

中国文字研究

ZHONGGUO WENZI YANJIU

教育部人文社会科学重点研究基地

华东师范大学中国文字研究与应用中心 编

广西教育出版社



第四辑

上海市政府重点学科基金资助

中國文字研究

(第四輯)

教育部人文社會科學重點研究基地
華東師範大學中國文字研究與應用中心 編

廣西教育出版社

中国文字研究

第四辑

教育部人文社会科学重点研究基地
华东师范大学中国文字研究与应用中心 编

☆

广西教育出版社出版

南宁市鲤湾路8号

邮政编码:530022 电话:5865797

本社网址 <http://www.gep.com.cn>

读者电子信箱 master@gep.com.cn

全国新华书店经销 广西民族印刷厂印刷

*

开本 890×1240 1/32 8.75印张 340千字

2003年12月第1版 2003年12月第1次印刷

印数:1-1000册

ISBN 7-5435-3771-0/H·92 定价:35.00元

如发现印装质量问题,影响阅读,请与承印厂联系调换

《中國文字研究》(第四輯)的說明

1. 本刊由國家教育部人文社會科學重點研究基地華東師範大學中國文字研究與應用中心主辦。作為中心的專業學術期刊,將嚴格遵循國家教育部關於重點研究基地創辦發行學術刊物規定的原則。

2. 本刊集中刊載中國文字學研究與應用的最新成果,竭誠為海內外文字學界投稿者服務。來稿的文獻數據請按國際標準期刊規範處理。來稿涉及古文字請集中處理,並提供軟盤和文本。由於本刊實行匿名評審制度,請將有關作者的通訊姓名、工作單位等背景數據另頁單獨填寫;審稿處理意見一般分如下三種形式:(1)全文發表;(2)修改發表;(3)不宜發表。

3. 本刊編輯部通訊地址:

200062 中國上海市中山北路 3663 號

(校內文科大樓 13 樓 1301)

華東師範大學中國文字研究與應用中心)

電話:021-62232050

傳真:021-62232050

網址:www.sinogram.com.cn

E-MAIL:wenzi@ecnu.edu.cn

目 錄

搭建中華字符集大平臺	李宇明(1)
關於異體字整理和《規範漢字表》的研製	王鐵琨(9)
中國正統文字的發端	
——殷商甲骨文字在中國文字發展史上的地位	
.....	李 園(17)
字符・結構	臧克和(30)
二十世紀文字改革爭鳴綜述	詹鄞鑫(34)
商王武丁的末期:中國上古年代學的重構實驗	
..... [美]夏含夷著 張德劭譯(51)	
甲骨文考釋方法的統計分析	張德劭(77)
《甲骨文合集補編》第一冊釋文商榷之一	
——與沈培、謝濟二先生的商榷	劉海琴(84)
《甲骨文合集補編・釋文》刊誤	沈之傑(92)
商周時期金屬稱量貨幣的自名名稱及其嬗變	趙平安(103)
王獻唐重要遺稿《國史金石志稿》整理劄記	王文耀(108)
西周金文字素表聲功能二題	張再興(111)
西周金文中的賓語前置句和主謂倒裝句	潘玉坤(116)
從青銅器銘文看商周的祭祀活動	秦建文(121)
《魚鼎匕》銘文性質及“下民無智”的有關問題	
..... [德]史克禮(130)	
關於商周金文中的“賜貝”及與貨幣相關問題	陶霞波(136)
敦煌懸泉漢簡詞義劄記	徐莉莉(141)
從出土文獻書跡看“捺”的成因	劉志基(145)
郭店簡考釋二則	徐在國(148)
釋包山楚簡中的“受期”	蘇 杰(151)
讀楚簡偶識	吳建偉(156)
石鼓文“會聲字”調查	吳 昊(158)
古璽文字中的省簡字初探	朱 疆(165)

