

如何使用電腦處理古今文字的銜接—以小篆為例

莊德明 中央研究院資訊科學研究所研究助理
derming@gate.sinica.edu.tw

許永成 中央研究院資訊科學研究所研究助理
yuenchen@gate.sinica.edu.tw

謝清俊 中央研究院資訊科學研究所兼任研究員
hsieh@sinica.edu.tw

摘要

中央研究院資訊所文獻處理實驗室自 1993 年起，即研擬漢字在電腦中的制式表達，以登錄漢字的字形結構，並建置漢字構形資料庫。經過多年來的持續努力，目前漢字構形資料庫收錄的字形已涵蓋《漢語大字典》的 54,678 個楷體字形及 12,208 組異體字、《說文解字詁林》的 11,100 個小篆字形及重文。

在漢字構形資料庫中，小篆與楷體等古今文字間的銜接問題，包括字形對映、字形分析及字義對映。小篆與楷體的字形對映，可參考《漢語大字典》單字條目下的字形源流演變。小篆字形的分析，是依《說文》的解形；而楷體字形結構的登錄，則參考《康熙字典》的部首，再將部件以橫連、直連、包含三個法則拆分。至於字義對映，可參考《漢語大字典》的〈異體字表〉。

漢字構形資料庫目前除了提供說文小篆字型及四萬多字的標楷體、細明體外字，還提供界面程式讓使用者利用部件來檢索字形、查詢字形結構及異體字。本文的重點僅在以小篆為例來說明如何使用電腦處理古今文字的銜接，其他的相關技術及說明可參考本實驗室先前所發表的論文及漢字構形資料庫使用手冊。

壹、前言

中央研究院資訊所文獻處理實驗室自 1993 年起，有鑑於中文電腦裡的漢字資訊嚴重不足，即開始研擬漢字在電腦中的制式表達，作為登錄漢字結構的依據，並建置漢字構形資料庫。漢字構形資料庫早期收錄的字形是以楷書的現代印刷字體為主，並且用來解決中文電腦的缺字問題。楷體字形結構的登錄，是參考《康熙字典》的部首，再將字形以橫連、直連、包含三個法則拆分，這和《說文》對小篆的構形分析是不同的。



《說文》中的小篆字形不但已形成一個嚴密而有規律的構形系統，而且小篆更是研究古今文字的過渡橋樑，無論對古文字的研究考釋，還是對今文字的整理，都起了很大的作用[1]。要在漢字構形資料庫增收小篆，不僅要納入《說文》的構形分析，更須先解決小篆和楷體的對映問題。相關問題的探討及解決，可在《漢語大字典》[3]的單字條目下找到重要的線索。圖一、圖二是摘錄自「楮」、「音」兩字在《漢語大字典》的單字條目，說明如下：

一、字形源流演變。例如字頭「楮」、「音」兩個楷體字分別為小篆「楮」、「音」的隸定字。

二、解形。例如「楮」字在《說文》的字形解說為『稅也。从木，音聲。』，而「音」字為『相與語唾而不受也。从、，从否，否亦聲。』



三、釋義及異體字。例如「楮」的字義是『棍，杖。』，後作「棒」。異體字的對映還可參考《漢語大字典》的〈異體字表〉。

字形源流演變中的小篆

字頭		說文·木部		居延簡甲一九九一
解形	《說文》：“楮，稅也。从木，音聲。”			
釋義	① 棍，杖。後作“棒”。《說文·木部》：“楮，稅也。”段玉裁注：“楮、棒，正俗字。”《廣雅·釋器》：“楮，杖也。”《淮南			

圖一、《漢語大字典》的「楮」字

字形源流演變中的小篆

字頭		說文·、部		睡虎地簡五〇九一
解形	《說文》：“音，相與語唾而不受也。从、，从否，否亦聲。𠂔，音或从豆，从欠。”			

圖二、《漢語大字典》的「音」字

《漢語大字典》單字條目下的字形源流演變及〈異體字表〉是解決小篆和楷體字形、字義對映問題的主要依據。字形源流演變是反映漢字形體的關係，異體

字則反映漢字在使用上的關聯。例如小篆「𦵏」隸定成「楮」，而它的字義則和「棒」相同。因此，在建置小篆構形資料庫時，可用「楮」替代「𦵏」；而與「棒」意對映的小篆，可透過〈異體字表〉的「楮」再找到「𦵏」。至於字形分析，小篆可依《說文》的解形，而楷體字則沿用先前的橫連、直連、包含三個法則來拆分。例如「商」的直接部件為「丨」、「商」，而「音」的直接部件為「立」、「口」。

在漢字構形資料庫中，小篆和楷體等古今文字可採用不同的構形分析，並建立彼此的字形、字義對映，如此不僅可透過現代漢字來認識古文字，更可藉由古文字而加深對現代漢字的理解，這也是利用電腦來處理古今文字銜接的主要目標。

自 1999 年起，我們陸續和北京師範大學及台灣師範大學合作，小篆的構形分析、和楷體字的對映如今即將完成。漢字構形資料庫目前收錄的字形已涵蓋《漢語大字典》的 54,678 個楷體字及 12,208 組異體字、《說文解字詁林》(以下簡稱《詁林》)的 11,100 個小篆及重文。以下各節將分別說明小篆和楷體的字形、字義對映，字形分析及古今文字銜接在電腦上的應用。

