

# 後設資料與內容標誌： 淺談數位化的文章和其內容標誌處理（上）

謝清俊 銘傳大學講座教授

【摘要】本文為 2007 年 5 月臺灣大學佛學研究中心舉辦「佛學專題與數位資源之運用工作坊」研習課程之講稿，主題為「後設資料與內容標誌」，談數位化的文章和語意之處理，分別就文章數位化之後其呈現、內容表現手法與語文的應用等生態變遷，來探討後設資料與內容標誌在數位化文章的重要性，並以《心經》實作為例，說明後設資料與內容標誌的重要。

關鍵詞：後設語言；後設資料；內容標誌；情境；語意；數位化文章

## 前言

本次課程的主題是後設資料（metadata）與內容標誌（content markup）。本次課程之前的課已介紹了標誌（markup）和 TEI（Text Encoding Initiative，即文本加碼推動計畫，此處之加碼即指標誌）。TEI 就是一種標誌。最近，標誌很流行，做數位化的人都知道標誌，也有很多人把很多事都用標誌做。今天所講的課，比較偏向於思考方面。談談何謂「後設資料」？何謂「內容標誌」？這兩者有何關係？有何一樣或不一樣之處。

## 一、從數位化看文章生態的變遷

今天的課從數位化看文章的生態談起，因為文章的生態從數位化以後就開始改變了；接著談

文章與後設資料的問題、內容標誌與文章的問題，最後再演繹《心經》版本和內容標誌之間的關係。

大家知道數位化改變了溝通的生態。請大家想想，在沒有網路的時候，沒有用手機的時候，沒有用數位化產品的時候，我們的生活是什麼樣的情形，而現在又是什麼樣的情形。這樣一比較就知道數位化對我們的生活，不管是工作、食衣住行、娛樂，都產生很大的影響。這影響主要的原因是來自溝通（communication）的改變。有人說網際網路真厲害，數位化產品真厲害；其實，是網際網路厲害嗎？不是！是數位化產品厲害嗎？也不是！是網際網路帶來的溝通厲害；它的成本大幅降低，它的功能急遽增長，而使得各種



前所未有的應用，出人意表的層出不窮。

因為數位化改變了溝通的生態，擔任溝通中介的文字紀錄或文章，其生態也必然風行草偃。如：現在網際網路上有很多「輕薄短小」的文章；很多人寫電郵（e-mail），碰到標點符號不好打就換行，所以現在很多電郵寫起來像新詩的形式，那麼美麗。這好像很酷（cool）！但是，如果看看電郵的內容，簡直是支離破碎、慘不忍睹。也有些迎合青少年口味的出版品，圖文夾雜，「圖多字少，膚淺花俏」，甚至於做立體書。諸如此類，無非是受到數位化、網路改變的結果。此外，在學術界，包括教育、傳播、心理學、認知科學、人工智能等，有關青少年閱讀習慣和認知行為變遷的研究報告指出，青少年閱讀習慣和認知行為的改變，可能會改變青少年腦袋裡認知所產生的結構。換言之，時下青少年和我們這種 LKK（老人家）的腦袋，構造都可能不太一樣。像我這一代（近古稀之年），大概都看文字，在我腦袋裡的知識結構都是文字關係建構，比較少圖形建構；但是現在青少年就可能不一樣。同樣的，電腦中的文章經數位化後，存在資料庫裡，這也和傳統的文章生態有所不同，而首先引起異議的，是文章定義（界定）和範疇的問題。

### （一）文章定義與範疇的問題

前面說的，都是數位化帶給文章生態的改變。但是，我們今天不再講這些，只談一個問題，那就是：文章經過數位化存到電腦中時，與傳統的文章究竟有什麼不一樣。

我們把一篇文章存到電腦裡，必須把一些有關背景的情境資料（context information）也存起來，如作者、出版處、出版時間，諸如此類，並

與文章作適當的連接。如果沒有情境資料，所存的文章將無法分類，無法檢索，也就沒辦法構成資料庫，所以，也就沒辦法管理和利用了。所以，文章存到電腦裡以後，就已經和我們一般傳統的文章不一樣。