《說文》諧聲偏旁補正·····	侯占虎(170)
《說文解字》大徐反切今音標注存在的問題·····	王 平(175)
《漢語大字典》引《說文解字》注音辨正 ·····	蔡夢麒(182)
《說文》《玉篇》《萬象名義》部首之比較分析 ·····	郭 瑞(191)
大徐本《說文解字》中 UCS 所無字初步調查分析 ·····	樓 蘭(194)
《三國志》難字獻疑·····	吳金華(198)
《玉篇》讀剗·····	朱葆華(205)
略論《玄應音義》在文字學研究上的學術價值·····	徐時儀(209)
《名義》義項與其字頭同例考辨·····	呂 浩(218)
《龍龕手鏡》俗字雜考·····	蔣冀騁 鄭賢章(223)
“𪚩”字考釋	
——兼及𪚩、𪚩、𪚩·····	劉中富(228)
瑪麗瑪莎文兩次調查所得單字的比較及其文字學意義	
·····	王元鹿(231)
建立東巴文象形字典語料庫的構想 ·····	甘 露(234)
文字類型學研究的意義、現狀及設想	
——兼談各民族文字資料庫的建立給文字	
類型學研究帶來的契機 ·····	朱建軍(240)
試論假借表達與假借字的區別 ·····	黃亞平(246)
常用形聲字聲符表音度分析和識字教學 ·····	竺海燕(251)
試論簡化字中的聲旁混同現象 ·····	夏 軍(261)
“交綬”“死綬”考辨 ·····	汪少華(264)
從“曰若稽古”說開去·····	李春曉(270)
關於古代漢字字際關係的確定	
——以“顧”及相關字為例·····	黃德寬(275)

搭建中華字符集大平臺*

教育部語言文字信息管理司 李宇明

一、中國漢字處理的簡單回顧

計算機語言處理的內容非常廣泛，漢字處理是其中的重要內容之一。20 世紀 50 年代，中國就有先行者開始研究俄漢機器自動翻譯問題，並採用當時的電報碼或四角號碼充當漢字編碼。^[1]70 年代，漢字的計算機處理問題開始受到重視，鍵盤編碼、漢字的點陣描述和輸出等，成為當時的“必戰之役”。^[2]80 年代中期，國家站在時代前沿，正式把語言文字的信息處理納入國家語言文字工作的重要內容。在 1986 年 1 月召開的全國語言文字工作會議上就提出：“漢語漢字的信息處理是一門新興的邊緣學科，有廣闊的前景，加強這方面的研究，對經濟、文化、科學技術的發展具有長遠的意義。因此，當前語言文字工作的任務必須包括這項內容。”^[3]

我國計算機漢字處理，近 30 年來獲得了重大成就。^[4]陸續開發的漢字鍵盤輸入方法，解決了漢字進入計算機的難題；漢字輸出實現了多字體^[5]、多字號；漢字字庫的製作也由點陣字庫、矢量字庫逐步過渡到曲綫輪廓字庫；漢字自動識別技術達到國際先進水平，並有商品投入市場；ISO/IEC10646 的 CJK 字符集^[6]，由早期的 6763 個通用漢字逐步擴充，擴充集 A 和擴充集 B 已經完成，現在正討論擴充集 C₁。待擴充集 C₁ 完成後，中日韓編碼漢字已經 7 萬有零。^[7]這些成就，促進了中國計算機的普及，為發展計算機應用技術和信息化創造了基本條件。

但在計算機漢字處理方面，還存在一些需要解決的問題。例如：

1. “萬碼奔騰”的局面，令業內人士深感擔憂。隨著計算機語言處理由“字詞”向語言理解的深度進展，當前提高鍵盤輸入的質量，已不僅僅取決於編碼的本身，而主要取決於與語言理解有關的各種資源的建設，如深度開發的各種語料庫，各種電子詞典、語法典的編纂等。用這些資源支持鍵盤輸入，纔

* 本文曾在第二屆肯特崗國際漢語語言學圓桌會議（2002 年 11 月 26—29 日·新加坡）上宣讀，並得到厲兵、張書岩、陳雙新先生的指教。此文曾發表在《中文信息學報》2003 年第 2 期。這次發表又做了一些修改。

能使鍵盤輸入出現跨越式的發展。

2. 怎樣進一步提高漢字識別質量,特別是脫機手寫體識別的突破。

3. 語音與文字的自動轉換,包括語音輸入的文字顯示和文字的語音輸出,應用前景廣闊,但是當前還存在著較多的技術難題,需要重點攻關。漢字與漢語拼音、國際音標等的自動轉換、標注,在技術上雖然難度不大,但並沒有引起軟件製作者的重視。

4. 漢字簡繁體的自動轉換還沒有完全實現自動化,特別是由簡體向繁體的轉換,尚需較多的人工干預。

5. 各組異體字之間的關聯亟待實現。

6. 宋體、仿宋體、楷體、黑體這四種常用印刷字體的字形標準應擴展到更大範圍,而且還應建立其他字體的規範或規範原則,應建立掌上電腦等小型計算機上使用的低點陣字形的省減規範。此外,還應研究手寫體及書法藝術的計算機表現問題。

