

講者：謝清俊教授

各位先生、各位女士，今天很榮幸有機會到這裡向大家報告。主辦單位找我的時候，我已經從中央研究院退休了，今天報告的系統已經完全由莊德明先生繼續發展與維護。這個系統的發展有相當長的歷史，民國六十一年份的，1972 年交通大學的字根系統是本系統的前身。現在我們看到的系統是從 1993 年開始在中央研究院重新做的。這個系統能做出來要感謝很多人，第一個要感謝的是北京師範大學的王寧教授，以前有一段時間我們雙方合作很愉快；然後要感謝中央研究院歷史語言研究所的古文字小組，他們和我們合作做了小篆跟鐘鼎文，目前還包括在甲骨文。所以本系統包括歷史上各種文字字形。我們要了解歷史、文化時，對每一個不同的時代的字形，以及它的字音、字義都要做相當的了解。我們這個漢字構形資料庫依據漢字構形的理論，希望收集歷代的漢字。今天時間很少，等一下請莊德明先生跟各位作簡要的說明。最後要聲明：這個資料庫一直是放在網路上，讓大家免費下載的，因為我們覺得文化的東西，最好是讓大家來共享。如果台北市文化局有興趣的話，我們也可以贈送給台北市文化局轉贈給所有的中小學免費使用。謝謝。

講者：莊德明先生

先感謝謝清俊教授給我一個機會，來這邊報告我們在中研院資訊所十一年來建置漢字構形資料庫的心得跟一些經驗，也跟大家分享一些成果，那就像謝老師剛開始講的我們的漢字構形資料庫，目前已到 2.2 版，它的主要內容包含大概六萬個楷體字形，那裡面有一個比較重要的是異體字表，異體字表有一萬兩千多組；然後古文字方面是小篆跟金文。小篆加上重文的話是一萬一千一百個，金文目前大概將近三千五百個，甲骨文還有楚系文字目前還在進行中。

這是個國際會議，目前我們的系統還是只適用在微軟中文視窗的繁體字版，整個資料庫的下載網址在論文裡面就有。假設各位家裡沒有安裝寬頻網路，因而下載不方便的話，可以用 e-mail 跟我們聯繫，我們也可以寄送這個光碟或手冊給您。假設有些聽眾需要用到古文字而又不太懂電腦，平常心可打電話或寫信過來，我們也可以寄過去，這是我們系統的使用情況。今天我帶了二十套光碟跟手冊，有興趣的話大家可以向我索取。

目前的漢字構形資料庫的建置情形是這樣：我們最原先建置的是個楷書的構形資料庫，接著建置小篆構形資料庫，再來是金文，楚系文字跟甲骨文。每個資料庫事實上都是這樣，就是說在建置的時候你要找到一些你要建置的字形，比如楷書，可從《漢語大字典》來找，小篆的話，從《說文解字》，金文和楚系文字也是一樣。找到適當的字形以後，接下來就繼續做分析，把這個字形分析成部件，部件再分析成字根（或叫做基礎部件）。建置的每一個字集都有它自己的異體字表，每個字形在電腦裡都有對映的電腦字型來呈現，最重要的是這五個構形資料庫其實不是分開的，它們的字形都是銜接在一塊的。那我們再往下看，這是楷書，主要的字書為《漢語大字典》；這是小篆，主要的字書為《說文解字詁林》，楚系文字為《楚系簡帛文字編》，金文為《金文編》，甲骨文為《殷墟甲骨刻辭類纂》。這些是字書裡面的字頭數，還有字形數。