貳、小篆和楷體的字形對映

本節討論漢字構形資料庫中的小篆和楷體的字形對映。

《詁林》總共收錄 9831 個小篆及 1269 個重文，合計 11,100 個字形。這些字形絕大多數可透過《漢語大字典》的字形源流演變找到對映的楷體字頭，有些小篆對映的字頭可能不只一個，有些仍然沒有對映的字頭，有些則對映到同一個字頭。字形對映的情形說明如下：

一、部分《詁林》的字形，由於重複或是相似，以致對映到同一個楷體字頭。重複的字形有 31 個，包含 8 個小篆及 23 個重文；相似的字形則有 5 個。以上合計 36 個。表一列出重複的小篆在《詁林》的編號、出現的冊、頁，以及對映的楷體。重複的重文可見表二(見下頁)，相似的字形則見表三(見下頁)。

表一、《詁林》重複出現的 8 個小篆

組號	小篆	楷體	編號	備註	冊-頁	組號	小篆	楷體	編號	備註	冊-頁
1	𦵏	藍	282		2-512	5	𦵏	吹	828		2-1147
	𦵏		594	傳誤	2-849		5499		議刪	7-790	
2	𠄎	右	873	議刪	2-1205	6	𠄎	愷	3076		4-1313
	𠄎		1894		3-1000		6682		議刪	8-1123	
3	𠄎	吁	912		2-1248	7	𦵏	塗	7384	議刪	9-594
	𠄎		3053	議刪	4-1272		9010			10-1141	
4	𦵏	教	2502	議刪	4-559	8	𦵏	否	933	議刪	2-1271
	𦵏		3846		5-1013		7665			9-950	

表二、《話林》重複出現的 23 個重文

組號	小篆	楷體	編號	重文	冊-頁	組號	小篆	楷體	編號	重文	冊-頁
1	灑	涇	821	或體	2-1140	13	櫛	椶	3640		5-708
	灑		7073		9-265		櫛		9320	或體	11-102
2	愬	愬	837	或體	2-1158	14	叡	叡	103	籀文	2-263
	愬		6702		8-1140		叡		2517		4-587
3	紉	紉	954	或體	2-1290	15	𦉳	𦉳	866	籀文	2-1194
	紉		5553		7-842		𦉳		5524		7-813
4	孳	孳	1878	或體	3-979	16	變	變	8159	籀文	10-102
	孳		7805		9-1149		變		8223		10-174
5	𠂔	𠂔	2035	或體	3-1254	17	百	百	2221	古文	4-153
	𠂔		2756		4-855		百		5663		7-975
6	𦉳	𦉳	3379	或體	5-379	18	𦉳	𦉳	1826		3-904
	𦉳		7902		9-1245		𦉳		3072	古文	4-1305
7	𦉳	𦉳	3477		5-515	19	𦉳	𦉳	3190	古文	5-61
	𦉳		3730	或體	5-830		𦉳		6781		8-1229
8	𦉳	院	4545	或體	6-659	20	𦉳	𦉳	2346		4-345
	𦉳		9656		11-536		𦉳		5814	古文	7-1205
9	𦉳	𦉳	4808	或體	6-971	21	𦉳	沿	6991	古文	9-114
	𦉳		9538		2-708		沿		7245		9-443
10	𦉳	踞	1387		3-336	22	章	章	3310		5-238
	𦉳		5376	或體	7-609		章		9035	古文	10-1172
11	𦉳	難	474		2-708	23	𦉳	𦉳	2495	古文	4-545
	𦉳		6387	或體	8-720		𦉳		8471	古文	10-511
12	𦉳	螫	8814		10-884						
	𦉳		8925	或體	10-1004						

表三、5 個《話林》相似的字形

組號	小篆	楷體	編號	重文	冊-頁	組號	小篆	楷體	編號	重文	冊-頁
1	𠂔	𠂔	1908	古文	3-1030	4	𦉳	𦉳	3842		5-1003
	𠂔		8322		10-272		𦉳		9027	古文	10-1157
2	𦉳	普	4235		6-107	5	𦉳	布	6045	古文	8-283
	𦉳		6662	或體	8-1080		𦉳		9831	古文	11-884
3	𦉳	寒	6723		8-1164						
	𦉳		6825	或體	8-1271						

二、仍有 8 個《詁林》的古文，目前並無對映的楷體字頭。表四列出這些古文及小篆字頭。

表四、8 個仍無對映楷體字頭的《詁林》古文

組號	小篆	楷體	編號	重文	冊-頁	組號	小篆	楷體	編號	重文	冊-頁
1	𣎵	𣎵	3451		5-488	5	𣎵	𣎵	5567		7-861
	𣎵			古文							
2	𣎵	𣎵	3807		5-940	6	𣎵	𣎵	7536		9-805
	𣎵			古文							
3	日	日	4168		6-1	7	𣎵	𣎵	8399		10-417
	日			古文							
4	𣎵	𣎵	4259		6-133	8	𣎵	𣎵	8913		10-985
	𣎵			古文							

三、同一個小篆，可能同時存在著楷化字、隸定字或異寫字，以致對映的楷體字頭不只一個。目前至少有 1278 個小篆在《漢語大字典》找到一個以上的對映楷體字頭，合計 2658 個。例如《詁林》以「𣎵(音)」為直接部件的 19 個小篆中，有 5 個同時存在著楷化字及隸定字；以「𣎵(攸)」為直接部件的 15 個小篆中，有 5 個多了異寫字。表五列出這些小篆及對映楷體字頭出現在《漢語大字典》中的冊、頁、字。

