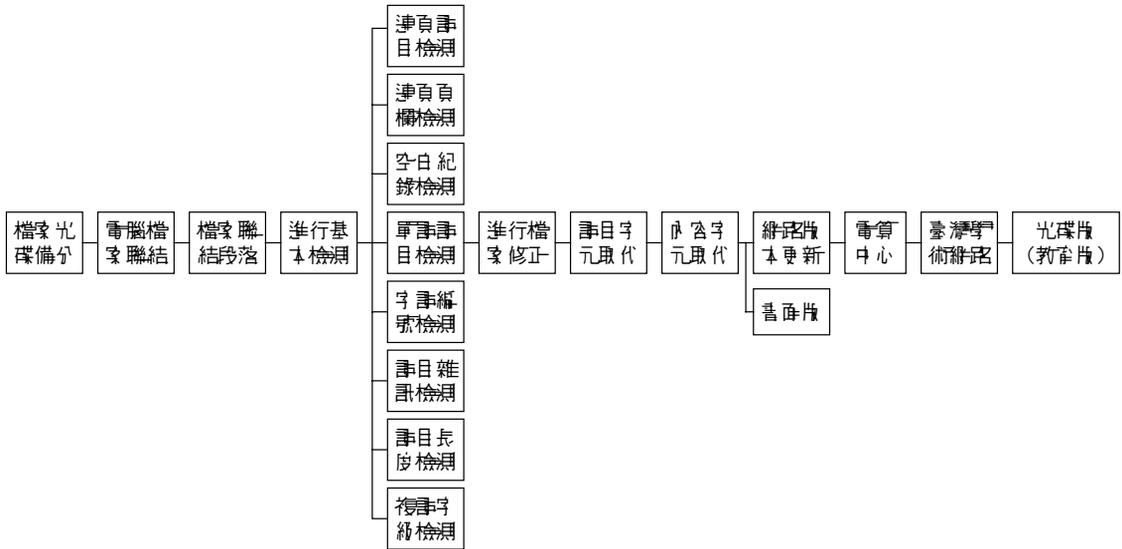
(五) 定稿流程



民國八十二年編輯成果逐漸完成，意外地，原有出版社無意出版紙面版，剛好當時臺灣學術網路「架構完成，正在積極尋求資源，於是我們改和教育部電算中心合作，將成果委由中研院資管所的 Csmart 小組開發網路版本 (<http://140.111.1.22/clc/dict/>)。工作的流程如下：

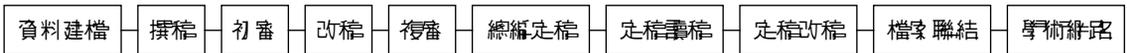


(六) 網路版編製流程



在此階段我們作了許多檢測，一則是因為要解決中文系統帶來的問題，二則是嘗試利用程式來輔助校對。如果欲知整部詞典編輯的總流程，概如下列：

(七) 工作總流程



從以上這些步驟可以了解一部詞典編輯的過程。筆者將過去編輯紙面版的經驗，結合資料庫管理的觀念，利用電腦將成果編出。後來網路版及光碟版的開發，證明當初的選擇是正確的。只有數位化的檔案才能靈活地做不同的利用。在完成這項工作後，筆者在《國語辭典簡編本》(<http://140.111.1.22/cle/jdict>)的編輯上面，更嘗試將聲音檔及圖片檔結合進來，豐富了中文詞典的內涵。這也使得這兩部詞典在近幾十年的詞典編輯觀念上，有了一些指標性質的成績。(註2)

三、異體字字典的編輯經驗

筆者在國語辭典系列工作後，曾參與了教育部《異體字字典》的編輯。這是一部以蒐羅字體字形為主的語文工具書，於民國八十四年七月開始編輯，八十九年推出網路試用版本，九十年六月推出網路正式版本(<http://140.111.1.40/>)，九十一年十月並發行光碟版本。資料庫總容量超過30G，是一部規模空前的字典，編輯作業也當是空前的繁瑣。筆者在這部字典的編輯流程規劃上，除基本檔案須於紙上作業外，其餘完全於線上完