除了以上問題之外,漢字處理方面還需要考慮的就是建立中華字符集的問題。本文討論中華字符集的內容、建立中華字符集的意義、需要解決的技術問題等等,希望引起大家對這一關係到中華文化在信息時代發展的宏大事業的關注和支持。

二、中華字符集及其意義

20世紀是人類科技大踏步前進的世紀,計算機和國際互聯網,應該是20世紀人類最重要的發明之一,特別是國際互聯網,給人類構造了一個全新的虛擬空間,展示了信息時代的光明前景。但是,計算機和國際互聯網的出現,也在全世界範圍內形成了巨大的“數字鴻溝”,帶來了各種新的國際性的社會問題。現代國家在制定語言文字規劃之時,不能不考慮計算機的語言文字處理問題,不能不考慮網絡上的語言文字交際問題,不能不考慮怎麼樣通過語言文字的信息化來縮小數字鴻溝的問題。

互聯網不僅是技術的,而且更是文化的。克服“數字鴻溝”,不僅需要技術,而且也更需要文化。文化是人類文明的積澱,是人類生活的底蘊。文化的傳承與開發,是人類生存與發展的重要方面。以計算機和互聯網為基本技術支撐的數字化信息時代,文化的傳承與開發更加重要。中華文化是由多民族文化構成的源遠流長的燦爛文化,在世界文化發展史上具有不可替代的重要性。促進中華文化在信息時代發揮更為重大的作用,是我們的歷史責任。

承載中華文化的文字與符號的總和,稱為“中華字符集”。中華字符集大致包含如下九個方面的內容:

1. 漢字發生隸變之後的記錄主流文化的今漢字,包括簡體字、繁體字、傳承字和異體字等。^[8]

2. 方俗漢字,包括漢語方言用字、古今民間使用的俗字以及碑別字、錯訛字等。

3. 古漢字,包括甲骨文、金文、戰國文字、簡帛璽印字、小篆及漢字隸變之前的其他文字。

4. 漢字系的古今少數民族文字,如古壯文、布依文、侗文、毛南文、仡佬文、老苗文、瑤文、白文、哈尼文、西夏文、女真文、契丹大小字等。^[9]

5. 古今少數民族非漢字係的表意文字,如納西的東巴文、水文、爾蘇沙巴文等。^[10]

6. 古今少數民族的表音文字,如蒙文、藏文、維吾爾文、哈薩克文、朝鮮文、彝文、傣文(數種)、滿文和柏格裏苗文、八思巴文、佉盧文、回鶻文、察合臺文、突厥文、于闐文、焉耆—龜茲文、粟特文等。

7. 記錄中國古今語言(普通話、方言、古代漢語,民族語)所需的國際音標,以及為中國語言所設計的各種拼音符號等。^[11]

8. 文字(主要是漢字係文字)的部首、筆劃、偏旁及其他構件。

9. 具有文化意義的其他各種符號,如西安半坡出土的陶器上的符號、64卦圖符、道教符號、古代數學史和音樂史等涉及的符號。

建立中華字符集,在信息化時代,對於中華文化的保存和發展具有極為重要的意義。

第一,以中華字符集作平臺,中華古今文獻在電子世界中,就可以有一個可進行文字加工處理的本面目的永久保存本。如果沒有中華字符集,中華文獻(特別是古文獻)在電子空間的保存,只能採取圖像方式,不能進行文字加工;或者將文獻轉寫為現在字符集(比如擴充 C₁完成後的 7 萬餘字的範圍內)中的字,但是原來的文字面貌失去了,相應地也就失去了許多語言文字信息和文化信息。

第二,為數字化圖書館、博物館、檔案館的建設提供基礎。20 世紀 90 年代以後,美、英、法、德、日和俄羅斯等近 20 個國家和地區相繼投入巨資,開展數字圖書館研究。據報導,美國投入到數字圖書館研究的經費已經超過 8 億美元;俄羅斯在經濟尚未全面恢復的情況下,政府計劃從 1999—2004 年,每年出資 2 億盧布支持數字圖書館研究;日本不僅投入了 15 億日元開發日文文獻數據庫,而且還以建設國會圖書館分館——關西電子圖書館為契機,投入 4 億美元,擬將其建設成為亞洲文獻中心。我國隨著電子政務、電子商務和遠程教育的發展,數字化圖書館、博物館、檔案館的建設已經啓動,並制定了《中國數字圖書館工程建設一期規劃(2000—2005)》。^[12]但是要看到,漢字的特點決定了沒有中華字符集,中國的高水平數字化圖書館、博物館、檔案館的建設,是不可能實現的。

