

## Explore the Unknown

### Episode 2\_5 智慧型中文語言：數位化的漢字之美

人類思想上的活動，絕大多數都要倚靠文字來完成，它不僅是表達意念互相溝通的工具，也是傳承歷史文化的重要載體。中國人運用漢字的歷史已有數千年，隨著時代潮流的推動，在西方科技文明的衝擊下，始於圖像的漢字該如何與電腦計算機並存共容，繼續為中華民族發聲，乃當代科學家致力突破的課題。在這段磨合的陣痛期，中研院資訊所的科學家不但建構漢字在電腦世界立足的知識體系，也發掘出自身與文化間密不可分的情感，而從資源共享的角度來看，他們努力導致的成果，也為未來中華文化的發揚奠定良好的基礎。

1<sup>st</sup> edition 2006/06/03 編劇：歐陽百麟

段落	影部	聲部	長度
S1	文字的重要性		1' 00
	<p>△ 山洞中的壁畫，溶接甲骨文拓印，鐘鼎上的文字、碑文的拓印等等</p> <p>△ 各種古書古籍的資料畫面</p> <p>△ 前景溶接到若干古中國的出土文物，並在文物上有著各式各樣不同年代的漢字</p> <p>△ 以類戲劇的方式重現倉頡造字的情景</p> <p>△ 倉頡造的字經過動畫演變後成為現今文字</p> <p>△ 各式各樣的大自然圖樣與象形文字交疊呈現，最後畫面轉黑</p> <p>△ 上子集名稱： 智慧型中文語言：數位</p>	<p>SE：悠揚的中國國樂</p> <p>OS：許多人類思想上的活動，像是意念的表達、歷史的記載、文化的傳承，都要靠文字才能完成。</p> <p>從宏觀的角度來看文字，它不僅是表達思想的工具，更是彰顯與發揚一個種族文明重要的載體。</p> <p>SE：音樂轉場</p> <p>OS：從倉頡造字的一刻起，中國人使用漢字延續了數千年，在漫長的時間演變下，漢字蘊含了豐沛的生命力。</p> <p>漢字的起源始於圖像，每個字都是數千年前古人生活的速寫，就像一張張照片一樣，訴說著自己的故事…</p>	1' 00

	化的漢字之美		
S2	漢字的優美與漢字數位化的動機		1' 00
S2	<p>△ 動畫：象形的”日”→生活實景的日</p> <p>△ 動畫：象形的”自”(手指著自己的鼻子)</p> <p>△ 採訪謝清俊老師</p> <p>△ 請書法家在宣紙上寫祭祀的”祭”(確認文體)</p> <p>△ 書法的祭→動態畫圖</p> <p>△ 請書法家在宣紙上寫寒冷的”寒”(確認文體)</p> <p>△ 書法的寒→動態畫圖</p> <p>△ 採訪謝清俊老師</p> <p>△ INS 畫面建議：許許多多書有甲骨文字的古代文物或器皿</p> <p>△ 向謝老師情商，借調一些當時研究中文數位</p>	<p>《中研院資訊所特聘研究員 謝清俊》</p> <p>漢字從這整個角度來看，它就是意義的表達，但是這個意義的表達是用我們我所知道的一些生活常識，把這些圖像兜起來後，做一個知識的表達+不管是具象的或是抽象的，都可以做這樣的表達。</p> <p>SE：(書法家運筆寫字)</p> <p>OS：漢字所呈現出的，是漢民族文化的知識體系，而且這個知識體系能讓人一目了然。</p> <p>比方說祭祀的祭，就是一隻手把一塊豬肉放在供桌上；</p> <p>SE：(書法家運筆寫字)</p> <p>OS：寒冷的寒，就是在結冰的窯洞裡，有一個人躲在草堆中瑟縮著。</p> <p>漢字的線條，生動地勾勒出一面生活圖畫。</p> <p>《中研院資訊研究所 特聘研究員 謝清俊 博士》</p> <p>古時候的字現在我們雖然很少用，但是我們有時候要傳達一些情意的時候，我還是要去用到那個字，那你用到那個字的時候就活回來了，你像甲骨文，甲骨文是不是死字?不是，你寫一篇甲骨文的書法，甲骨文就活回來了，這跟歷史整個傳承有關係，我們如果不要忘記我們的過去，我們要了解歷史的源流，飲水思源的話，文字是必須的。</p> <p>OS：謝清俊教授，是國內發展漢字數位化的先驅。早在 1971 年，他跟隨著當時台大</p>	