成。流程之複雜遠超過《國語辭典》系列的編輯。下列試舉其中幾項的流程為例：（註3）

（一）總流程

〔上游收發：形體表製作、填注正字屬性〕→〔初審：收錄異體字形〕→〔上游收發：填注異體屬性、發稿〕→〔中游收發：交小組撰稿及委員研訂〕→〔小組整理委員稿〕→〔中游收發彙整委員意見，製作委員會資料〕→〔委員會議〕→〔小組清稿〕→〔編輯成果〕

在整個流程中，負責資料收發的人也就是資料庫的建構與管理者。他們是整個工作的總樞紐。在過去筆者負責的編輯工作，都只設一組「收發」，此次設了兩組，分享上游的基礎資料和中游的過程資料，而且把整體工作分為上中下三層，將撰稿人員當成下游。在這種管理模式中，「上游」作為基礎資料的總管，有如整體工作的播種者；「中游」作為體例、進度、品質的掌控者，有如整體工作的耕作者；「下游」作為成果的完成者，有如整體工作的收成者。每個階段各設工作小組，各訂工作原則，但整體進度須聽從中游的監控指揮。如此一來，每個階段把問題解決，相互承接，相互監管，讓成果編輯問題儘量減少，而且人力運用可以發揮最大效果。

（二）成果網頁編輯流程

A 正字釋義部分：〔正字釋義〕→〔資料庫轉成 word ㄩ書檔〕→〔排版〕→〔轉成 htm 檔〕→〔加注對應訊息及連結點〕

B 異體字部分：〔關鍵ㄩ獻、研訂說明及附錄字清稿資料庫加注 html 語法〕→〔利用程式轉成 txt ㄩ書檔〕→〔副檔名更名為.htm 檔〕

C 形體資料表部分：〔形體資料表資料庫建檔〕→〔罕用字引書體例建檔〕→〔利用程式加注 html 語法並轉成 txt ㄩ書檔〕→〔副檔名更名為.htm 檔〕

D 索引部分：〔建立部首筆畫、形根及四角號碼等屬性資料庫〕→〔利用程式將形根轉成筆順序〕→〔資料庫排序〕→〔驗算筆畫數〕→〔利用程式加注 html 語法並轉成 txt ㄩ書檔〕〔副檔名更名為.htm 檔〕

（三）形體表編輯流程

A 〔形體表掃描〕→〔形體表切割〕→〔建立資料表〕→〔校對〕→〔結合至成果〕

B 〔守形轉寫發稿〕→〔守形謄寫收稿〕→〔守形校對〕→〔掃描〕→〔切割〕→〔結合至成果〕

C 〔ㄩ獻守形切割〕→〔結合至成果〕

D 〔罕用字形下載〕→〔結合至成果〕

這部字典的工作十分繁雜，一共規劃了七個流程。每一個流程就像一條生產線，由收發組統一指揮，運作流轉。由原始資料進而編輯成成果網頁，環環相扣。如果規劃得宜，往往可收事半功倍之效。例如本字典實際全程參與的專任編輯只有十四名，其餘都是兼任的人力。六年中，以一百萬資料單位計，平均每位專任編輯要處理七萬多個資料單位，每年要處理一萬多個資料單位，每個工作天要處理約六十個資料單位。處理結果包括形體表、釋義內容、異體字形、研訂內容、連結點、網頁等等的完成，事實上幾乎是不可能的任務。所以這其中藉用了許多「工具程式」，例如固定內ㄩ的填注、固定連結點的設定、固定網頁的編寫等等，都是利用程式來完成。在套裝軟體方面，編輯過程中只運用了如



foxpro、frontpage、photoimpact、excel、word、dbase 等通行的軟體。(註4)

四、網路語典編輯要點綜述

從以上兩部網路字詞典的編輯介紹，大概可以了解，在編輯過程中有幾個要件是必然存在的。拙作〈字辭典編輯經驗談〉一文中曾歸納如下幾點：(註5)