第三,為用於知識發掘的數據庫建設提供堅實的平臺。智能計算機的發展,依賴於各種可進行知識發掘的數據庫,通過數據庫中知識的發掘來增強計

算機的智能。要使計算機“精通”中華文化，必須建設龐大的可供計算機“學習”的中華文化數據庫。科學研究，特別是人文科學研究，數據庫逐漸會成為主要的研究手段，通過文獻數據庫檢索文獻信息，通過素材數據庫統計數據、發掘知識、總結規律、驗證結果。中華字符集使得用於知識開掘的中華文化數據庫的建設成為可能。

三、中華字符集的實現

中華字符集，是自計算機對文字進行處理以來人類所提出的最為龐大的字符集，在實現中必然會遇到許多技術難題，必須認真研究，妥善解決。

第一，字符種選擇。這裏所謂的字符種，包括需要進入中華字符集的字種和各種符號。中華字符集的字符種，涉及到古今漢字學、古今民族文字學、語音學、方言學、民俗學、文獻學及其他眾多學科，必須廣泛動員各相關學科的學者，遵照“有見必收”的原則，全方位地拉網式地進行搜集。在字符種搜集過程中，要充分利用已有的各種字典和相關研究成果，利用已有的古文獻庫和專業語料庫、文獻庫，同時對散佈在民間的俗字和一些少數民族的文字，還需要進行田野訪查，爭取最大限度地搜集字符種。

在廣泛搜集的基礎上，還需組織專家對這些字符種進行甄別、查重(chóng)、篩選，最後擇定。

第二，形體規範。對選定的字符種，要根據不同文字系統和符號的特點進行形體規範，給出典型的或標準的字形或符號形體。被選定的字符種中，有許多沒有印刷體形式，如一些民間俗字、碑別字、錯訛字，一些敦煌變文俗字、地下出土的竹帛字，一些古代的少數民族文字，等等。這就需要將手寫字體“印刷體化”。對於甲骨文、金文、篆文等古漢字，不僅要給出符合這些形體規範的字形，而且可能還需要隸定楷化，並在隸定楷化前後的兩種形體之間建立關聯。

第三，造字。根據字符種的形體規範，造出適合計算機字庫使用的形體，並輸入計算機。如果條件允許，還應逐漸造出不同字體的字。鑒於造字量大且技術複雜，應當設計出便捷的計算機造字軟件，藉助計算機完成造字任務。

第四，存儲。中華字符集具有數量大、非勻質的特點。中華字符集的非勻質性表現在：1. 它是多種文字系統和符號的集合，在符號性質上具有非勻質性；2. 字符的使用範圍和使用頻率上有巨大差異，有的是現代語言生活所必需，有的只在特殊的領域中使用，在使用上具有非勻質性。

因中華字符集字符種數量巨大，現有的音序、形序等排序方法，都無法適應。以何為順序將字符種排列存儲起來，是需要研究解決的。因中華字符集的非勻質的特點，必須考慮將全字符集分為若干子集，以滿足不同領域、不同人群的使用需要。子集的劃分可考慮三大原則：

1. 民族原則。某一民族的文字或某些形體來源有關係的民族文字盡量結成一個子集。如藏文應當結成一個子集；維吾爾文、哈薩克文、柯爾克孜文因文字來源關係應當結成一個子集；漢字同漢字系的民族文字應當結成一個子集，等等。

2. 古今原則。古今文字的用途往往有較大差別，比如隸變之前的古漢字，主要供專業人員使用，一般人很少涉及這些文種；隸變之後的今漢字，有許多只在古籍中使用，現代交際中不使用或很少使用。所以，在依照民族原則劃分的子集中，往往還需要以古今再劃分子集，當然，像西夏文、女真文、契丹大小字等今天已經不具交際活力的文字，不需以古今原則劃分子集。

3. 字頻原則。字符集應當考慮依照不同的使用頻率再劃分子集。據華東師大中國文字研究與應用中心研究，西周金文，字頻在 10 次以上有 612 字，僅占總字數的 19%，但是這些字的總字頻卻達 65792，占西周金文字頻總量 69289 的 92%。西周金文中，字頻在 100 以上的常用字，已識字的比例為 99.2%，已識字比未識字的使用率高。這說明古文字需要考慮字頻問題。今天仍然使用的今文字，特別是非拼音文字，有些字符種已經退出現代交際，只具有備用性質，使用頻率很低，應建立備用子集。現代交際中還在使用的字符種，如現代語言生活中使用的漢字，還應當再以字頻所形成的級次劃分子集。