有一位法國的文學家 Julia Kristeva，她曾提出一個「互為文本（inter-textuality）」的理論，其主要的概念是：世界上沒有一段文字是可以獨立存在的，任何一段文字一定和其他文章一些相關的文字有某些關係。「互為文本」以我們中國人看來並不是新的概念，然而，這個理論可說明：文章存在電腦以後，它不是一個孤島；它的內容、字串必須與其他的文章有點關係。這個關係是怎麼建立的呢？當然，就是加標誌。這也表示：文章存在電腦裡經標誌後，其結構也就變了，變得和傳統紙面的文章不一樣。

另外，文章內容與外界的聯繫（hyperlinks），也可以用標誌來表達。這聯繫不一定是文章和文章之間有聯繫；文章和實物也可以有聯繫。比方說，可以把《紅樓夢》裡的食譜跟台北有名餐館提出類似的餐點連結起來，那麼我們去吃飯的時候就可以試點一道《紅樓夢》裡的菜；又，《紅樓夢》裡提到很多植物、花、草，也可以跟植物園、植物研究所的東西連起來，彼此便有個照應、參照。所以數位化後，文章內容與外界的聯繫也成了文章的一部分，這也引起數位文章的定義與範疇的問題。

### （二）表現內容的手法與語文的應用

一般來講，我們在電腦裡面表現文章內容的手法和以下這些有關：

第一是呈現（presentation）。文章放在電腦裡



以後，怎樣呈現出來：怎麼樣輸出到各種不同的週邊設備、網路設備、螢幕、列印設備……，這是呈現的問題。

第二是內容之外化。舉個例子，中央研究院的二十五史資料庫，採取的是標點版。因為，如果不採取標點版，看起來會很辛苦。因此，文章存在電腦裡以後，標點符號、句讀、標題、文章的章節段落……，還有版面的位置、字體、色彩、加網加邊、美工加工……，以及標誌與內容標誌（content markup），包括通用結構（DTD，如：版面的、結構的、語文的、內容上的……）、標籤集（tag set），諸如此類，都必須增添。為了增添這些東西，就需要用到一個非常重要的工具，叫做後設語言（meta-language）。關於細節本文以下會陸續介紹。

### （三）數位化文章的界定

數位化的文章，該不該包含傳統文章外的情境信息？這問題其實不用回答，因為目前已經這樣做了，例如：

第一，文章與其後設資料，通常都會在資料庫裡面。也就是說，我們把一篇文章、一本書或是一本雜誌放到電腦裡去以後，一定有一個資料叫做書目資料。這個書目資料包括作者、寫作日期、標題、出版處、發行人、字數、……，這些都是所謂的後設資料，不是書的內容資料，是和書的背景有關的資料。

第二，以前出書的時候，不必提供書目資料。一本書進入圖書館，圖書館要做的第一件事就是要編目，編目後才能上架。編目就是整理書目資料。現在的圖書館員比較方便了，現在的出版公司，出書時大都提供一個電子檔的書目資料，入

館時不用再做編目。這是出版界和圖書館界推動資訊共享的成果。

第三，以前投稿一篇文章，就只是投稿一篇文章；現在投稿一篇文章需附摘要、檢索詞和作者的信息（如：作者的姓名、年齡、籍貫、地址、工作職位）等，這些都是背景資料，也就是後設資料。

以上的例子，都和電腦化、數位化有關，這可說明：在電腦中，只有文章本身是不夠的，必須和情境信息一起打包，文章才算完成。

### （四）情境（context）

許多人以為 context 只是「上下文」的意思；其實「上下文」是 context 在語言修辭「情境」下的意義。看一篇文章裡的一個詞、一句話，最接近的情境就是文章裡面該字句的上下文。所以，以「上下文」譯 context，就是在那個情境之下 context 的意思。