表五、同一個小篆對映到不同的楷體字頭

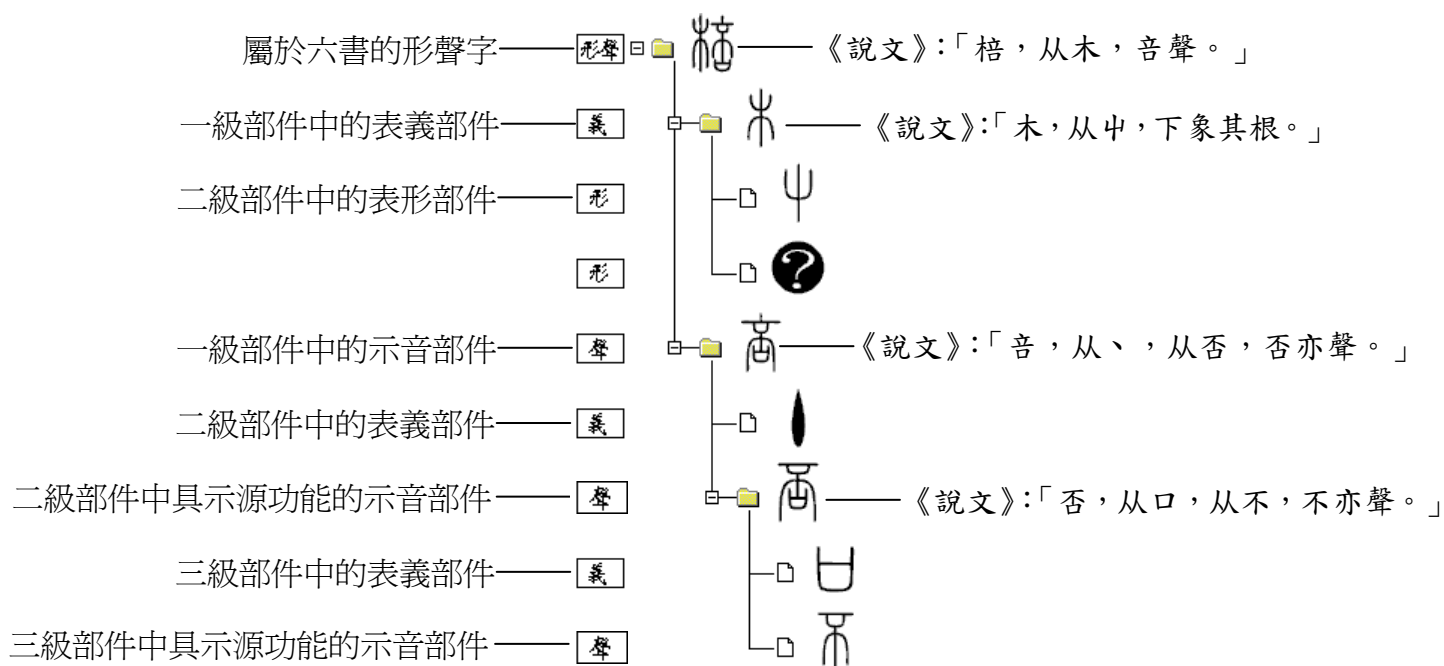
序號	小篆	楷體	冊-頁-字	序號	小篆	楷體	冊-頁-字
1	𣎵	剖	1-345-13	6	𣎵	倏	1-218-8
		刮	8-10-18			筵	6-3709-13
2	𣎵	倍	1-182-8	7	𣎵	脩	1-189-5
		倍	1-183-4			脩	4-2491-11
3	𣎵	涪	3-1659-1	8	𣎵	脩	1-176-4
		涪	3-1659-7			脩	1-189-2
4	𣎵	培	1-456-7	9	𣎵	條	1-202-15
		培	1-457-2			條	1-210-15
5	𣎵	醅	6-3587-13	10	𣎵	鑿	6-4209-5
		醅	6-3588-1			鑿	6-4209-6

綜合上述，《詁林》收錄的 11,100 個小篆及重文，扣除重複對映的 36 個，仍無對映的 8 個外，其餘的 11,056 個字形，目前對映的楷體字頭合計為 12,437 個。

參、小篆和楷體的構形分析

本節分別討論漢字構形資料庫中的小篆構形分析以及楷體字形在電腦中的表達。

首先討論小篆的構形分析。小篆字形結構的登錄是依據《說文》的解形，而部件功能的區分及構形模式的歸類，則參考《漢字漢語基礎》第二章[1]。圖三為小篆「梔(梧)」字的構形分析，以下分別說明。



圖三、小篆「梔(梧)」的構形分析

一、小篆部件登錄的順序是依據《說文》解形的順序。例如「梧」的直接部件先登錄「木」，再登錄「音」；「否」的直接部件先登錄「口」，再登錄「不」。

二、非字部件或《詁林》未收錄的成字部件，目前一律以「❓」代表。例如「朮(木)」的「下象其根」的非字部件；「壯」的成字部件「𠂔」，《詁林》未收錄。

三、《詁林》9831 個小篆使用的部件總數為 1,947 個，其中 1,831 個部件為小篆，其他的 116 個部件為重文。另外，《詁林》未收錄的成字部件共有 57 個，詳見表六。

表六、《詁林》未收錄的成字部件

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
二	𠂔	晶	𠂔	由	劉	𠂔	稊	泚	矩	妥	昏	筮	备	諤	佐	𠂔	𠂔	𠂔	𠂔
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
免	希	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57			
𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔

四、部件的功能區分。部件可依承擔的構意類別，區分成表形、表義、示音、標示及替代五種功能，說明及字例可見表七。

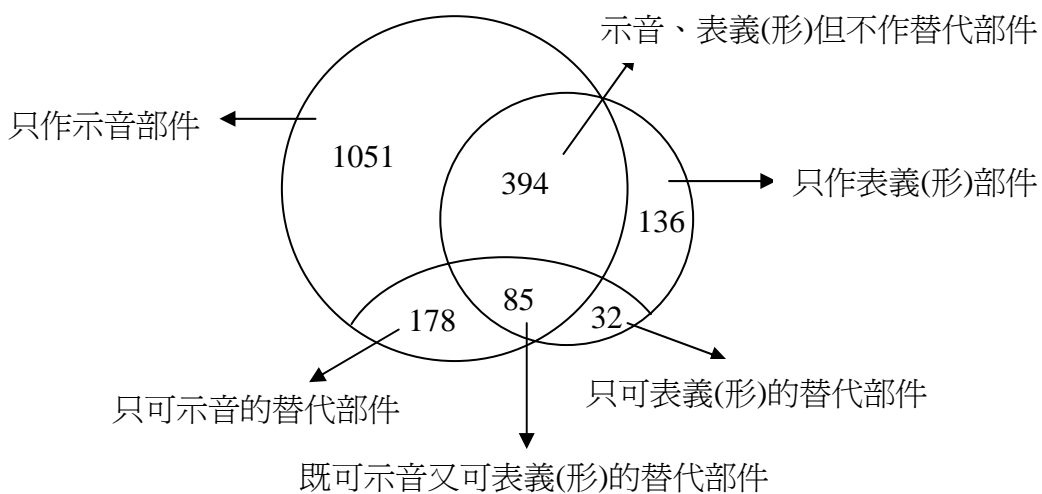
表七、部件的功能區分、說明及字例（摘錄自《漢字漢語基礎》）

編號	部件功能	說明	字例
1	表形功能	使用與物象相似的形體來體現構意的部件	「𠂇(巾)」為「𣎵(木)」的表形部件
2	表義功能	使用它在獨用時所記錄的詞義來體現構意的部件	「木」為「梧」的表義部件
3	示音功能	體現示音構意的部件	「音」為「梧」的示音部件
4	標示功能	不獨立存在，而是附加在另一個部件上，起區別和指示作用的部件。	「末」上面的一橫為標示部件
5	替代功能	部件本身不表示構意，而是作為另一部件的替代物。	「祝」的部件「兄」是替代部件「祝」。《說文》：「祝，从木，祝省聲。」

在示音部件中，有一部分還同時可以提示詞源意義，也就是具有示源功能。例如「丩」是「否」的示音部件，同時又可以提示「否」的意義與「丩」有關。但提示意義來源只是示音部件附帶的功能，不能單獨存在，所以不單分一類。

另外，漢字的大多數部件都有多種功能，這些功能在不同的結構環境中分別顯示出來。例如「刀」在「召」中有示音功能，而在「切」中具有表義功能。

《詁林》9831 個小篆使用的 1,947 個成字部件中，示音部件可依解形中的「从某某聲」、「省聲」、「亦聲」等來標記，計有 1,708 個。1,708 個示音部件中，有 215 個具有示源功能，即「亦聲」的部件。可被替代的部件可依「从某省」、「省聲」等來標記，計有 295 個。若把 215 個具有示源功能的示音部件也計算在內，其他尚未標記的表形、表義及標示部件則共有 647 個，由於標示部件數量很有限，這些絕大多數都是表義或表形部件。各類部件的重疊的情形可見圖四。



圖四、示音部件、表義(形)部件、替代部件的重疊情形

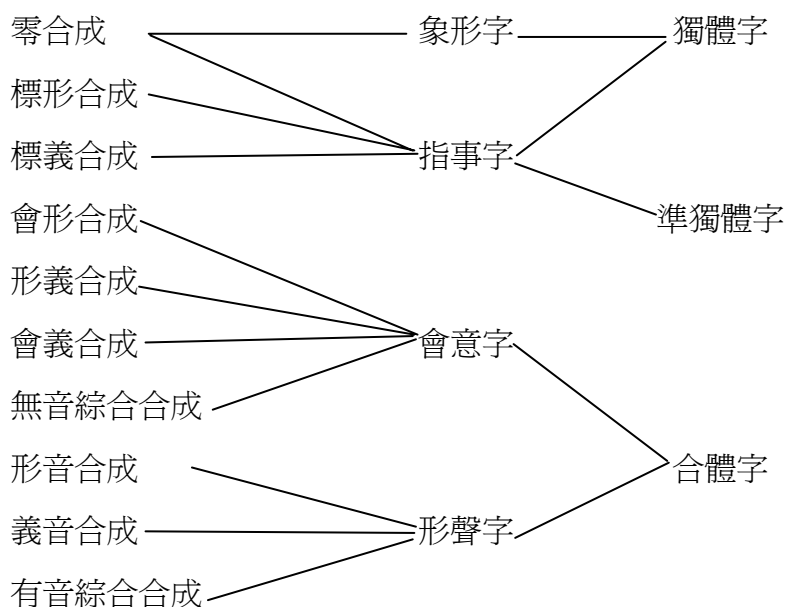
五、小篆的的構形模式與六書。六書是漢字發展到小篆時，許慎對漢字構形模式的概括。嚴格來說，只有象形、指事、會意、形聲前四書才是用來說解形義關係的條例。表八是摘錄自《說文解字·序》的六書說明。

表八、六書說明（摘錄自《說文解字·序》）

編號	構形模式	說明	字例
1	象形	畫成其物，隨體詰訕。	日、月
2	指事	視而可識，察而可見。	上、下
3	會意	比類合宜，以見指撝。	武、信
4	形聲	以事爲名，取譬相成。	江、河
5	轉注	建類一首，同意相受。	考、老
6	假借	本無其字，依聲托事。	令、長