首先要講的就是小篆的構形資料庫目前都已建置完成，所有字形的結構都已經登錄了，這在 2003 年 3 月我們就做完了。在金文方面，《金文編》收錄的這些字形有一個特色，它的異寫字很多，金文跟楚系文字的字形將近都有兩萬多個，異構字的話可能只有三、四千個，其他都是異寫字形。除了異寫字形外，金文的另外一個比較特別的是將近一千字的圖形文字，這兩個資料庫我們大概今年底都可以建置完畢。關於金文的異寫字，早期我在跟史語所金文工作室討論字形的時候，當初總覺得只須收一個異構字就好，但是這兩年來持續的接觸古文字，最後也被他們同化，覺得這些異寫字每個字都應該放到電腦裡才對。金文的圖形文字非常精彩，非常豐富，我們今年也要放到電腦裡。

剛剛講的那五個構形資料庫，事實上他們之間是有關係的，他們字形是互相銜接的。比如說我介紹這邊一個拓片，拓片上面有一個字是這個樣子的，這個字跟《金文編》去做對照，發現可對映到《金文編》裡面的兩個字頭。這兩個字頭，都是小篆的字頭，那底下這兩個金文，一個是「乍」，一個是「作」。講到字形的銜接，就是說這些金文、小篆、楷體字等字形在電腦裡是怎麼銜接，怎麼呈現？目前我們做法是這樣，至少從我們使用電腦的概念是這樣，大家可能比較容易瞭解。比如說在電腦裡面使用的內碼，不管是 big5 或 unicode，目前這個內碼我們看到楷體字形是這個，在字形的銜接上，我們的機制是這樣的：不同的古今字形使用的電腦內碼都是相同的。比如說這個「乍」切換到不同的電腦字型，它的形體就不一樣、比如說換到北師大說文小篆，或換到我們自己製作的中研院金文，它的字形就是不一樣。然而另外一個「作」，其實它的內碼是不一樣的，字形是不同的，金文裡面沒有「作」這個形。這是一個字形銜接的機制，就是說在使用上，像在 word 的使用上你們只要換字型就好，這些字型是在電腦裡用的內碼都是一樣的，在資料庫裡是沒有什麼區分的，都是一樣的。

另外一個就是字義上的銜接。在字義的銜接上，剛剛看到的兩組金文，這個「乍」在《金文編》裡面有兩個字頭，一個是當「作」來看，一個是「乍」本身。這是我們資料庫裡面一個界面所呈現的畫面，也就是說在《金文編》裡面這個「乍」分屬兩組異體字，我們把這個架構做出來了。但是另外一個問題是在對映楷體字

的異體字表上。比如說《漢語大字典》異體字表收錄的一萬多組異體字我們都已經收到漢字構形資料庫，裡面有一組是這樣，字頭是「作」，底下也有一個異體字是「乍」，所以這個時候我們還要對這個收錄在《漢語大字典》的異體字「乍」加一個金文的標示「金」，表示這是對映的金文。假使在字形演變中找不對映的金文，你可能要到異體字來找。這是我們針對《漢語大字典》異體字表所做的第一個增補——標示的增補，第二個增補則是異體字的增補。看到第二組這個「乍」在《金文編》裡還多收了四個形，它的異構字。這四個異構字並未收錄在《漢語大字典》的異體字表裡，所以我們必需增補，這些增補的異體字同時也加上金文標示。古今文字的字義銜接，我們目前的做法是這樣子。

除了金文標示外，其實我們還有小篆等古漢字標示。比如說這個「燃」字有異體字「然」，而「然」有個「篆」的古漢字標記，這個意思是這樣，「然」字在《說文解字》裡的意思是燃燒，就是後來的「燃」，於是我們就補了這個標記。當我們用滑鼠在這個標記的地方按一下，你發現在字形演變的畫面裡有一個小篆的字形，還有一個金文的字形，這個銜接是這樣的，一個小篆字形、金文字形在這裡，一個楚系文字在這邊，他的意義代表說，我按一下金文字形「然」，這個「然」在《金文編》裡面是擺在小篆這個字頭底下。那在金文裡面其實他還有一個異構字在這個地方；然後楚系文字也是一樣，也有一個異構字在這個地方，這是我們的一個字義的銜接機制。除了這些標示外，根據《說文》還有說文或體、說文古文、籀文，還有金文、甲骨文、楚系文字，這些是目前加的標記。另外可從異體字表得知古漢字的標示字數。比如加上說文小篆標示的異體字數有 2,585 個。這個表的異體字其實跟剛剛講的是類似的，例如這個主體字「燃」沒有直接對映的古文字，但是可以從我們異體字表找到這個「然」，這是古漢字的標示。那剛剛講的是一個古今文字的銜接機制，這個銜接機制包含字形的銜接跟字義的銜接。