這些具有並列關係、不同層次包容關係和在不同層次上形成交叉關係的各種子集，便形成中華字符集錯綜複雜的內部結構。根據這種內部結構來處理中華字符集的儲存問題，有利於字符集的應用。

第五，檢索與輸入。運用傳統的檢索、輸入方法難以適應中華字符集。傳統的檢索、輸入方法主要有語音方法和形體方法。運用語音方法進行檢索和輸入難以奏效，因為：

- a. 字符集中同音字太多；
- b. 字符集的語音系統不統一；
- c. 有許多字符有形無音。

運用形體方法進行檢索和輸入也困難重重，因為：

- a. 字符集形體系統差異顯著；
- b. 構件切分不具統一性；
- c. 部首難以起到統領作用。

檢索和輸入是使字符集發揮作用的最為重要的兩項操作，必須集中精力研究準確、快捷的檢索方法和方便的輸入方法。

四、中華字符集與國際標準

互聯網是人類信息交換的最先進的手段，它所構造的虛擬世界，將成為人類新的十分重要的交際空間。互聯網的交際活動，有各種即時的互動交際和

網絡出版等等。在互聯網已經將整個世界勾連起來的今天,建立中華字符集必須考慮互聯網的交際問題,甚至說首先應考慮到互聯網的交際問題。

如果不考慮互聯網交際的話,中華字符集的內碼可以有多種,即可以允許每個公司的產品有一套自己的編碼系統,或是國家間、地區間使用不同的編碼系統。這樣的技術路綫導致的必然結果是,涉及中華字符集的電子產品,必須攜帶一個龐大的專用字庫;用戶使用這樣的電子產品時,必須同時裝載專用字庫,或是通過編程的方式建立某種特定的字庫間的映射關係。計算機存儲空間的急劇增大和運算速度的快速發展,允許採用這樣的技術路綫,但是其弊端也十分明顯:

1. 增加了產品的開發成本;
2. 增加了用戶下載字庫或在多個編碼系統之間來回轉換的麻煩;
3. 網絡交換的信息不穩定,使用戶常受亂碼干擾之苦。

爲了克服網絡交際中的亂碼、空碼等問題,保證信息傳輸的穩定性,實現網絡的無障礙交際,爲了減去自帶字庫、多個編碼系統來回轉換的麻煩,降低開發成本,增加使用的便利,中華字符集的內碼必須走國際化的技術路綫,實行國際統一編碼。

國際統一編碼,會使中華字符集具有兩大特性:1. 中華字符集的每個字符種的代碼是惟一的;2. 全世界只使用一套統一的編碼系統。

ISO/IEC 國際標準化組織,爲世界各種文字和符號提供了 $128 \times 256 \times 65536$ 個字符的足夠大的空間。這些空間分爲 128 組(0 ~ 127 組),每組由 256 個平面(0 ~ 255 平面)構成,每個平面可以存放 65536 個字符。當前已經定義了第 0 組的 17 個平面(0 ~ 16 平面),其中 0 ~ 14 平面用來存放字符,15、16 兩個平面留作特殊用途。第 0 組的第 0 平面是多文種平面(BMP - basic multilingual plane),當前這一平面已基本用完^[13],字符集再擴充就需要啓用其他平面。ISO/IEC/JTC1/SC2/WG2/IRG(表意文字)工作組,負責中國、日本、韓國等國家和地區使用的漢字的國際編碼,如前所述,即將完成的編碼漢字已逾 7 萬。此外,ISO/IEC10646 已經建立了藏文、維吾爾文、朝鮮文、四川彝文等的國際標準,蒙文、傣文的國際標準即將通過。這 7 萬多漢字和藏、維、朝、彝等民族文字,已經能夠滿足日常語言生活交流和一般漢文古籍的出版,功莫大焉。^[14]但是,在此基礎上要完成中華字符集的國際編碼,並使其成爲國際標準,還有大量的工作要做。

五、餘 言

教育部語言文字信息管理司根據許多專家的倡議提出建立中華字符集的設想,該設想受到了語言學界、信息學界的專家學者和社會有關的部門、行業、公司的大力支持。2002 年 3 月 15 日,教育部語言文字信息管理司在北京召