	<p>化的照片，最好有馬志欽教授在其中</p> <p>△ 1970 年代，當時巨大的電腦資料畫面</p> <p>△ 許多研究人員在巨大的電腦中行走，一大堆擁塞的中文字</p> <p>△ 續前，一堆的漢字，像是零件般被拆解開來</p> <p>△ 交大字根系統當時發表的新聞報紙，或可拍攝交大字根系統</p> <p>△ 496 個字根像基因密碼般跑動</p> <p>△ 用電腦打出字根組成的文字</p> <p>△ 古書一個奇怪的缺字 當時電腦無法輸入</p> <p>△ 採訪謝清俊教授</p> <p>△ 二十五史的古書</p> <p>△ 特寫其中一個奇怪的古字，從電腦上找不到這個缺字</p> <p>△ 數個步驟呈現電腦造出這個二十五史上的缺字</p>	<p>電機系的馬志欽教授，研究如何讓電腦來處理中文字(加上目前使用的中文字數)。電腦是西方的產物，先天就欠缺處理中文的能力，而且那時一整間屋子大小的電腦主機，記憶體卻只有一個 MB，處理的速度很慢，如何在有限的電腦記憶體容量內處理大量中文，的確是棘手的問題。</p> <p>SE：音樂轉場</p> <p>OS：漢字包含字音、字義和字型三部分，其中漢字字型的組合結構，被科學家應用在電腦的構字系統。</p> <p>1973 年，交大字根系統首度發表問世，所謂的字根，就像我們使用的字典中的部首、偏旁等漢字的組合元件。</p> <p>OS：當時科學家將所有中文字拆解出 496 個字根，來組成四萬多個中文字。這看似符合一般使用者的需求，但一碰到古文研究或是建立數位資料時時，便會發現嚴重的缺字問題。</p> <p>《中研院資訊研究所 特聘研究員 謝清俊 博士》</p> <p>中研院做二十五史就發現缺字缺的很嚴重因為那時候電腦才一萬三千零五十一個字，古書裡面有很多字體、字型就不在這些裡面，我們盡量要維持古時候的字型+在電腦裡面怎麼辦呢？就只好缺一個字，造一個字+你怎麼樣告訴電腦這是一個造字？你必須要給它一個編碼，字體，這樣電腦才能在螢幕上找到</p>	
--	--	---	--

		字型，我們才能指到這缺字。	
S3	漢字數位化的過程及其碰到的困難		1' 00
S3	<p>△ 史語所的研究人員操作電腦應用漢字輸入的畫面，當他們再 KEY IN 一些古漢字時，面臨找不到字的窘境</p> <p>△ 當研究者以電腦傳輸古漢字時，接收的對象將會看不到古漢字的形體而產生不解其文的困擾</p> <p>△ 採訪莊德明老師</p> <p>△ 紀錄一小段莊老師在研究中心內建構漢字構型資料庫的過程</p> <p>△ 特寫該資料庫中大量的楷書字型</p> <p>△ 以動畫方式呈現漢字構型中上下、左右、包圍的組合方式</p> <p>△ 實際的文字部件，以不同的方式排列組合</p>	<p>OS：電腦普及是漢字數位化起初的動機。若數千年悠久歷史、數量龐大的漢字，無法與西方科技的電腦相容並存，所產生的不只是缺字這麼單純的問題；站在資源共享的角度，也不利文化的傳輸交流。因此科學家著手研究，將漢字與電腦共存相容的方法。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>漢字它是表意文字。表意文字它是因意構型，也就是說它的部件在這個字裡面它都有它的功能，我們只是依漢字本身的特性忠實的表達漢字的知識。+對我們而言，我們系統的核心是個部件，但對電腦而言，它是字，就是所謂的字碼。不管怎樣，我們還是要現在有的架構裡面來建立我們的系統。</p> <p>OS：1995 年，莊德明投入了漢字數位化的研究，並開始著手建構「漢字構形資料庫」。資料庫的建置是從楷書開始，當時應用的需求還僅止於楷書。</p> <p>由於部件是構成漢字的最基本單位，透過部件與部件間上下排列、左右排列、或是一個部件包含另一個部件等組合方式，可以在電腦上顯現絕大多數(是否有比</p>	