漢字構形資料庫目前收錄了小篆，接著還要處理金文及甲骨文。由於漢字在發展時形體的遞變，《說文》對部分字形的解形已和金文或甲骨文不同，因而造成漢字在構形模式上的爭議。目前的做法是在漢字構形資料庫中，依不同的時空同時保留小篆及金文等古文字形及分析方法，再根據各自的解形來確認六書的對映。例如「示」的小篆「𠄎」，《說文》的解形爲「天垂象，見古凶，所以示人也。」是指事字，但是「示」的甲骨文字形表「地祇」，則爲象形字。

即使可依字形解說不同來確認六書的對映，然而六書之間的分際仍嫌模糊，缺乏一個較制式化的定義。《漢字漢語基礎》另外根據漢字的合成情況及部件在組構中的五種功能，從歷代漢字的實際狀況出發，而總結出表九(見下頁)的十種漢字構形模式。這十種漢字構形模式和六書的對映如圖五，透過這樣的對映，大多數的小篆可依示音部件的有無、部件組合的個數來確認出形聲字及會意字。



圖五、十種漢字構形模式、構形三分法及六書之間的對映

表九、十種漢字構形模式（摘錄自《漢字漢語基礎》）

編號	構形模式	說明	字例
1	零合成字	由一個單獨的成字部件構成，即獨體字。	「鳥(鳥)」的形、音、義信息都由本身提供
2	標形合成字	由一個表形成字部件加上標示部件合成	「眉(眉)」是由表形部件「目」及「畫出眉形」的標示部件合成
3	標義合成字	由一個表義部件加上標示部件合成	「音(音)」是由表義部件「言(言)」及標示部件「一」合成
4	會形合成字	由兩個以上的表形部件組合在一起	「弄(弄)」是由「王(玉)」、「冫(升)」兩個表形部件合成，以兩手把玩玉，表示玩弄。
5	形義合成字	用表義與表形部件組合在一起	「興(興)」，四手相對是表形部件，中間加表示「共同」的表義部件「同」，有「起來」的意思。
6	會義合成字	用兩個以上的表義部件組合在一起	「解」字由「角」、「牛」、「刀」三個表義部件合成，以用刀剖解牛角表示「解析」。
7	形音合成字	用表形部件與示音部件組合	甲骨文的「星」原為象形字，後又增加示音部件「生」，使字面所含的信息更為豐滿。
8	義音合成字	用表義部件與示音部件組合，即形聲字。	「論」字由表義部件「言」及示音部件「侖」組成
9	無音綜合合成字	表形、表義與標示部件的一次合成，但沒有示音部件介入。	「葬(葬)」由表示草的表形部件「艸」，表示「死」的表義部件和標示部件「一」組合而成。
10	有音綜合合成字	由多個表形、示音、表義、標示部件一次合成。	「春(春)」由表形部件「艸」，表義部件「日」及示音部件「屯」組合而成。

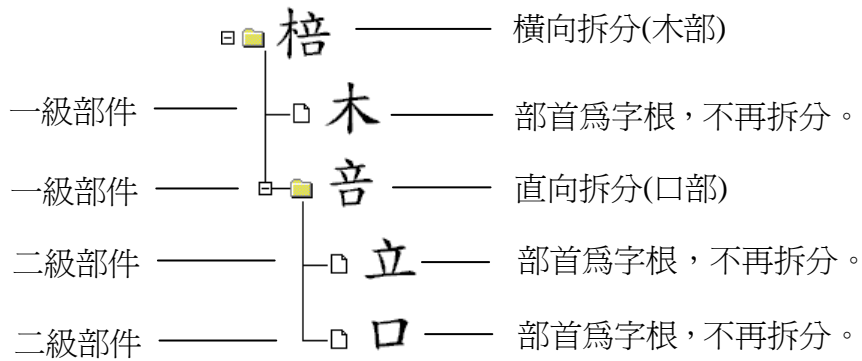
表十為圖四提到的獨體、準獨體及合體的構形三分法。

表十、構形三分法（摘錄自《漢字漢語基礎》）

編號	構形模式	說明
1	獨體字	由一個成字部件構成
2	準獨體字	由一個成字部件和一個非字部件構成
3	合體字	由兩個以上的成字部件構成

六書的標記目前雖未開始進行，但是由於示音部件已標記完畢，根據圖四的對映，包含示音部件的組合即可定為形聲字。在《詁林》中 9,823 個小篆中(扣除 8 個重複字形)，形聲字的個數為 8,433 個，佔 85.85%。至於其他的 1,390 個字形中，還有 77 個包含兩個以上的部件，64 個各包含一個表義(形)部件及替代部件，這 141 個都是會意字。最後剩下的 1,249 個字則留待人工標記。

其次討論楷體字形在電腦中的表達。在漢字構形資料庫中，楷體字形結構的登錄，是參考《康熙字典》的部首，再將字形以橫連、直連、包含三個法則拆分。圖六為「梧」的字形結構，以下分別說明。



圖六、「梧」的字形結構

一、直接部件的確認是參考《康熙字典》的部首，因此部首就是字根。例如「梧」為「木」部，直接部件為「木」、「音」；「音」為「口」部，直接部件為「立」、「口」，而部首「木」、「立」、「口」都是字根。