開了國際標準“信息技術通用多八位編碼字符集”會議，向一些部門、行業徵集“集外字”。7月13—16日在哈爾濱召開了“全漢字庫暨國際音標問題商討會”，研究建立中華字符集的必要性、緊迫性、可行性及其操作思路。7月26日和9月2日，在北京兩次召集了少數民族語言文字學家座談會，專門研究少數民族古今文字進入中華字符集的問題。10月30—31日，邀集大陸、香港、澳門、臺灣等中國四方專家和官員在福州開會，溝通想法，在進一步討論擴充集C₁問題之外，初步討論了中華字符集的子集及以後四方協調的有關問題，達成了共識，並成立了兩岸四地中文數字化合作論壇（Chinese Digitization Forum，簡稱CDF）。中華字符集的建設工作已經起步，正進行字符種收集和子集設計，並已將中華字符集意向及設計方案提交有關的國際會議。

此外，2001年5月，國家語言文字工作委員會再次立項研製《規範漢字表》^[15]，可望於2003年完成初稿並向社會公開徵求意見。課題組在研究過程中，還分別在上海、井岡山、合肥、煙臺召開了四次專題性的學術討論會^[16]，2002年12月26日又在北京聽取了有關專家的意見。《規範漢字表》包括現代漢語用字、人名地名用字、科技用字等，可以滿足現代語言生活中漢字交際的基本需要，解決現代漢字生活中最基本的問題。根據使用頻率和使用領域，字表內部還要分不同的級次，並要進行讀音、字形的規範等。

《規範漢字表》也許可以看做中華字符集的一個最為基本的集合。^[17]一般漢字用戶，只需在計算機中使用《規範漢字表》字庫即可，特殊用戶可以採取在《規範漢字表》基礎上再掛上他所需要的中華字符集某個或某些子集，形成“1+X”的解決方案。

中華字符集規模宏大，難以一蹴而就，許多問題需要分期分批解決，而且需要多方的參與和支援。希望海內外人士關注這一問題，並為中華字符集的實現獻策出力。

注釋：

[1] 見傅永和《中文信息處理》，廣東教育出版社，1999年版，第51頁。

[2] 參見石雲程《語言信息處理的新任務》，《語文建設》，1986（1~2）。

[3] 劉導生《新時期的語言文字工作》，《語文建設》1986（1~2）。

[4] 詳見中華人民共和國教育部《跨世紀的中國教育》，高等教育出版社，2002年版，第254~259頁。許嘉璐《語言文字學及其應用研究》，廣東教育出版社，1999年版，第117~130、202~203頁。馮志偉《應用語言學綜論》，廣東教育出版社，1999年版。傅永和《中文信息處理》，廣東教育出版社，1999年版。

[5] 計算機漢字字體，現在已多達120餘種。

[6] C：中國；J：日本；K：韓國。

[7] 現在的提交稿為 70205 個漢字。

[8] 包括 1977 年開始試用、1978 年停止試用、1986 年正式宣佈廢止的《第二次漢字簡化方案(草案)》中的 248 個“二簡字”。

[9] 女書、傣僑文也可收進到此類字符中。這類字符大約有 25000 個。

[10] 這類字符大約有 17000 個。

[11] 包括漢語拼音、註音字母以及清末以來主要的拼音文字或拼音方案的符號形體。

[12] 資料來自 www.d-library.com.cn。

[13] 大約還有 1000 多個字符空間。

[14] 參見張軸材《國際標準編碼字符集與 CJK 的演進》，信息時代語言文字規範標準建設工作會議。(2002-09-23 武漢)上的學術報告。

[15] 該項目已被教育部列為 2003 年工作要點。見《中國教育報》2003 年 1 月 2 日第 2 版。

[16] 2001 年 12 月 21—22 日，在上海召開了漢字規範問題學術討論會。2002 年 5 月 16—17 日，在井岡山召開了異體字問題學術討論會。2002 年 6 月 22—23 日，在合肥召開了簡化字問題學術討論會。2002 年 8 月 22—23 日，在煙臺召開了印刷字體字形問題學術討論會。

[17] 中華字符集的基本子集，如果作為 CJK 的國際標準，還需要在《規範漢字表》的基礎上加上日韓的常用漢字。

參考文獻：

馮志偉. 應用語言學綜論. 廣州: 廣東教育出版社, 1999

傅永和. 中文信息處理. 廣州: 廣東教育出版社, 1999

胡鞍鋼, 周紹傑. 新的全球貧富差距: 日益擴大的數字鴻溝. 中國社會科學, 2002, (3)