	<p>△ 莊老師在筆記型電腦上面示範漢字構型資料庫，拍攝青字部件所產生出來的各個與青相關的漢字</p> <p>△ 出土的殷商文物，特寫上面的文字</p> <p>△ 各種不同的碑體，或是草書、行書等不同的字體</p> <p>△ 採訪莊德明老師，INS畫面建議：拍攝莊老師與研究人員就著電腦討論互動的畫面，並特寫他們在處理小篆、金文、甲骨文等文字的情形</p> <p>△ 史語所內的研究人員面對不斷缺字的現象產生懊惱與不知該如何處理的情形</p> <p>△ 莊老師與史語所的研究人員互動的畫面</p>	<p>例、數據?)的漢字。</p> <p>OS：舉例來說，青色的青就是一個部件，如果我們在「漢字構形資料庫」內鍵入這個青字，透過部件組合的原理，便可發現清水的清、感情的情、晴朗的晴、眼睛的睛、精華的精等字，一次就可以認識一批與青這個部件相關的漢字。</p> <p>OS：但一套完整的資訊系統，光會處理楷體字是不夠的，從殷商的甲骨文算起，漢字的使用歷史已達三千四百年之久，其中因時代演變而產生的繁複變化、因書寫方式不同而導致異體字種類之多，也不斷地在考驗著電腦處理漢字的能力。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>當初我們在拆分字型的時候，我們有蠻多疑問的，後來我們才去看文字部件的書籍，就找到小篆，小篆裡面它有很多，包含這個字怎麼去拆解，+加完小篆以後，我們就可以慢慢加金文、甲骨，因為你會更想要找到他文字的源頭是什麼地方，這也是文字學上做研究它常常需要用到這些字，這些字它早晚都是要放進來的。</p> <p>OS：漢字知識庫建構的過程中，不斷有新的缺字產生，資訊專家發現當初設計這套漢字知識庫系統時，對於漢字組成和演變的知識不足。於是科學家轉向借重文字學家的專長。</p>	
S4	透過說文解字中提供字根部件組合的方式來解決缺字的過程		1' 00
S4	△ 應該要先接一小段，莊		