情境對文章而言，可泛指文章作成時所有相關的背景，包括：1. 與其他文章相關的背景，如：一篇學術文章後面一定要有參考資料、一定要有註；2. 作者相關的背景，如：作者生平、成書時間、著作時的身心狀態……，舉個比方，如果我們曉得當初曹雪芹寫《紅樓夢》時的心情、際遇是如何？那麼我們再來看《紅樓夢》，理解就會比較深刻一點；3. 時代相關的背景，如：政治背景、經濟背景、軍事背景、天災人禍、社會重大事件……；4. 文化相關的背景，如：三武之亂、滿清的辯子……；5. 社會情境的變遷，如：有電視、手機、網路之後；……。所以，情境這個詞，可泛指除了本體之外，所有外部和這個本體曾有的關聯。



這個情境有一個很重要的特徵，就是：一旦作品完成，情境信息即已固定，且恆久不變。正因如此，這些情境的信息，也就是文章的這些背景信息，可以和文章永遠連在一起，不會改變；而這個合在一起的記錄，就是一個文化歷史的見證。這是一個很重要的觀念，此所以文物之可以做文化記錄，這也是數位化的紮根理論。

情境會影響我們對作品的瞭解，也影響我們對作品的解釋。意義是依情境而定的，情境既已固定，則作者創作的原意亦隨之固定。如：佛陀所說的經，究竟其真實義是什麼，有時我們必須設法把自己的心意放在佛說法的那個情境之下去理解。

文章的意義可有兩層，一是作者的原意和讀者理解到的意思。作者的原意和讀者理解的意思，經常並不完全相同，而且有時可以差很遠。這兩者都屬文章本義的範疇。其次，還有文章的延伸義，也就是說對該文章所作的詮釋意義。例如，我們現在有時候故意把這些古詩詞拿來做另一種解釋，比方說，一個人學成或開悟的境界，像是「眾裡尋他千百度，驀然回首，那人正在燈火闌珊處」。辛棄疾這首詞原來並不是說學成或開悟的狀況，但是王國維把它拿來如此詮釋時，大家都覺得他說得真好。這是王國維對辛棄疾這首詞的詮釋義。

對文章意義的詮釋是對原有文字賦予新的意義，是閱聽者可以依閱聽的情境、生活周遭的情境而作的創新。（注：詮釋是有章法的，可參考釋義學或詮釋學的書籍）理想上，文章數位化後應設法表達文章的這兩層意義，而這兩者相關的情境卻南轅北轍，大不相同。

那麼現在我們可不可以允許在電腦裡的文章由我來詮釋？現在沒有一個電腦資料庫可以做到這樣的程度，因為目前電腦沒有描述情境的功能。情境的描述和語意是非常有關係的。我們了解作者產生作品時的情境，可以幫助我們了解作者表達的原意。如果電腦可以描繪情境，才可能去描述意義、處理意義；若無法描述情境，則無法真正處理語意，而目前的電腦就是笨電腦，根本不懂什麼叫意義。例如，電腦會做 $1+1=2$ ，它了解 $1+1$ 的形式以後，它就說答案 $=2$ ，但是它不了解 $1+1$ 的意思。所以，電腦能不能描繪情境，就是目前一個非常挑戰的研究工作。

情境並不是很陌生的信息，我剛剛講過：像後設資料、書目資料都是情境的一種。然而，如果有一天，我們電腦可以描述一般的情境，這些這些情境存在電腦裡一定要遵從一些標準或規範。為什麼？因為情境信息必須不分國家、種族，甚至於不分電腦機種，都要能夠處理。不能說英文的書目資料是這樣，中文的書目和英文的書目資料不能夠合併，那就很糟糕（早期確實是如此，中、英文沒辦法合併）。所以，當電腦裡要處理情境的時候，需要用一種電腦會處理的通用人工語言（general artificial language），也就是後設語言（meta-language）來描述，而不用自然語言，不用中文、英文、法文、德文……，因為用那些語文，都有隔閡。