李字明. 信息時代需要更高水準的語言文字規範. 術語標準化與信息技術, 2001, (3)

李字明. 信息時代的中國語言問題. 語言文字應用, 2003, (1)

劉導生. 新時期的語言文字工作. 語文建設, 1986, (1~2)

石雲程. 語言信息處理的新任務. 語文建設, 1986, (1~2)

許嘉璐. 語言文字學及其應用研究. 廣州: 廣東教育出版社, 1999

姚亞平. 中國計算語言學. 南昌: 江西科學技術出版社, 1997

張軸材. 國際標準編碼字符集與 CJK 的演進. 信息時代語言文字規範標準建設工作會議(2002 年 9 月 23—25 日·武漢)上的學術報告

中華人民共和國教育部. 跨世紀的中國教育. 北京: 高等教育出版社, 2002

周慶生主編. 國外語言政策與語言規劃進程. 北京: 語文出版社, 2001

關於異體字整理和 《規範漢字表》的研製^{*}

教育部語言文字信息管理司 王鐵琨

這次學術研討會人氣很旺。剛才的自由發言，有的談參會感受，有的提工作建議、意見，發言踴躍，言辭懇切，表達了對國家語言文字主管部門組織進行的異體字整理工作和《規範漢字表》研製工作的殷切期望，聽後深受鼓舞。同時，我們更感受到肩上責任重大。課題組決心博采衆家之長，與全國有志於《規範漢字表》研製工作的同志一道，把這項功在當代、利及子孫的好事情做好。國家語委的老領導、語言學前輩曹先擢先生為這次會議所作的學術總結，高屋建瓴，深刻透闢，切中要害，是一次精彩的學術演講，對我們扎實、穩妥地做好異體字整理和《規範漢字表》研製工作具有指導意義。曹先生一再強調決策要科學化、民主化。這幾天，曹先生在與課題組交談時也告誡我們：“問題要想得多一點，材料要準備得充分一點，考慮問題要周到一點，處理的時候要簡單一點。”這些意見都很中肯，是個重要的提醒。課題組要深入地思考、領會曹先擢先生的意見，並貫徹到字表研製的全過程中去。

借此機會，我談三個問題：1. 對會議的基本估價；2. 談談進一步提高對異體字整理工作的認識，中心想法是立足現實，把握形勢，與時俱進；3. 根據大家的建議，通報一下《規範漢字表》研製工作的進展情況以及下一步工作的打算。

一、對會議的基本估價

這次“異體字問題學術研討會”的召開是必要和及時的。經過兩天緊張而熱烈的研討，代表們都認為會議開得很成功，很有特色，取得了一定的成效

* 本文根據作者在異體字問題學術研討會（2002年5月，江西井岡山）閉幕式上的發言刪改整理而成。原題為《發揚學術民主，做好異體字整理和〈規範字表〉研製工作》。

和收穫。

會議的特色主要體現在三個方面：

一是會議目的明確，主題集中。確定一個大家興奮點比較集中的會議主題，是會議成功的一半。這次會議，討論的是學術界近年來較為關注的異體字問題，而且召開這次會議的直接目的是為編制《規範漢字表》服務的，這樣就把大家的興奮點和政府行為有機地結合了起來，學者們有話要說，不吐不快。儘管這是一次比較專門的會議，但從提交的論文和會議討論情況看，大家的參與意識還是比較強的。據初步統計，與會代表 30 人，提交論文提要 25 篇，報告論文 19 篇，這 19 篇中還不包括提交了論文但因故未能到會的遠在日本、上海、北京的 5 位先生的論文。一個小型的學術會議而能有 24 篇論文，內容涉及異體字的定義與整理異體字的原則、正異關係的變化與異體字整理工作、對《第一批異體字整理表》（以下簡稱《異體字表》）的評價以及對《規範漢字表》處理異體字問題的建議、結合辭書編纂的經驗和做法談異體字整理工作、整理異體字與海峽兩岸漢字字形的統一等，論題廣泛而相對集中。一些學者不約而同、不謀而合地選擇了相同的論題，比如異體字的概念界定問題，說明學者們的目光還是比較敏銳的。