<p>老師走進歷史與研所和學者討論</p> <p>△ 採訪陳昭容老師，INS 畫面建議：陳老師和研究人員們拿著論文報告互相討論切磋的情景</p> <p>△ 有人拿起幾個剪紙的部件排列組合成一個字”雲/雞”(隨便舉例)</p> <p>△ 在電腦上也鍵出這個字”雲/雞”</p> <p>△ diss 真的雲飄走/雞跑走</p> <p>△ 莊老師與史語所的研究人員共同針對某類文字進行討論的畫面</p> <p>△ 秦始皇時代書同文的代表文字</p> <p>△ 許慎的說文解字及其內頁的拍攝</p> <p>△ 採訪陳昭容老師</p> <p>△ 許慎說文解字中，拍攝幾個拆解的例子</p> <p>△ 甲骨和金文的拆解</p>	<p>(採用一段中研院歷史語言所的現場討論)</p> <p>《中研院歷史語言研究所 研究員 陳昭容 博士》</p> <p>漢字有一個很大的特色，它其實就是最早的是類似圖像，把不同的部件集合在一起，那這本來就是漢字的內在的邏輯關係。+怎麼樣利用這組有限的部件去做各種的結合去展現無限的字型，在我們都了解和字的結構原理就會知道說其實它的部件並沒有多少，我們就知道可利用部件跟部件的組合，變成一個我們在螢幕上看的到的我們需要的字。</p> <p>(是否有原資訊專家的某文字拆解，後來徵詢文字學者發現其他部件的拆解法)</p> <p>OS：文字學家的投入和協助，彷彿為「漢字構形資料庫」注入一劑強心針，讓整套系統活了起來。</p> <p>文字乃成於眾人之手，漫長的使用年代間，幾乎每個朝代都會去做大規模的整理，像是秦始皇的書同文便是一例，這種整理將會把漢字歸納出一套組字成形的邏輯來，而其中最具成效的，就是東漢許慎所著的「說文解字」。</p> <p>《中研院歷史語言研究所 研究員 陳昭容 博士》</p> <p>許慎最早提出六書，他認為六書就是漢字的結構，所謂象形、指示、會意、形聲、轉注、假借，並不是說人們按照這六種方式去造一個字，而是說它對當時的文字的一種整理，他認為就是這六種方式，對甲骨、金文都可以這樣子拆解，這是沒有問題，而許慎做這樣的拆解，</p>	
---	--	--

	<p>△ 以某個字為例，呈現出不同異體字(員?或有其他例子)</p> <p>△ 莊老師和陳老師互動討論解決問題方法的畫面</p> <p>△ 特寫莊老師在建構資料庫時專注的神情</p> <p>△ 採訪莊德明老師，在筆記型電腦上面講解𣦵這個字的畫面</p> <p>△ 以電腦動畫呈現𣦵這個字的演變</p> <p>△ 採訪莊德明老師</p> <p>△ 史語所的研究人員使用漢字構型資料庫的</p>	<p>事實上就是因為他掌握了漢字的結構跟特色。</p> <p>OS：然而，漢字組合邏輯和電腦資料庫的設計架構還是有所不同。早期的漢字並沒有所謂的定型，常常同一個字卻有不同的形體，也就是所謂的異體字，部件的組合也不見得相同，但電腦的架構邏輯就是一對一，當要把不是標準文字的古漢字納入資料庫時，就必須回頭調整資料庫的設計方式。對原本沒有涉入過古漢字專業領域的莊德明來說，這趟漢字與電腦互相撞擊的衝突歷程，帶給他許多的刺激和反思。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>我第一次知道這個𣦵它的意思覺得印象非常深刻，這𣦵是現在加了一個石頭旁，這石頭旁是後來才加的，在金文甲骨文字型是蠻有意思，這旁邊四點，其實是螢火，一點一點發光的螢火，螢火通常是人死的，我們以前說鬼火，這𣦵在小篆本來就是鬼火的意思，然後後來這個大或許有人是說這是巫師他做法的時候加點𣦵灑一點𣦵在上去，在祈禱的時候他能產生更好的效果出來，那四點就是鬼火。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>古漢字的知識都是從跟他們（史語所）討論的過程中得來的，或者是有時候建置的過程可以自己去翻古漢字書，但是跟他們討論的過程的確是收穫蠻大，因為它們都把心得跟我們講，所以學電腦有時候也是蠻有福氣，因為它是一個系統整合，你可以接觸到不同領域，吸收</p>	
--	---	---	--