首先是「整體的需求」方面，當注意以下幾點：

1. 語彙整理需要本科的專才。
2. 辭典編輯需要跨科際的知識。
3. 成果展現需要新領域的技術。
4. 理想工具書需要環境的配合。

再來是「編輯的條件」方面，需要具有以下的觀念：

1. 編輯目標需要銳利的眼光。
2. 編輯態度需要脫俗的平凡。
3. 編輯精神需要長期的毅力。
4. 編輯模式需要資管的理念。

至於如何著手去編一部詞典，試以《國語辭典》的編輯工作為例，筆者以為當注意如下要點：

(一) 《國語辭典》編輯流程要點

1. 擬訂目標。
2. 設計體例。
3. 規劃流程。
4. 分析技術。
5. 安排人力。
6. 管理資料。
7. 編寫成果。

(二) 《國語辭典》編輯技術要點

1. 傳統編輯與電腦編輯的結合。
2. 編輯系統的開發。
3. 資管理念的發揮。
4. 編輯資料庫的建構。
5. 電子版成果的展現。

(三) 《國語辭典》編輯資料庫的建立要點

1. 基本檔案資料庫。
2. 參考檔案資料庫。
3. 參考書目檔案資料庫。
4. 引書體例資料庫。
5. 成果編輯資料庫。

如果拿《異體字字典》的編輯來看，那就更複雜了。尤其要考慮如何建構以下幾個資料庫，然後再將它們結合起來：

1. 字形資料庫。
2. 字音資料庫。
3. 字義資料庫。
4. 單字屬性資料庫。
5. 方言用字資料庫。
6. 民俗文獻用字資料庫。

《異體字字典》全書利用網頁語言編寫，連結點超過百萬，在編輯時也要一併考慮工具程式的運用。

五、發行人網路版語典可能遇到的問題

在電腦尚未利用至詞典編輯之前，詞典編輯主要靠著紙筆來編。利用卡片、剪刀、漿糊來蒐集資料。以筆者經驗來說，三十餘年的編輯工作，前面十餘年可以說是剪刀漿糊的歲月。像《大學



字典》收了近萬個單字，在編輯索引時，一套字頭卡片要從部首筆畫序變成筆畫部首序，一個編輯人力埋首於卡片的整理，大概要花上七個工作天。改用電腦建立資料庫，無論何種排序的索引，數分鐘完成。工作效益不可同日而語。在網際網路架構之前，詞典的成果只能利用紙面印刷，於是從手寫稿到打字條稿，進再到拼成機樣稿，照相排版裝訂，不但費時費錢，而且印本有限。網路版則不同，編輯完成，找部伺服器，立即向全世界公開，展現力量完全不同。而且將資料數位化後，文字、影像、聲音，蒐羅殆盡。資料內容也可以隨時維護。如此看來，傳統紙面版本該可休矣，往後古口具書的世界當只有電子版本天下！自《國語辭典》網路版推出至今，使用人次已逾六百萬，每天數千人的利用，的確證明了網路版本的力量。

但是，這幾年來，筆者在回覆眾多讀者來信的問題後，也累積了一些想法。這些想法當是國內網路版詞典初次上路的珍貴經驗，所以願意在此提出來供開發網路版本的編輯者作參考。

（一）資料印象凝固的問題

過去將資料印成書面，無論刷幾版，印數有限，使用者也有限。而且平面資料翻查不易，尤其是具關聯性質者，除非編者註記，否則難見線索。詞典使用者，往往是讀書遇有疑難者，問題討論範圍自然有所侷限。但是網路版則不然。一經上網，立即成為網路世界一員，使用者幾無限制，而且利用搜尋引擎，關聯全。這本來是網路版優勢，但是卻也產生一個問題。那就是資料一經公開，使用者既是全面性的，價值觀念自也是參差不齊，於是詞典收詞、釋義、舉例，不再