後設語言不僅僅可以描述情境信息，文章內容的注疏、註釋，以及文章之間彼此的參照（hyperlink），甚至於文章內容與實物之間的聯繫關係等，也都可以用後設語言描述。後設語言大家也都不陌生，像 HTML、XML 都是。



### （五）數位化文章的表達

文章的結構在電腦中產生了根本的改變，變成以自然語言和後設語言相輔表達的雙重結構：以自然語言寫文章本身、以後設語言描述數位化的文章與外界各種關係。

一篇文章到了電腦裡去以後，就變成兩種語言的表達，文章本身是以自然語言寫，如：用中文寫、用英文寫，但是這篇文章寫好了以後，它和其他文章與外界種種關係，就必須要用後設語言來描述。（見表一）

表一、文章數位化後信息之表達

數位化之文章		表現系統
文章本身		自然語言
文章與外界的關係	情境描述 (metadata)	後設語言
	參照聯繫 (hyperlinks)	
	內容詮注 (content markup)	

表一的向度有數位化之文章、文章本身、文章與外界的關係；文章與外界的關係有情境描述 (metadata)、參照聯繫 (hyperlinks)、內容詮注 (content markup)；使用的表現系統有文章本身的自然語言、文章與外界的關係是用後設語言。所以，文章數位化以後，就形成雙重語言的描述結構，一是自然語言，二是後設語言。像大家用 TEI 標誌一些資料庫的內容或文章，事實上也是這樣雙重的語言結構。

有些學者認為，後設語言愈來愈重要，往後的年輕人除了母語外，其次最重要的不是外語，而是後設語言。年輕人如果有空多學一些後設語

言，多學一些 XML、HTML、TEI，一輩子可以享用不盡，因為它是超越國界、不分種族、文化，能跨越時空、讓你充分表達思想、意念的人類共同工具。

## 二、後設資料

### （一）後設資料的認知

目前一般人對後設資料的認知不是很正確的，所以特別在此和各位說明一下。有人引據國外的文章，說後設資料就是「資料的資料 (data of data 或 data about data)」。有了這樣的說法，許多人便認定：「除了文物數位化的本身之外，所有其他的資料都屬後設資料」。又如，有些老師教學生時，他會把後設資料定義說成「後設資料就是資料的資料」。這些都錯了，都是對後設資料錯誤的認知，都犯了嚴重的錯誤。

說後設資料是「資料的資料」，沒有錯，可是要明白：只是為了闡明後設資料這個概念的性質，並不可將後設資料定義為「資料的資料」。因為，後設資料固然是「資料的資料」，可是並不是所有的「資料的資料」都是後設資料。將後設資料界定為「資料的資料」這種認知，與「不吃豬肉的都是回教徒」犯了同樣的邏輯錯誤。

### （二）後設資料的範疇

現行的任何後設資料，有其固定表達的方式、訂定的規格，以及標籤 (tag) 或欄位 (field) 的選擇和數目等，都限制了後設資料的範疇。這很明顯表示：不是所有的「資料的資料」都是後設資料。要明白數位化的後設資料，不能把資料二分為資料和「資料的資料」這樣籠統的概念去



理解。如果你把電腦裡的資料分成兩種，一個是資料，一個是資料的資料，那你在電腦裡所發展的系統、發展的程式，保證會出問題，因為在邏輯上這變成自我參照，自我參照的邏輯有時候就會出現矛盾。