	<p>情景</p> <p>△ 資料庫中的構字系統解決了研究人員在列印古文字時的困擾</p> <p>訪陳昭容老師，INS 畫面建議：拍攝一段陳老師在研究所上課的情形，同學在台上專注聽講的畫面</p> <p>△ 同一個畫面裡出現各種同一字祖的字，構成一幅大張的圖畫 (如: 木、林、樟、楓、森…；或是雨、雲、雷雪…)</p>	<p>到不同知識。</p> <p>OS：透過漢字構字原理，讓電腦在組合漢字時有了依歸。這讓從古至今的漢字，皆能順利鍵入電腦中，這項創舉，不僅解決電腦缺字的問題，也讓研究古文的學者，有了更便利更容易掌控的工具。</p> <p>《中研院歷史語言研究所 研究員 陳昭容 博士》</p> <p>漢字有一些特色，如果某些時候，某些字，它們共有某些部件，往往意思上會有些相關連。+我們在做文字學的來源的時候，會發現說這些字是同一個字祖。+而且可以從甲骨的演化到各個不同的時代，形體發展我們都可以知道它的脈絡來，那麼在教學跟推廣漢字這個部分我覺得它有很大的幫助。</p> <p>(在此可接一段應用漢字知識庫的教學現場)</p>	
S5	舉實例來說明如何構字	1' 00	
S5	<p>△ 拍攝中研院資訊所的網頁，點入漢字構型資料庫中的畫面</p> <p>△ 莊德明老師示範講解漢字構型資料庫的應</p>	<p>OS：漢字構形資料庫的建置至今已快滿十三年，目前收錄楷書字形 60082 個，小篆 11100 個，金文 3459 個、楚系簡帛文字 2627 個，並包含了銜接古今漢字以表達字形的源流演變及古今字間的字義隸屬，記錄了不同歷史時期的漢字結構，並利用結構中的各級部件來檢索字形，同時制定構字式以解決楷書缺字的編碼問題，而研擬中的風格碼，則可用來解決古漢字的編碼問題等四大特色。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>像甲骨文著電腦裡面的“老”字，它的</p>	

	用	<p>象形味道就很濃，上面就是人的頭髮，留著長髮，下面就是人駝著背，杵著拐杖，到現在象形味道就比較沒那麼明顯了。這是老這個字。在系統裡面，甲骨文也是可以利用這部件來檢索。</p> <p>上面這老人留著長髮，駝著背，拿著拐杖。這“鬚”，其實就是鬍鬚的鬚了，這邊是鬍鬚的鬚，現在已經寫成這個鬚了。</p> <p>這“欠”是打呵欠。“眉”就很明顯了，畫個眼睛，上面是眉毛的眉。這“見”也是看，象形味道很重。這“聽”有個人，耳朵，這個口。“聞”也是一樣。這“臭”以前其實是嗅的味道；這“臭”其實底下是一個犬，“自”是鼻子，邊走還到處聞。其實這是在編件的時候過程覺得很有意思，我就把它挑出來。</p>	
S6	留下優美的漢字文化	1' 00	
S6	<p>△ 古文字藝術家在使用漢字來創造一些藝術作品的畫面</p> <p>△ 特寫古漢字的藝術圖畫</p> <p>△ 採訪謝清俊老師</p> <p>△ 古漢字藝術家上網去</p>	<p>OS：文字是有生命力的，對漢字來說尤其如此。透過漢字構形資料庫的建置，我們發現的不只是漢字本身的意義、樣貌和各種形體的變化，更可藉由這座時光隧道，方便快捷地穿梭於古今之間，並從這過程中去理解領略屬於中華文化中最美的一環。</p> <p>《中研院資訊研究所 特聘研究員 謝清俊 博士》</p> <p>如果中國字的形音意的構造不是那麼好的話，怎麼可能綿延？老早就被其他東西取代了，那麼中國歷史上又有那麼光輝的思想文化，種種的東西，我們不是講西元 1800 年以後，從有歷史到公元 1800 年都是世界上領先的國家，跟語言</p>	