針對不同歷史時期的字形，我們還做了構形分析。關於構形分析，我舉我這邊一個例子。我本身並不是學文字學，引用了一些參考資料，講錯的話請大家不吝指正。比如說這樣一個「磷」字，它的意思是熒火或鬼火，這個可以說我們為什麼要做這個古文字，這個熒火或鬼火它這邊有四點，好像熒火跟鬼火閃爍的意思，但是這樣一個字形到小篆以後，看起來就像兩個火，雖然還是鬼火的意思，但是分析的時候就變成「炎」，本來是一點一點閃爍，現在火突然變大了，就變成「炎」了。當然另外一個字形跟它相關的，這個「磷」後來寫成這個樣子，上半部變成「米」，連「火」的意思都不見了，最後我們又看到現在這個「磷」，又加一個石頭旁的，這幾個不同的字形在我們系統其實是作不同的分析。比如說，這邊有看到的金文部件為一個「大」、一個「舛」，小篆部件為上面一個「炎」、一個「舛」，楷體部件為一個「米」、一個「舛」，這些分析都不一樣。那所做的這些分析，有一些是跟構形理據有關係，有一些可能沒有什麼直接關係，但是為什麼這樣做呢？我們可以從一個觀點來看，就是說對於各個不同時期的漢字所做的不同構形分析，其中有一個功用就是你至少可以根據分析後的部件來檢索這些

字形。比如說在《說文解字》裡面這個「磷」編在「炎」部，你可以從「火」來找這個字，找到「炎」後再找「磷」，這個是我們的架構。你在這邊可以從「火」來找，根據這個小篆的架構，你可以看到字形的銜接，那楷體字呢？楷體字在利用部件找字的時候，你不可能從「炎」來找，而可以從「米」或「舛」來找這個字，《康熙字典》也編在「米」部。我們還看到異體字表裡面還有這樣一組，這個是小篆的楷化字。不同的構形分析你有各種不同的找字方式，因為你看到的是那個形，就用那個形的部件來找，這是很自然的。比如我看到這個「磷」，就從「米」來找，你這樣分析，我們就可以這樣找，這在檢索上有一定的功用。

以下是一個金文的例子，比如說我們從金文的「貝」找到這個「得」，這個從「又」、從「貝」的金文「得」，然後這邊列出金文「得」的異體字，金文的「得」有好幾個，在我們處理金文的時候，其中有一個很大的收穫，是收錄到很多異構字，從異構字裡可看到這些漢字的不同生成意義。在這邊是一個字形的銜接，你可以看到貝殼的「貝」到小篆就寫成見面的「見」。這個是目前還在做的一個甲骨文的例子，看看甲骨文這個「孕」字，所以不常看古文字的看這個甲骨文的「孕」字就覺得很傳神。這個是楚系文字，使用的界面都是一樣的。查到這些字以後，其中一個對我們目前研究很有幫助的是很多字書的索引都可同時查得到，包含每天我都在查的《漢語大字典》，這對我來講其實很重要。我舉一個例子，有天一個金文工作室的同事來跟我訴苦，他說花了好多時間才在《殷周金文集成引得》中找到「乘」字，因為「乘」字在這本書中是依據《說文》的部首編在「桀」部。在字書裡找古文字有時候是一個挑戰，因為它可能按照《說文》的部首編排，你最好要知道這個字在《說文》的結構，才能找到這個字。這當然和從《漢語大字典》來找字是不一樣。也就是說，這個索引對研究古文字而須天天翻書的人實在幫助很大。這個是目前建置的索引，剛剛是對於漢字構形資料庫這個字形、字義的銜接，還有一些字形的屬性所做的一些說明。