只為讀書疑難者服務，可能得接受網路各界的檢視。結果勢必造成問題紛雜，諸多非關學術。加上網路資料，除非卸除，否則終將凝固於數位環境中。此點猶如報紙資料一經上網，則某人一時犯錯，縱然早已服刑完畢，但犯罪訊息卻永遠留存網路，一經查尋立即呈現，有如新犯。所以面對網路環境的特質，詞典編輯如何古專業化和社會化尺吋中拿捏，將接受更嚴格的考驗。

（二）資料不容錯的問題

網路版的詞典幾乎不容錯誤。但是詞典出於眾人之手，或者體例不一，或者釋義輕重有別，或者資料互見參差。在過去紙面版，此種疏誤較難發現，但如今結合全。檢索，相關資料，一眼看盡，無可隱藏。於是使用者上網，發現錯誤，一經反應，往往會有馬上更新的期盼。原因古於網路給人一種即時的、最新的印象。但是詞典正誤，為求慎重，自有流程，無法同步。如果置之不理，拖延時日，終將造成詞典權威漸失。所以要在網路環境中呈現詞典編輯成果，品質恐要有更嚴格的要求，而且建立立即反應，隨機答詢的機制也是和過去紙面版不同的。

（三）網路習慣配合的問題

配合網路環境使用習慣，可從兩方面來談。從正面來看，網路使用習慣包括系統、技術的新潮流，也包括使用者上網欲求直接、快速、豐富的企圖心。於是將性質穩重的口具書置上網路，古成果包裝的設計上是需要多費心思的。過於花俏，易流於輕浮；過於保守，則又違反潮流。系統老舊，會被淘汰；系統太新，又難照顧到所有平臺。這些都是過去紙面版不會遇到的問題，雖



非關學術，但卻是切合網路環境與否的關鍵。古另一方面，網路環境有著脫法、快速、隱匿的幾個特性。古網路遊走的駭客，藉此特性，縱橫這個虛擬的世界。於是古網路上公布成果，須作許多保護，尤其資料備份更形重要。一遭破壞，立即遞補上綱。有些損壞可能來自本身軟硬體的問題，所以整體軟硬體須有維護上的縝密考慮。到底電子設備並非絕對可靠。詞典一經上綱，建立信譽後，是不宜中斷的。

(四) 持續更新的問題

一部工具書的生命要長久，持續修訂是必需的。紙面版如此，網路版更應如此。但是一般編輯的習慣是書編完後，工作群解散，待下次修訂，再行組成。這古過以紙面版，長久修訂一次或許可以，對網路版本來說恐有不足。前面說過，古使用者對網路版有立即更新的期盼下，成立一個常態維護小組及開發線上維護系統是重要的。對於一般屬於詞典的明顯缺欠或急待補充的資料，就由小組透過維護系統來補正。如此一來，對這部詞典的工具功能可保持一定的水準，並能充分發揮網路即時反應的特色。小組要訂下一套維護的標準作業模式，持續地從不同的角度來檢測資料庫及成果系統，並吸收新的資訊，逐漸古既有基礎上累積新的成分，以作為下一次大規模修訂的依據。

(五) 立即消失的問題

網路版是將資料庫及運作系統置放於伺服器上，立即公開，立即使用，更新時亦然。但是反過來說，只要將資料從伺服器上卸除，或直接将伺服器的電源關掉，則原有詞典也會立即消失，

從此古網路上了無蹤影。這點和紙面版有很大的差異。紙面版一刷千套，除非聚古一起，一把火燒盡，否則總會古某個圖書館書架上發現蹤影。所以網路版本實存有著立即消失的可能性。這種可能性包括了政策的變更，負責單位的裁撤等因素。過以古書傳鈔，尚有海內孤本之說，一部網路詞典的消失，極可能連孤本都不復存古。這正說明了，要編輯網路版詞典時，不得不以為這部詞典的永續生命預作制度考慮。也許配合著光碟版的發行，也許配合著有條件的資料庫開放都是可行的辦法。