二、部件拆分的方式可分橫向(由左到右)、直向(由上到下)、包含(由外到內)三種拆分。部件若因重疊而無法以橫向、直向或包含來拆分的，也視為字根。例如「梧」橫向拆分成部件「木」、「音」；「音」直向拆分成部件「立」、「口」；「因」是由部件「口」包含「大」；「東」雖為「木」部，但是部件「木」、「日」重疊，因此也是字根。

三、部件登錄的順序是依據部件結合的順序。例如「梧」的直接部件登錄順序為「木」、「音」，「音」為「立」、「口」，「因」為「口」、「大」。

四、異寫部件的登錄必須反映形體。例如部件「衣」在「初」登錄成「衤」、部件「刀」在「刻」中登錄成「刂」，如此才能區分「初」、「剗」及「刻」、「劊」等異寫字。

五、若分別以連接符號△、△、△表示橫連、直連、包含，則「梧」、「音」、「因」的字形結構表達式可寫成『木△音』、『立△口』、『口△大』。這樣的表達式稱作構字式。

六、由於楷體字形結構登錄並不依照《說文》理據，以致非字部件大量增加。非字部件的取舍是以確保 99.9%的字形都可由構字式組成為原則(即留下 0.1%的殘留缺字用其他的方法解決，以增進系統的效益)[9]。例如「侵」的非字部件「彳」使用頻度高，應予保留；而「牖」字右邊的非字部件則可捨棄。

捨棄後的非字部件，必須改用次級部件來登錄，連帶捨棄構字符號，並在部件的前後加上起始符號⊖、終止符號⊕。例如「牖」字的構字式改寫成『⊖片戶』。

甫㇏」，這也恰好符合「牖」字在《說文》的解形：「从片、戶、甫。」

七、爲了方便和小篆部件作比較，若以《詁林》9,831 個小篆對映的楷體字形爲統計對象，這些字形使用的部件總數爲 2,137 個，其中 1,835 個爲《漢語大字典》的字，非字部件爲 302 個。2,137 個部件中的字根個數爲 556 個，其中 397 個爲字，非字字根爲 159 個。

八、構字式可用來識別字形，並當作電腦缺字的交換碼。例如《詁林》中以「𠂔（音）」爲直接部件的 18 個字形中，有 6 個字所對映的楷體字形是五大碼的缺字，這些字形即可透過表十一中的構字式來交換。

表十一、使用構字式來交換缺字

小篆	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔
楷體	𠂔	𠂔	𠂔	𠂔	𠂔	𠂔
構字式	走△音	月△音	竹△音	缶△音	髟△音	彡△音

九、構字式是以楷體字形結構爲依據，和小篆字形結構無關。例如表十二「祝」、「況」、「悅」、「祝」四個字的部件「兄」在《說文》的解形都不相同，但是在構字式中一律當作「兄」。

表十二、「祝」、「況」、「悅」、「祝」四個字的解形及構字式

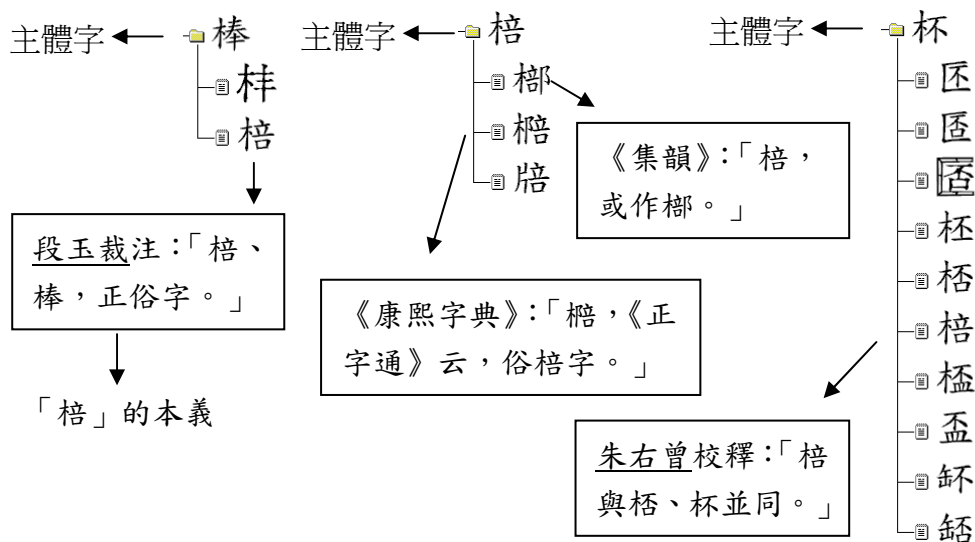
小篆	祝	況	悅	祝
楷體	祝	況	悅	祝
解形	从木，祝省聲	从水，兄聲	从心，況省聲	从示，从人、口
構字式	木△兄	氵△兄	忄△兄	礻△兄