二是研究比較充分，論文學術含量高。一些學者早就在關注異體字的問題，有備而來，準備充分。比如有的先生長期在比較大的字量範圍內較為全面地摸過異體字，總結出一些規律性的東西，其論文側重理論方面的深入思考探索。有的先生對《異體字表》前 410 組進行了測查分析，得出了若干有說服力的資料，分析了該表中存在的問題，進而提出了繼續整理異體字的一些想法。這種數理統計的科學方法在漢字的整理研究中是非常需要的。有的論文側重實際工作的歸納和總結，而實際工作中提出的問題又常常會使理論工作者受到某些啓發。有的論文既有理論的指導作用，又有應用實踐的意義；既有宏觀的闡述，又有微觀的分析（如一些典型字例的剖析等）；既注意到共時，又把目光放遠到歷時……我感到大家研究問題都很認真、細緻、深入，顯示出為參加這次會議作了精心準備。這正是本次會議的論文學術含量較高的一個重要原因。

三是會風好，會議研討氣氛比較濃。會議期間，代表們積極思考，發言非常踴躍，學術觀點有碰撞，研討有深度，從會場到餐廳再到宿舍，大家意猶未盡，討論不斷，談興不減。從專家們的發言中，課題組深受啓發，許多學術思想、學術見解和意見是花錢也難以買到的，而大家毫無保留地貢獻了出來，實在難能可貴。這種良好的會風應該提倡。學術上有不同的看法始終是正常的，學術上的不同看法需要通過平等討論來取得比較一致的意見。我們應該不惟上，不惟書，只惟實。所以在以後的簡化字問題學術研討會和印刷字形問題學術研討會上，以及將要召開的其他一些學術研討會上，我們特別希望各位專家能夠繼續踴躍發表意見，充分地討論，甚至有一些交鋒更好。這樣才能形

成良好的會風。課題組只有充分地看到、聽到了來自各個方面的不同意見，“博采衆家之長”才不是一句空話。

除了會議有特色之外，會議的收穫也是多方面的。我試著歸納出四個方面的收穫：一是出了一批階段性的成果，對異體字整理和《規範漢字表》的研製具有參考價值。二是瞭解和發現了一些在異體字研究上學有專長的人才，如有可能，我們將和這些同志商談具體合作的事宜，比如承擔《規範漢字表》研製的某些子項目等。當然，也不排除其他的合作方式。三是在一些基本概念的理解上取得了較為一致的意見。比如異體字的定義，多數同志贊成採用嚴式定義，即狹義的定義，當然對定義的具體表述還需要進一步探索完善。由定義問題，我想到我參加過的地名用字管理的討論，地名的界定也有個寬嚴尺度問題，大家的看法不盡相同。《中華人民共和國國家通用語言文字法》正在貫徹實施，但該法中涉及的“規範漢字”的定義到今天學術界仍然有不同的看法。並不是拿不出定義來，學術界對定義有好些種不同的表述，大家覺得都不是太滿意，但是沒有合適的學術定義並不影響大家對“規範漢字”的理解。昨晚課題組商量了一下，認為嚴式定義是會上比較一致的看法，我們也決定採用嚴式定義。當然概念的界定和實際操作中的認定還是會有距離的，就像有的先生說的，我們不能拿一個筐，什麼都往裏裝，講得非常好，目前的《異體字表》就在一定程度上存在這種情況。四是在異體字整理的原則、方法上大家的看法比較趨於一致。比如課題組負責人張書岩先生提出的一些具體的處理方式，大家沒有什麼異議，這些處理意見是經過課題組討論過的。課題組準備按照這些原則、方法，即按照嚴式定義歸納出幾條以後，先嘗試著做一下，看看有什麼問題沒有，若有問題再尋找其他的解決方案。總之，《規範漢字表》涉及的異體字整理的結果，在正式拿出去之前一定要反復聽取學術界包括在座的各位專家的意見。請大家一個字一個字地幫助認定一下，隨時提出問題隨時研究，避免簡單化。

以上是對會議的估價。我們認為這次會議開了一個學術科學化、民主化的好頭。

二、立足現實，把握形勢，進一步提高 對異體字整理工作的認識

我國上世紀50年代曾經大規模地進行過異體字整理工作，取得了明顯的成績。《異體字表》發佈後發揮了較好的作用，但也逐漸暴露出一些不容忽視的問題。《異體字表》中的不完善之處，在這兩天的討論中，大家都充分地談到了，這裏不再贅述。我想說的是，任何一個規範（標準），要想十全十美都是不可能的，何況《異體字表》是46年前發佈的規範。本來一個規範（標準）發佈後，5年之內需要復審、修訂一次的，《異體字表》將近50年基本沒變，“標