	<p>下載漢字構型資料庫的流程，以及透過這個資料庫將優美的古漢字列印出來的畫面</p> <p>△ 採訪古漢字藝術家或者招牌工程業者</p> <p>△ 採訪陳昭容老師，INS 畫面建議：學子們使用漢字構型資料庫，或是老師帶領小朋友使用的畫面</p>	<p>文字很有關係，所以這是一個中國系統，世界上的瑰寶，我們應該好好去保存、利用。</p> <p>OS：原本只是一套基礎的電腦中文化應用工具，但在其背後卻蘊含了當初對此投入心力的科學家們，對於漢字文化發揚的期待。目前這套人人可以上網下載的漢字構形資料庫，也真的便利許多對漢字文化有興趣的人們。</p> <p><b>【這裡請加入一段真的使用漢字構形資料庫的用戶心聲，講講他們對這套系統的感受，不論是漢字藝術家或招牌工程的商家皆可】</b></p> <p>《中研院歷史語言研究所 研究員 陳昭容 博士》</p> <p>當這樣的學習變成是一個很有趣的，而提供這樣的材料的背景是一個正確的古文字學知識，它不是一個亂七八糟學習的東西，每一個東西我們都是有根據有來源的，它可以學習到正確的知識，它可以很有興趣的去學習，我想對於這樣的文化，他自然會喜愛的。所以我們看到展場上面的拓片、文字、他自然會想要去了解，這就是我們中華文化重要的一個部分。</p>	
S7	其他有關漢字數位化的研究	1' 00	
S7	<p>△ 陳克健老師在研究室內鑽研中文電腦斷詞研究的情景</p> <p>△ 採訪陳克健老師</p>	<p>OS：除了漢字構型資料庫的建立之外，在中研院資訊所內，還有許多專家學者投入教導電腦學習中文的研究，其中陳克健主持的中文斷詞研究，就是希望電腦從單字的學習，進展到詞的理解。</p> <p>《中研院資訊研究所 研究員 陳克健 博士》</p> <p>詞是了解文章的第一步，因為詞是文章裡面最小的意義單位，雖然字…中文字</p>	

<p>△ 在電腦上打入把手這兩個字，以及請你把手舉起來這個句子，以簡易動畫方式呈現這兩個字（例如閃光）</p> <p>△ 採訪陳克健老師，INS 畫面建議：陳老師在斷詞研究的網站上鍵入關鍵字模樣</p> <p>△ 陳克健老師與研究人員針對電腦上的斷詞進行討論與加強的工作</p> <p>△ 研究人員上網搜尋中文資料的畫面，可以拿在 GOOGLE 網頁上鍵入土地公三個字來做例子</p> <p>△ 採訪陳克健老師</p>	<p>也是有意義的，但是有些時候，比如說名字，拆開來其實沒有什麼意義，三個單位，就是所以詞是最基本的一個單位，電腦認出這是一個人名，把它當成一個單位來處理。基於這個原因，中文你必須先懂詞介在哪裡，每一個詞的意思知道以後，你才能去組合句子的意思，才能了解文章的意思，所以第一步是做斷詞的工作。</p> <p>OS：舉一個明顯的例子來說，「把手」在中文裡有可能是一個名詞，例如鍋子的把手，可是如果將這兩個字放在一個句子裡時，像是「請你把手舉起來」這個句子，這兩個字的意義是完全不同的。</p> <p>《中研院資訊研究所 研究員 陳克健 博士》</p> <p>我們處理的方式說，我必須有些訓練的原料，我有大量的文章，這些大量文章已經切好詞，先有一些基礎工具查詞典去切，切好之後，他一定有錯，人工修改以後，做大量原料，這大量原料可以從裡頭去做 machine learning 讓機器去學習，資料去學，最簡單的方法就是去計算，統計機率，碰到什麼狀況，用機率去計算出來。</p> <p>OS：中研院的電腦中文斷詞研究，就像在教導完全不懂中文的小孩學習一樣，只是電腦無法像孩童一樣有智慧、能舉一反三，和教導者互動。電腦只能被動地接收研究者從鍵盤輸入的知識，直到電腦達到某種中文程度後，才能夠從網路上去擷取閱讀大量的文章資訊，化被動為主動的學習。</p> <p>《中研院資訊研究所 研究員 陳克健 博士》</p>	
---	---	--