肆、小篆和楷體的字義對映

本節討論漢字構形資料庫中的小篆和楷體的字義對映。

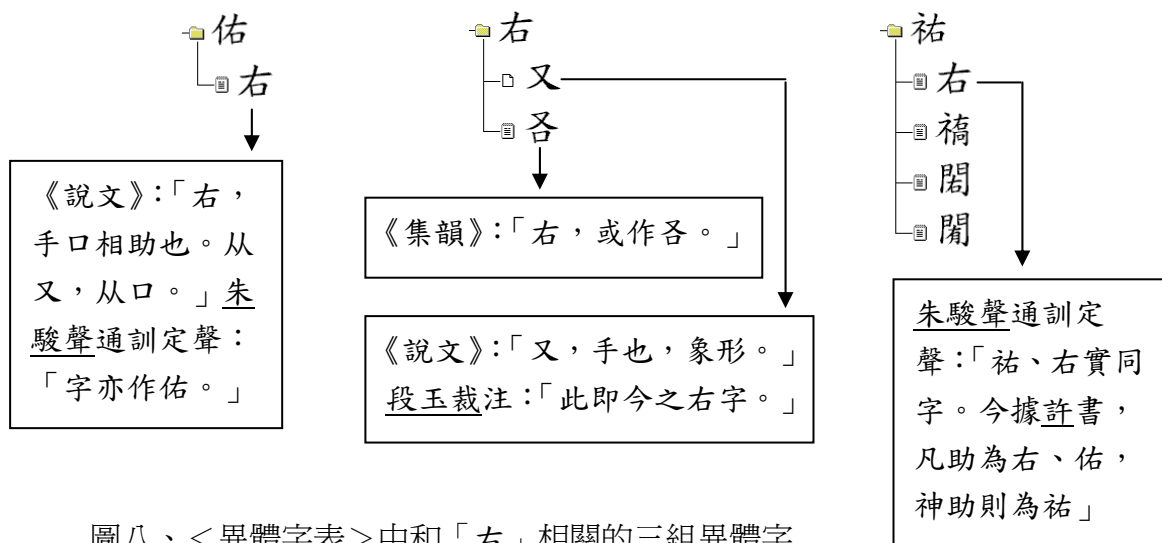
小篆和楷體的字義對映，可參考《漢語大字典》的〈異體字表〉。〈異體字表〉的編排是採用主體字統領異體字的方式，將同一主體字統領的簡化字、古今字、全同異體字（指音義全同而形體不同的字）、非全同異體字（指音義部分相同的異體字），集中在該主體字下編爲一組。〈異體字表〉目前收錄 12,208 組，涵蓋 36,309 個字形。

《詁林》收錄的 11,100 個字形中，有 9,278 個所對映的楷體字形出現在〈異體字表〉中。圖七（見下頁）列出〈異體字表〉和「楮（𣎵）」相關的三組異體字。其中「楮」不但是「槲」等字的主體字，還是「棒」和「杯」的異體字。注意的是，「𣎵」在《說文》的意義是「棒」，並不是「杯」，所以「棒」才是「楮」在《說文》的本義。



圖七、〈異體字表〉中和「梧」相關的三組異體字

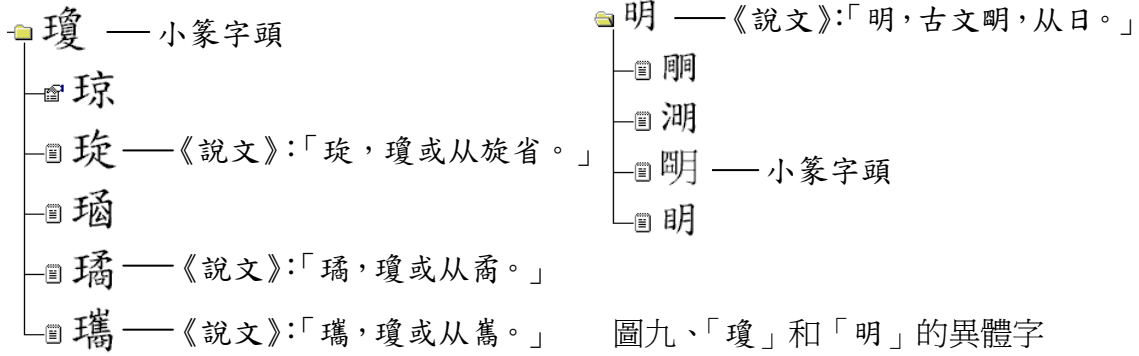
圖八列出和「右」相關的三組異體字。從字形來看，小篆「𠄎」隸定成「右」，楷化成「各」。但由字義來看，「𠄎」是「助」的意思，即楷體的「佑」字。然而「又(𠄎)」在《說文》的本義為「右手」，即楷體的「右」字。



圖八、〈異體字表〉中和「右」相關的三組異體字

《詁林》的 9,278 個出現在〈異體字表〉中的對映楷體字形，只作主體字的有 5,088 個，只作異體字的有 2,302 個，其他的 1,888 個則同時當作主體字及異體字。使用〈異體字表〉必須注意，主、異體字只是表達某一時空下兩個字使用上的關聯，至於是否為該異體字在《說文》的本義，〈異體字表〉並未標明。

由於〈異體字表〉為古今漢字的集中反映，所以《詁林》的 1,269 個重文所對映的楷體字形，也都收錄在〈異體字表〉中。重文並不一定都作異體字，有時反而是主體字。圖九(見下頁)列出「瓊(瓊)」及「明(明)」的異體字，其中異體字「琤(琤)」、「瑠(瑠)」、「瑠(瑠)」都是小篆「瓊(瓊)」在《詁林》的或體字；而主體字「明(明)」反而是小篆「明(明)」在《詁林》的古文。