(六) 中文系統相容的問題

中文系統相容和收字不足的問題，向來是電腦編輯的困擾。於是造字檔依然存古。古此方面，筆者曾使用了幾個方法：

1. 下載造字檔安裝
2. 將造字檔轉成圖片
3. 手寫圖片直接顯示

各有優缺點。至於系統方面，從早期的MSDOS+ET 中文到微軟視窗 31、95、98、me、2K、XP，讓人感覺有一點追著微軟趕路。如果再考慮麥金塔或 linux 環境，則問題越來越複雜。古網路版上解決這些問題還較簡單，但光碟版上的考慮，可能不那麼容易。也許將兩個版本環境的一致，把網路環境直接移置光碟是個可以選擇的方向。除此之外，還有繁簡的困擾。這些都是電子版本古成果展現上不得不注意的問題。

(七) 網路試刊版的問題

網路版詞典古公開之前，比較妥當的作法是先推出試用版本。這種作法有兩個好處：一則藉



此檢驗原來的設計是否適切。因為無論事如何考慮，總缺少「現場實況」，一經上網，往往可以看出原來無法察覺的缺失。二則是以凡徵求使用者意見，藉以集思廣益，讓正式版本更臻理想。因為編者是無法完全代替讀者來看事情的，而且當詞典編完後，詞典的主人事實上已轉為讀者了。至於試用版推出的時間點，宜納入整體工作來作規劃，也因此古工作初期就要有整體系統環境運作的考量。

六、結語

網際網路的世界是個新世界，要在這個新世界開疆闢土，靠的不是船堅炮利，而是一個個強大的資料庫。詞典就是資料庫，品質優良的詞典就是強而有力的開疆闢土者。詞典的編輯，可藉一人之力成就，卻更適合群策群力。網路詞典有著資料呈現快速，眾所共享的優勢，所以力求完美是編輯上唯一的選擇。集結眾人智慧，適切規劃，逐步漸進，聚沙成塔，則各領域有各領域的成就，各學門有各學門崢嶸頭角的光輝。這其中的關鍵除了要有環境的配合外，更要有股恆毅的堅持。筆者曾呼籲國內詞典環境要快成熟起來，在此環境中，會具有如下的條件：（註6）

1. 各種相關語料的普遍整理。
2. 各種相關學理的普遍探討。
3. 各種類型及不同主題詞典的普遍編輯。
4. 各階層人士的廣泛使用。
5. 國家政策的有力支持。

在此環境中，有人能了解詞典多樣編輯的重要，有人願意長期從事編輯工作，也有人能認同詞典水準反映出全民知識水準。要待此環境成熟

後，薈萃各種理想條件的工具書才會普遍出現。

筆者從事詞典編輯工作時間雖長，仍在學習中。利用本文，把一些過山勇於嘗試的經驗提供出來，也許不夠成熟，卻當可算是個人對成熟詞典環境的期盼所盡的一份心力。

【附註】

- 註 1：參《重編國語辭典編輯總報告書·流程編》，教育部國語文教育叢書 20。
- 註 2：參曾榮汾〈試論國內語文工具書編輯觀念之成就〉一文，2003 年，訓詁學會論叢。
- 註 3：參《異體字字典編輯總報告書·流程編》，教育部國語文教育叢書 46。
- 註 4：較詳細的介紹可參拙作〈教育部異體字字典之析介〉，2003 年，文字學會論叢。
- 註 5：刊載於《師生論壇》，第一輯，2003 年 5 月，集科大漢學資料整理研究所。
- 註 6：同註 2。

參考著作

1. 曾榮汾，《辭典編輯學研究》，台北市：世界文物出版社，民 78 年。
2. 教育部，《重編國語辭典修訂本編輯總報告書》，台北市：教育部，民 86 年。
3. 教育部，《異體字字典編輯總報告書》，台北市：教育部，民 91 年。
4. 曾榮汾，〈試論國內語文工具書編輯觀念之成就〉，訓詁學學會論叢，民 92 年。
5. 曾榮汾，〈教育部異體字字典之析介〉，文字學會論叢，民 92 年。
6. 曾榮汾，〈字辭典編輯經驗談〉，《師生論壇》，1 輯（民 92 年 5 月），頁 94-100。