### （三）時下後設資料的性質

目前的後設資料都是為了某類文物訂定的。比方說，書目資料是一般書籍的後設資料，新聞有新聞的後設資料，玉器、青銅器、畫作、雕刻……等都有各自的後設資料。後設資料既然是描述「某類」文物的資料，那麼就有它的特徵和它的侷限：它適合敘述某類文物的共同現象（共相），而無法顧及個別現象（別相）。

一般而言後設資料敘述的多屬事實、屬性這類較客觀可考的共相資料，也就是說後設資料所描述的都是一些情境資料，是相關的事實、相關的屬性這類比較客觀可以考據的資料，不涉及文本內容的理解、感受、比較、批評，以及詮釋等（別相）。所以，後設資料是可以由具技術專業人士查訪、考證的；但是，它不可以作詮釋。比方說，我們可以考證《紅樓夢》的作者是誰，卻不能詮釋《紅樓夢》的作者是誰。所以，時下的後設資料都是一些屬於事實的陳述和屬性的陳述，是事實陳述和屬性的陳述，這些東西是可以查訪、考證的，但是這些東西是不可以去理解、感受、批評、比較、詮釋。

後設資料的內容是依應用的目的而異，一件數位文物的後設資料可以有許多種。例如新聞稿，對記者而言，有一種後設資料；對報社來說，同一則新聞有編輯用的、管理用的，甚至於是與其他通訊社交流用的各種後設資料。這些後設資

料之間，會有些重複，但也有獨特之處。所以，應用時會要求獨特者能彼此互通，重複者需彼此一致。

人世間的事情常有變化，所以後設資料不會是固定不變的，它是一個時間的函數，會與時遷移，需要花很多力氣更新、維護和保養。有人說在電腦裡建一個資料庫，或是在網路裡建一個資料庫，這個負擔比生一個小孩還沉重。為什麼？因為小孩二十多歲就能夠獨立了，但是做一個資料庫，就要照顧它一輩子，說不定你的兒子、兒子的兒子、孫子，還要去照顧它。

後設資料既然如此複雜，就不是電腦常用的欄位結構可以處理得了的。所以，描述後設資料現在都用後設語言。只有語言才有能力描述後設資料的種種規格和後設資料之間的相容關係，以符合應用的需求。

### 三、內容標誌（content markup）

了解了後設資料再來談內容標誌就方便了，因為內容標誌要照顧的，正是後設資料無法觸及，關於文物個別內容描述的這一部分。以文章而言，對文本內容的理解（解釋）、感受、比較、批評、詮釋等，正是內容標誌的主要工作。所以，內容標誌的重要不言而喻。內容標誌的工作觸及人文、歷史、社會、美學、哲學等學門的核心問題。這些工作需要真正了解內容的專業人士為之。如：故宮有一幅宋朝的畫，把它數位化放在電腦裡，關於這一幅宋朝的畫，它的內容怎麼去理解、去欣賞、去感受，這一幅畫和別的畫有什麼關係，再怎麼去批評這幅畫好在哪裡、不好在哪裡，怎麼去對畫裡的東西做詮釋，是需要教授



級的人物去做。所以，內容標誌的工作，坦白來講，是需要該相關學門真正了解的專業人士來做。這倒不像後設資料只要專業人士做考證、考據就可以了。專業人士做考證的時候，可以去考據這一幅畫的作者是誰，但是他對於這一幅畫，對於美術、對於藝術、對於畫派的理解不見得很內行，他只要考據專業的工夫就勉強可以了。

內容標誌，無論作理解（解釋）、感受、比較、批評或詮釋，均觸及一個人文方面最根本的問題——意義（meaning）和了解（understanding）。談到意義和了解這兩個問題，就踩到了電腦的痛處。電腦碰到意義和了解，到目前為止是一籌莫展，這是認知科學、語言學、記號學等近幾年來致力研究的重點，也是電腦迄今未能處理的痛處。內容標誌正是為了解決這個困局而設。內容標誌事實上有個前提假設，認為電腦不可能直接去了解一篇文章的意義，也不可能直接去了解一篇文章。因此，電腦要能處理語意，這就必須和人充分合作：意義和了解的部分由人去做，後續的處理才由電腦去做。所以，發展電腦做內容標誌的前提是建議一個人機合作的構想，由人（專業人士）負責意義和了解的部分，再由機器來處理其餘的工作。