現在講到如何利用漢字構形資料庫來解決缺字問題。解決缺字問題是漢字構形資料庫的一個應用，不是說漢字構形資料庫只能用來解決缺字問題。缺字是什麼，缺字就是漢字交換碼裡字形不夠所引起的，比如說我目前電腦裡面只能放兩萬字，但是《漢語大字典》五萬多字怎麼放到電腦裡，放不進去的三萬多字就都是缺字。缺字問題的根本原因在哪裡？漢字是一個開放性質的字集，就字數而言，已不適合作固定數量的限定，這跟數量已定的西方字母集是不可一概而論的，但是現行漢字交換碼的結構，卻仿照西方語言字母集的結構來設計，這不能不說是削足適履。我做一個簡單的比喻好了，比如說我們現在要把《漢語大字典》這五萬多字擺到電腦裡，那應該怎麼擺？稍稍了解一下目前電腦編碼結構的人可能就會這樣說，就一個蘿蔔一個坑啊！現在冬天蘿蔔又好吃又便宜。假使你這樣說的話，我們從另外一個觀點來看，假設現在要把英文字典的單字擺到電腦裡，單字可能有十萬個，那我能不能也這樣做，一個蘿蔔一個坑？十萬個單字，電腦裡面有個位置我就塞一個單字，大家聽到這樣做的話，一定會覺得這樣做實在不可思議，貽笑大方，怎麼會把英文單字一個一個擺到一個字碼裡來，但問題是我

們目前整個漢字的編碼就是這個樣子，一個蘿蔔一個坑，這就是目前缺字問題的根源。這邊舉的是一個缺字的例子，我們在做金文的時候，比如說《金文編》裡面的「寶」字有二百七十三個，其中大概有二十個是異構字，它們的構形不一樣，這二十個異構字把他楷化，然後這邊藍色的，表示是 Big5 的缺字，黑色表示目前有這個字。

剛剛講的是缺字的問題，現在講解決缺字問題的方法。解決缺字問題的根本在於解決現行漢字交換碼的根本缺失，我們解決缺字問題的方法，是遵從漢字構形的原理，對漢字字形的結構作分析，一個字的字形結構是一個字的構字式，是該字極佳的識別符號跟工具，因為字形不一樣，結構就不同，反之結構相同，其形也相同。這是什麼意思？就好像我要介紹一個人給你認識，那我可能要講這個人的身高、體重或相貌的特徵，但是我不會跟你講這個人的身分證字號，因為你不能從身分證字號去認識這個人。我們現在要把字形介紹給電腦，把字形放在電腦裡，我們的做法也是這樣，我們不能只給電腦一個字碼，還是要把這個字形的特徵告訴電腦，這樣電腦才能把字形處理好。那底下是我們一個方式，就是說告訴電腦這個字形到底長的什麼樣子，而不要讓電腦只有存一個數字，這永遠沒辦法解決缺字的問題。