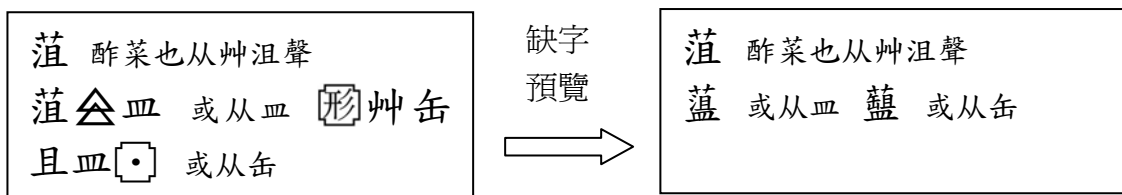
圖九、「瓊」和「明」的異體字

伍、古今文字銜接在電腦上的應用

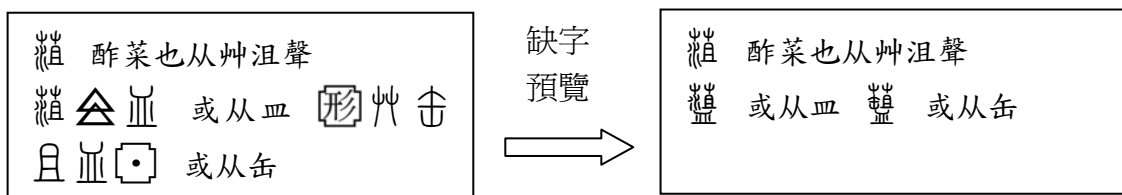
本節僅就古今文字銜接方面，舉例說明漢字構形資料庫在電腦上的應用。

漢字構形資料庫目前提供了五大碼缺字 41,658 個及小篆(含重文)11,100 個字形。缺字字型原先由葉健欣先生免費提供給學術界使用，去年我們開始重新造字，其中 35,838 字目前都已有標楷體和細明體可用。小篆字型的 11,100 個字形原為北京師範大學製作，我們透過小篆和楷體的字形對映，目前已重製成「北師大說文小篆」及「北師大說文重文」兩套字型，前者涵蓋《詁林》的 11,056 個字形，後者僅收錄重複對映的 44 個字形。我們同時在去年取得小篆字型的授權，現在免費提供給學術界使用。

圖十、圖十一說明如何在 Microsoft Office 處理缺字及小篆，兩個圖的差別僅在於使用的字型是楷體或是小篆，其中的缺字「菹」、「藟」分別用構字式『菹 艹皿』、『藟 艹且皿』來表達，再透過漢字構形資料庫的缺字增益集將構字式轉換成字形。圖十二(見下頁)說明如何透過漢字構形資料庫的界面程式，以部件「立」依序檢索出「音」、「楮」，並列出「楮」在〈異體字表〉中的三組異體字、「楮」的對映小篆「楮」及楷體字形結構。圖十三(見下頁)說明以部件「丩(不)」依序檢索出「否(否)」、「音(音)」、「楮(楮)」，並列出「楮」的對映楷體「楮」及《說文》的解形，而「楮(楮)」在《詁林》則沒有重文。



圖十、缺字預覽(楷體)



圖十一、缺字預覽(小篆)

以部件「立」依序檢索出「音」、「梧」

選取「梧」，即可列出異體字、字形結構及對映的小篆。

「梧」的楷體字形結構

「梧」在異體字表中的三組異體字

「梧」的對映小篆

圖十二、古今文字銜接(楷體)

以部件「丌(不)」依序檢索出「商(否)」、「商(音)」、「楠(梧)」

選取「楠」，即可列出異體字、字形結構及對映的楷體。

「楠(梧)」在《說文》的解形:「从木、音聲。」

「楠(梧)」在《詁林》並無重文

「楠」的對映楷體

圖十三、古今文字銜接(小篆)

陸、結語

漢字構形資料庫目前仍在持續建置中，只待小篆的六書標記完成後，就要開始處理金文了。利用電腦整理古今漢字，始終是一件繁重和冗長的工作，在此感謝中央研究院資訊所這十年來的支持，也希望這些成果能讓所有使用漢字的人來分享。由於文字學並非我們所長，本文若有疏失，也希望讀者不吝指正。

附錄、漢字構形資料庫光碟

漢字構形資料庫光碟內容如下：

- 涵蓋《漢語大字典》及 Big5 的楷體字形共 54,711 個。
- 收錄《漢語大字典》的異體字表，共 12,208 組，包含 36,309 個字形。
- 涵蓋《說文解字詁林》的小篆及重文字形共 11,000 個。
- 提供 4766 個部件以檢索字形，其中包含 1,324 個字根。
- 整理異體字根 296 組，共包含 716 個字根。
- 支援 Truetype 缺字及小篆字型。
- 擴充 Microsoft Office 的功能，可在 Office 下使用 54,711 個楷體字形及 11,100 個小篆及重文。
 - 開發漢字構形資料庫使用界面，讓使用者可以利用部件檢字及查詢異體字、字形結構及字形演變。
 - 提供處理網頁缺字的 Java Applet。

漢字構形資料庫光碟下載網址：<http://www.sinica.edu.tw/~cdp/>

參考文獻

編號	作者	書(目)/論文	出版社/研討會	出版年月
1	王寧等	漢字漢語基礎	北京科學出版社	1996年7月
2	李佳信	《說文》小篆字根研究	國立台灣師範大學國文研究所碩士論文	2000年7月
3	徐中舒等	遠東·漢語大字典	遠東圖書公司	1991年9月
4	莊德明	漢字印刷字形的整理	電子古籍中的文字問題研討會·台北	1999年6月
5	莊德明	中文電腦缺字解決方案	全國技專院校圖書館自動化規劃第七屆研討會·屏東	2001年12月
6	莊德明等	漢字構形資料庫使用手冊	中研院資訊所	2002年7月
7	楊家駱	說文解字詁林	鼎文書局	1994年3月
8	謝清俊	電子古籍中的缺字問題	第一屆中國文字學會學術討論會·天津	1996年8月
9	謝清俊等	中央研究院古籍全文資料庫解決缺字問題的方法	第二次兩岸古籍整理研究學術研討會·北京	1998年5月