## （一）內容標誌與語意

### 1. 標點符號

電腦目前不會處理語意，然而我們生活上卻使用了很多的語意工具，只是我們習而不察。例如：標點符號就是處理語意的工具。各位有沒有看過沒有標點符號的文章？文言文很多沒有標點符號，文言文為什麼沒有標點符號，是不是我們

的老祖宗比較笨，沒有辦法發明標點符號？其實不是，因為以前竹簡、木簡，甚至紙張都是很貴的，古人希望在有限的面積容納最大的訊息，所以有意省略了標點符號。試將古文做一個統計，用25史的標點版平均起來兩個標點符號中間字串的長度差不多是4.5；也就是說，如果你加標點符號的話，就會把版面從4.5加到5.5，幾乎加了18%的版面。這可能使「學富五車」要改成學「學富六車了」；或者說，原來一本5斤的書，加了標點符號就要變成6斤。所以，文言文裡省略了標點符號。省略了標點符號，就要付出文章講究修辭的代價：文言文不用標點符號寫，要寫得別人只能看出一種意思，不會解釋成另外一種意思。要學寫文言文最困難的就在這一點。

標點符號用的不同，同一字串的意思就可能不一樣。舉一個不太雅的例子，有一個人家的牆角常常有人去小便，主人很生氣，就在牆上貼了幾個字「閒雜人等不得在此小便」，意思是「閒雜人等，不得在此小便」。但是他沒有點標點符號。有個路人看了以後哈哈一笑就在牆角小便，剛好被主人看到。他很生氣的說：「你不識字嗎？」那個人回答：「我識字啊，你看：『閒雜人，等不得，在此小便』，是你叫我在這裡小便的呀！」所以說，標點符號能改變了文章內容的表現方式，是一種處理語意的工具。

標點符號最重要的作用是把一些內隱的、隱晦的語意，用標點符號把它標示清楚，變成外顯的語意。無標點符號的文章內容較為隱晦（implicit），需經分析、理解的過程才能窺見原意，讀的人水準要比較高。所以，標點符號是一個典型的工具可使語意表面化，讓我們一眼就可



以看得很清楚。換言之，標點符號使內容較外顯（explicit），使隱晦的語意（implicit meaning）變成外顯的語意（explicit meaning）。諸如：私名號的使用，已明顯的標出姓名或機構名稱，減少了斷詞的工作，句點、逗點、分號等，則已將斷句標明，表明前後文的關係。所以，標點符號有將部分文章內容由隱晦轉為外顯的功用。由此看來，標點符號是內容標誌的典型例子。

## 2. 句讀

除了標點符號之外，句讀也是文言文常用的內容標誌。句讀主要用途是作文章內容的標誌：標明明文中之美辭、佳句、警句，或文中不佳之處等；對詩詞韻文，也有用於標示韻腳和朗誦時的間歇者。有些影印的古書中，有前人的句讀：在精采的句子旁加紅圈，在警句旁加點，這些都是做內容標誌。換言之，句讀就是把前人對於文章內容的理解、感受、批評、比較等，以符號的形式點在文章裡面。

標點符號或句讀這類的標誌，都是設計來幫助讀者理解文章內容的；但是它還有另外一個功能，就是它也幫做標誌的人，把他們對文章的理解、詮釋、批評、比較，透過標誌記錄下來。如：我們現在可以買到以前蘇東坡先生做過句讀的一些古本書，甚至我們可以買到一些曾國藩先生做過句讀標點的書，這些對我們來講，我們所買的就不只是了一本書了，這還包括蘇東坡、曾國藩先生他們對這本書內容的理解、詮釋的資料。因此，我們所講的內容標誌和語意，與西方現在講的所謂標誌很不一樣。