不同的漢字有不同的結構，它的特徵在什麼地方？就是它的部件不一樣嘛，每個字形有不同的部件，但是有極少數字形的部件位置擺得不同，也可能就構成不同的字形。比如說落日餘暉的「暉」，「日」在左邊，「軍」在右邊；還有這個暈車的「暈」，上面是「日」，底下也是「軍」，這兩個字的部件相同，相對位置不同。但是這些字數並不是很多，通常它的部件都是兩個，相對位置可能是左右、上下跟內外，所以這邊是我們一個描述字形的機制。比如說這個「順」，它就是長得這個樣子，左邊這個是「川」，中間是一個左右連接的符號，右邊是個「頁」；這個「含」呢，它上面是一個「今」，底下是一個「口」；這個「圍」呢，外面是個圓圈的「口」，裡頭是這個「韋」。假使這個漢字的部件是三個以上，我們的描述方式就是把這個部件按照書寫的順序寫上去就好了，不需要再做更複雜的描述。這些字的部件排列可能是上面一個，底下兩個；或者是左邊一個，右邊兩個；或者是左邊一個，右邊三個。其實電腦雖然笨，但是還沒有那麼笨，你跟他講這三個部件或四個部件，大概已足可區分了，制定那麼多符號其實有時候只是把人弄的更糊塗。這是我們目前用的一個機制，這是告訴電腦，這字形長的是什麼樣子，他的特徵是什麼樣子，一個制式的表達方式，剛剛那些符號事實上就夠了。另外我們還制定了八個符號，讓大家表達起來更方便，因為中文字有很多是部件重疊上去的，像這四個火，這個你就這樣寫就可以，你就不用左右、上下描述，那個比較麻煩，而這樣表達比較直接。

以上是缺字在電腦裡的表達方式，那後來我們應使用者的需求做一些簡單的運用。其實這個系統可以不斷的擴大，不只是這個樣子，這是目前做的一些簡單的運用，一個查詢的界面，就好像我們剛剛看到的，你可以從部件來找字形。然後在 Microsoft office 裡我們也提供了一個缺字處理的增益集，用來處理 office 文

件上的缺字，因為這是大家寫論文最常碰到的。還有在網頁的應用方面，我們也加了一些機制來處理網頁缺字，另外我們還提供了幾套缺字字型跟字形產生器。這些都是簡單的應用，在這邊報告這些應用，主要是用來說明整個系統運作的機制。比如說你在安裝了我們的系統後，打開 word 就能看到你多了很多字型出來，包含古文字、標楷體缺字或細明體缺字，這是你形式上看到這些字型；使用 office 的時候你會看到多了一些介面出來，比如說我裡面有個構字式，你怎麼把它換成字形，把它換成圖片，換成圖片可以傳給對方處理比較方便。這只是一個運用的機制，真正使用的時候你就是這樣，告訴電腦這個是 big5 缺字，告訴電腦我要的這個字形長的是這個樣子，那透過我們剛剛的機制做一個轉換，就把這個字形轉出來，或者是圖片轉出來，轉成圖片傳給別人的電腦，他沒有我們的字形，也可以看的到，這是個應用的方式。

古文字應用是這樣，古今字形既然都銜接起來，你在使用古文字，基本上就是換成不同的字型嘛，對不對？比如說你要換到小篆，最好像我們從標楷體換到細明體，這邊換到北師大說文小篆，事實上是一樣的。那另外就是說，缺字處理也是一樣，比如說這個字透過我們剛剛的機制，交換碼裡面有那個字形，我們可以把那個古文字編到那個位置，交換碼裡面沒有那個字形怎麼辦呢？還是用我們的老方式，用這個構字式，構字式打出來，換成北師大說文小篆，這個字形換上去轉出來，就是小篆字形了，這個也一樣轉成圖片。金文的例子也是一樣，我們自己研發一個中研院金文字型，金文這個「射」比較傳神，有「弓」有「箭」；金文裡還有一個字，左邊是「虫」，右邊是「正」，這個字也可以楷化，Big5 也沒有這個字，那我們處理方式還是一樣，用構字式換成中研院的金文字型，把它轉換以後得到這個金文字形，所以使用手冊或是我寫的論文，裡頭所有的金文字形，所有的楷體字形，所有的小篆字形，全部都這樣轉出來的，當然，甲骨文和楚系文字也是這個樣子。

接著說明網頁缺字的處理機制，這個機制的重點在於使用者不必安裝我們的系統，他一樣可以看到由構字式轉換出來的的字形。所以在這個架構圖裡可以看到一個伺服器，當你的網頁送出來時可能還要加幾行指令，再送到使用者那邊，利用這些指令可透過這個伺服器把那個構字式換成字形，讓使用者可以看得到。那使用者不用安裝任何系統，就可以看到缺字。對於系統開發的人，他只要在碰到缺字的時候，採用構字式來表達，他不需要再去處理缺字這方面的事情。這個是使用的例子，比如說他可能說網頁這邊看到一個構字式，而這個構字式很快一閃就變成這個字形，這是一個處理網頁缺字的機制。