	<p>△ 陳老師與研究人員討論斷詞研究的畫面</p> <p>△ 研究人員利用斷詞研究，用鍵入關鍵字的方式與電腦互動溝通</p>	<p>語文這種東西，它是非常複雜，它是這麼多人這麼多年使用下來一個複雜的產物……但是對電腦來講，它是一個整體的，不管大小，常不常用它都要能夠處理，所以現在為什麼要用機器學習的方式，準備一些資料讓機器能夠學習，也許機器去學比較困難，所以要做資料的加工，作語料的加工，直接分出來詞的詞類標上去，有了這些資料以後電腦再去學可能可以學的更快。</p>	
S8	<b>漢字數位化未來的發展</b>		1' 00
S8	<p>△ 新聞資料影片：有愈來愈多的外國臉孔都在學習中文的畫面</p> <p>△ 前景溶接到莊德明老師在研究室內工作的情形</p> <p>△ 特寫漢字構形資料庫目前在電腦螢幕上呈現的狀況</p> <p>△ 採訪謝清俊老師</p> <p>△ 拍攝一般民眾在使用</p>	<p>OS：無論是漢字知識庫或中文斷詞研究，科學家的目的，都在幫助電腦增強中文的能力。隨著世界局勢與潮流的改變，未來勢必將有越來越多的人會使用到中文。站在這個歷史的關鍵點上，如何將漢字文化與電腦科技做更緊密的結合，將是未來數位中文發展的重要課題。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>大家知道漢字有字形、字音、字義，我們工作的長遠目標主要是文字學的數位化。文字學它有狹義化與廣義化，狹義指的就是字型這方面，廣義的話它包含字型、字音、字義，在狹義的文字學這方面主要是解決文字型的問題，廣義就慢慢加入字音、字義，逐漸的擴充漢字的字型，不斷的提升它的能力。</p> <p>《中研院資訊所特聘研究員 謝清俊》</p> <p>我們做漢字字型的分析，主要目的是解決中研院漢籍資料庫的缺字，但是這個只是諸多應用中間的一個而已……這些</p>	

	不同輸入法的畫面	知識經過構形資料庫整理出來以後，可以幫我們用來幫助我們做漢字的輸入，舉個比方說，我們把文字的知識越多輸入在電腦裡，電腦就愈可以幫助我們在輸入上的麻煩，我們對於漢字構形知道的越多，就越來可以以簡御繁。	
S9	結語		1' 00
S9	<p>△ 本及若干精采畫面剪輯，最後鏡頭落到某各上班族正在鍵盤上鍵入中文漢字的畫面</p> <p>△ 小朋友在學校內進行國語文課的畫面，並且圍在一起討論漢字構形資料庫的情形</p> <p>△ 最後以莊老師的採訪配合小朋友在使用網路的情形做結</p>	<p>OS：經過一趟漢字數位化的洗禮後，當我們再度打開電腦，一字一句鍵入漢字時，是否也能感受到自己與文字間緊密鏈結的情感？這種源自文化的漢字之美，將會透過科學家的努力，讓它與電腦相容共存，持續為我中華民族發聲，並廣為傳播到全世界。</p> <p>《中研院資訊研究所 研究助技師 莊德明》</p> <p>寫字寫久了已經把它當成符號了，事實上它裡面有非常豐富的內涵跟意義，透過我們的系統，透過我們的古漢字，或許，可以讓現在的學生對於這文字的情感可以增加，他們對整個文化的傳承應該有幫助吧！</p>	