電腦的文章標誌，是近年來西方資訊科技發展出來的，如：XML、TEI 等。而漢語文獻的標

誌，則是我們先人留下的智慧。目前的電腦標誌，很多是對文章的外形作標誌，如：通用結構（DTD），版面的、結構的、語文詞彙的……，標籤集（tag set）；傳統漢語文獻的標誌，則側重於文章內容的標誌。

## （二）語意處理的問題

談到文章內容的標誌，我們免不了要談語意處理的問題：也就是，希望電腦有一天能幫我們處理「意義」，幫我們做一些「了解」方面的例行公事。近年來，計算語言學和人工智能均致力於處理意義的研究，也取得一些成果。例如，詞網（word net）、主題圖（topic map）、知識本體結構（ontology）等。然而這些東西在西方所有研究中，還是只做文章的外觀形式方面，如：把一些詞與詞之間的關係標清楚。詞和詞之間的關係當然和語意有關，但是真正這些標誌、這些工作，還只是做形式上而沒有做內容上的標誌。這些研究現在都號稱為語意處理研究，如：word net、主題圖是知識結構，他們都號稱：「我們現在電腦裡已經可以處理一部分語意問題。」是不是呢？坦白講，答案：「是，也不是。」這就看我們對語意處理的界定在哪裡了。

它們將詞彙間的關係在電腦中作了適當的表達（representation），並構成資料庫和研發為數位工具。詞彙間的關係是語意中的一種，將它數位化，對意義的處理是有助益。可是助益有限，並沒有突破性的進展，原因是囿於形式和內容（意義）是一對一的前提，因此，並無能力處理意義的癥結——多義問題（ambiguity）。

如果詞和詞之間的一些關係，這個詞是那個詞的廣義詞、狹義詞、反義詞，這些都算是語意



關係的話，當然他們做的算是語意處理，但是真正的語意處理，不是指做形式方面的東西，是要處理多義的語意問題。ambiguity 很多工程人員把它翻成歧異的，就是不同、有衝突的意思，事實上翻譯成多義反而更恰當一點。

多義問題，是指一個詞依情境變化時有好多不同的意思。例如，作數目字時，「十、拾」通用，如：十元，也可寫成拾元；可是情境變為「路不拾遺」時，就不可以作「路不十遺」。為什麼「路不拾遺」的「拾」只能用「拾」，不能用「十」，因為「拾」這個字是多義詞，它在「路不拾遺」的情境裡面，只能用「撿起來」的意思，所以只能用「拾」；在數目字的情境下，十元、拾元都可以。所以，真正處理多義問題，關鍵就在一個情境表達的問題。在電腦裡，若沒有辦法區別不同的情境，就無法真正處理語意問題。所以，多義問題，簡單說，就是當一種形式可能對應到好幾種意義時，如何作正確選擇的問題。因此，電腦裡要做這些處理，便必須要能夠描述這個意義的情境。這種語意隨情境而轉移的現象，在語言學稱為「義隨境轉」。

人面對多義或義隨境轉問題並無太大難色，所有的自然語言都有濃厚的義隨境轉色彩，因為人很聰明，人可以在溝通的過程當中，或者在閱讀的過程當中，了解相關的背景、相關的情境，對「意義」會作適當的「了解」，可以去區別不同的環境作正確的解釋。所以，人才可以作適當的選擇，作適當的了解。但電腦顯然是沒有這方面的能力。如果電腦無法描述情境資料的話，它就無法作理解的事情，因為語意和情境是絕對非常密切的連在一起。所以，電腦處理意義問題的

先決條件，是要會表達情境，可是目前學界在這方面的努力，還沒有顯著的成績。

【編者按】本文為錄音謄稿，經講者撥冗審閱刊載。

（下期待續）