整個字形結構的描述也可以拿來製作字形產生器，字形產生器的基本功已經有，但是還沒有和系統整合，希望下一個版本就能整合。意思是說你有不同的字形，有時候可以造字，有時候來不及造的話，也可以透過字形結構自動產生這樣一個字形。畫面上的這幾個中文缺字其實都是異寫字，這邊的「皿」換成「血」，這邊「日」、「月」的「明」換成另外一個「明」的異體字，就可以換出不同的字形出來。在目前的電腦裡除了「盟」以外，其他就都是缺字了。

這邊我要再說明的是，從解決中文缺字的這個觀點來看，事實上電腦裡要擺的最基本東西是這樣，第一個就是常用字，接下來比較重要不是什麼次常用或罕用字，而是要把整個漢字的基礎部件擺上去，或者是說你把更大的合成部件都擺上去，再加上一些描述結構的構字符號。這樣一個系統，常用字就用這邊處理，其他的缺字就用這個結構式來描述，這樣子應該是一個中文電腦的最基本的字集。或許你的字形可以多加一些，比如說 Big5 的 13051 個字。

至於漢字構形資料庫的未來展望，我們希望今年應該可以把《金文編》跟《楚系簡帛文字編》裡面的所有字形都收錄，這些字形包含楚系文字的異寫字，也就是說《金文編》跟《楚系簡帛文字編》裡面各兩萬多字都應該要擺到電腦裡面來，而甲骨文的構形資料庫也在持續建置中。另外一個更重要的是，我們目前的資料庫偏重在構形，我們希望有機會能跟其他相關單位合作，例如把整個聲韻的系統，各種字書加到資料庫中。字書可包括最簡單的小學生的字典，到最複雜的《說文解字詁林》、《金文詁林》，都可以擺到我們的系統裡來，建構一個真正的漢字的知識庫，而不只是一個漢字的構形資料庫，那我今天報告就到這邊，謝謝大家。

問：

我想請教謝教授跟莊教授，我們都曉得，在中國的四周，有很多新字跟中國的當地不大一樣。如，台灣遺留日本的叮字很多。我現在要請教的問題就：我們招財進寶四個字喔，變成一個字的話，這在電腦如何去解決？謝謝！

答：

像這個招財進寶這四個字我們現在系統裡面自己已經可以把牠造出來，就是用剛剛莊先生介紹的這個構字的形式，只要你把構字的形式放入電腦，電腦就可以把招財進寶四個字一起合起來，這些對資訊技術來講都不是很困難的技術。

問：

剛才莊教授提到有關英文的這個拼音字母上面，每一個字是不用字碼在上面，但是對於漢字每一個字都要編上一個字碼，那我想到所以會產生很多缺字是因為我們一定要把每一個字形都放上字碼，我的電腦上面我缺的字我就自己編了一個字碼，我就用一個造字系統去把他編了一個字，然後放進去，然後等到我閱讀別人的文件的時候，別人的文件裡面，他也造了很多新字，就是造了他的電腦裡面原本沒有不存在的字碼，可是跑到我的電腦裡面並不叫缺字，謝教授您的系統裡面解決缺字，可是別人的文件到了我的電腦上面的時候，就不叫缺字了，就變成另外一個字形的字碼了，那於是就產生一個讀不懂的文字了，讀不懂的文章，這就是漢字的字形在不同的電腦裡面或者不同的系統下編碼的情況之下產生很多不同編碼的字形，所以我就想說我們是不是兩位教授，謝教授跟王教授你們是不是有研究過，如果中文這個或者說漢字的系統，漢字的電腦裡面，可以完全放棄字碼，把他像英文一樣，把他變成字母的編碼，或者說每一個部件把他編碼上

去，那雖然部件編碼上面還有字形結構的問題，字形結構也可以把他編碼上去，然後這樣子整個漢字系統裡面就沒有缺字的問題，也沒有這個錯字，不管錯字對字總之他有一個字形的編碼，字形構造的編碼……，這方面有沒有研究？

答：

有關你那個缺字造字的問題是這樣，每個交換碼都定一個字形，他所謂交換碼就是說我這個碼的對映字形長的是什麼樣子，但是你個人在造自己字的時候，這個碼和字形的關係是你自己定義的，你自己定義的那個碼及那個形在跟別人交換的時候就會有問題，意思是說那個形大家可能定義不一樣。但是從字形的結構上來看，一個字形一定有它的結構，它的結構是從它的部件，或者它的位置，根據這個結構定出來的碼可以說是交換碼的交換碼，他比交換碼還更具有這個交換的功能，因為他是根據這個字形結構本身來定義的。另外一個就是應用上，比如說，你剛剛講是不是有一個系統我們都純粹用這個字形結構來表示，當然也可以這樣做，你整個交換碼全部用字形結構來做，跟現在的交換碼是一個可以並行的方式，但是在我們實際的應用，因為我們的系統是架在這個商業系統上，通常是兩個並行的，比如說我們目前有的字形，交換碼有的字形我們就用，因為使用者一定會這樣用，交換碼沒有的字形，可以用我們的缺字表達式，當然，交換碼有的字也可以用我們的表達式，我們也可以把那個表達式再換成交換碼，因為這個對應關係我們都建在資料庫裡頭。

剛剛那位先生提的那個問題，在我們的系統裡面已經完全解決了，我想這個問題很複雜，在這裡一下子可能說不清楚，你假如要繼續了解的話，可能會後找莊先生，好不好。

問：

請教一下謝教授跟莊教授，就是非常高興知道你們已經做了這個形、義，就是所有中文字都已經做成一個非常理論的而且系統化的一個系統，那我想請教就是，我們中文字的特色是形音義合一，那這個音的部分非常複雜，因為我們有上古音、中古音，還有現代音及各地方言還殘存的，所以想知道你們資料庫裡有沒有蒐集這個方言的字形，因為我個人認為如果這個能解決，可以解決我們今天台灣那個吵鬧不休的方言問題，那非常感謝，想知道一下。

答：

謝謝你問這個問題，我想舉個例子來說明，我們實驗室曾經幫中央研究院龔煌城教授印這個西夏文，龔煌城教授當選院士以後出了兩本西夏文的論文集，這個西夏文大家都曉得是非常特殊的文字，我舉這個例子的意思就是說，我們有能力做這個事，也有能力去蒐集方言，那麼我們也跟語言所，事實上我現在也是語言所的兼任研究員，那我們也跟語言所在商談如何建構這個所謂歷史文字學，就是斷



代的文字現象跟詞彙的現象，而且這個工作呢已經在黃教授的引導之下開始做了，但是這個工作花的人力很大，各位知道，我們這個系統做了很久，這要非常感謝中央研究院，因為中央研究院支持這個系統支持了十幾年，才有今天看的這個成績，如果在其他機構做這個事的話，這個計劃老早就被取消掉了，因為我們得到的資源跟支持人力實在是非常有限，那麼因此這個歷史文字、歷史語言學的架構是要做的，不只是歷史的變化，共時的變化，方言的變化，各個時代的詞彙、文字、語意、語音，這些東西我們現在看到的一線曙光可以用資訊科技把他建構起來，這個在文字學、語言學方面應該都是很大的事情，謝謝你提這個問題，我們也正在朝這個方向走，謝謝。

我們的時間大概是到了，我感謝聽眾很合作，在問問題的時候很簡短，我們更要感謝黃老師、王老師、謝老師跟莊老師，做了精彩的報告，謝謝。