

文章的標誌

謝清俊 950518

現在，很多人的電子郵件不喜歡用標點符號。行文遇到需要分斷處，就乾脆「換行」。如此一來，信件的形式好像新詩，似乎別有風韻，可是依其內容看來，就顯得散漫、支離。這也許是「換行」的動作比輸入標點符號方便許多的緣故；圖個自己的方便嘛，那管讀信人的感覺。

傳統的古文大多沒有標點符號^①。現行的標點符號是五四運動推行白話文後訂定的。兩千多年來，古文雖然不用標點符號，並未使讀者誤解了文章的內容。這固然是讀者有比較高的理解能力使然——理解文章時能正確的斷詞、斷句，也是作者使用文字的能力高強的緣故——要高到寫出的文句只允許有一種斷詞、斷句和理解的可能。這是寫古文最基本的功夫——要能寫出這樣的文章才可避免語意混淆，才能派上用場。

古人為什麼不使用標點符號？是他們不夠聰明？可能不致如此吧！筆者曾對整個二十五史的全文^②做過統計：標點符號之間的平均字串長度約 4.5 個字。這個數字可以這麼理解：如果記錄的版面不變，使用標點符號將使該版面的記錄減少約 18.3%；以一篇文章而言，增加標點符號要增加約 22.2% 的版面。換言之，一個人的藏書若需 4.5 個書櫃存放，那麼用了標點符號後就要 5.5 個書櫃；如果一篇書簡^③的重量是 4.5 斤，用了標點符號後就成了 5.5 斤。這也許是古人沒有用標點符號的緣故吧。

讓我們再看看這平均數 4.5 個字的魅力，它也說明古文四、五個字就大致可以表示一個較完整的概念。這也許可以說明為什麼古詩多四言或五言，童謠為什麼常常四、五字成句。古文就可以這麼簡潔！

有了標點符號的確方便不少；至少免除了斷詞、斷句和理解的困擾，使讀寫文章能力的門檻降低許多，讀寫能力也就更為普及。可是這門檻的降低，也意味著寫作能力和理解能力的降低。若無標點符號，讀者需經斷詞、斷句的分析，有時還要參照該文章的情境，才能了解文章的內容。這本事，在使用標點符號後逐漸式微。

標點符號也改變了文章內容的表現方式。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。有了標點符號，則內容較外顯（explicit），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。所以標點符號有將部份文章內容由隱晦轉為外顯的功用。

古文雖然不用標點符號，然而有另一套常用的標誌系統：句讀^④。句讀不是作斷詞、斷句用的，它的主要用途是作文章內容的標誌——標明文中之美辭、佳句、警句，或文中之不佳處、敗筆等；對詩詞韻文，也有用於標示韻腳和朗誦時的間歇者。句讀的作用也可說是使文章的內容外顯。

標點符號或句讀這類的文章標誌，都是設計來幫助讀者理解文章內容的^⑤。它也幫做標誌人，把他們對文章的理解透過標誌記錄下來。所以，從漢語文獻的角度看來，文章的標誌有兩大功能：對讀者而言，它能將文章隱晦的內容外顯，方便讀者理解文章；

對專家而言，則是可將他們對文章的理解記錄下來留作後用。

把文章數位化存入電腦時，也需要做標誌（markup），不過此標誌和前述者有所不同。為電腦作文章標誌時，需使用一種電腦能夠解讀的標誌語言（markup language）^⑥，如 HTML、XML，將電腦無法理解的文章版面、篇章段落結構、圖、表、註解、參考文獻等格式標示清楚，讓電腦知道以使電腦能幫我們處理文章。所以這標誌是直接為電腦做的，其間接的受益者才是使用電腦的人。這種做法是典型的人機合作：電腦無法理解的事由人來做，人們不耐煩做的例行事務交給電腦執行，相輔相成。

目前為電腦做文章標誌，絕大部份還只做到將文章呈現的各種形式（如上段所述的各種例子）外顯—用標誌語言的標籤（tag）把這些形式標清楚讓電腦知道；對文章內容的標誌，則幾乎還沒有開始嚐試。

電腦的文章標誌是近年來西方資訊科技發展出來的，而漢語文獻的標誌，則是我們先人留下的智慧。比較這兩種文章標誌，雖然表現得南轅北轍相去甚遠，然而也顯示出合則兩利的前景。例如，電腦一直對文章內容的處理力有未逮，而漢語文獻的標誌卻正提供了一種處理內容的方法。漢語文獻的標誌顯示，文章標誌是可以處理語意的，而目前的電腦一遇到語意問題就一愁莫展。這種差異正表明二者的互補。

再者，做文章內容的標誌，不是電腦工程師可以做的，需要了解文章內容的專家來做。這情形正好提供人文學者一個絕佳的機會加入文獻數位化的行列。如果人文學者能用標籤把他們的知識，也就是對文章的理解、真知灼見，表達給電腦知道，那麼，久而久之電腦將匯集大量的人文知識。果真如此，那麼，一種嶄新形式的人工智能（artificial intelligence）即將誕生。

且讓我們拭目以待。

註：

- ① 古文獻有標點符號者不多，且標點符號較時下使用者簡略。可參考：嚴文郁，《中國書籍簡史》，臺北市，臺灣商務印書館，民國八十一年十一月初版，圖四和圖五：明朝三色和五色套印的《楚辭》和《文心雕龍》樣張，其上即有句號和幾種句讀符號。
- ② 這是十多年前利用中央研究院的「二十五史全文資料庫」所做的統計。該資料庫用內建的是標點版的二十五史，總共約四千萬字。
- ③ 指竹子或木片編成的簡冊。
- ④ 句讀中「讀」之讀音為「ㄉㄨˋ」，同「逗」。句讀之例請參考①中之圖四和圖五。
- ⑤ 古書中幫助讀者理解文章內容的，不僅僅是此處所談的文章標誌，像注疏、校勘、眉批、批改文章的符號、乃至編輯符號等，也都可視為是幫助讀者理解文章內容的。
- ⑥ 標誌語言是一種後設語言（meta-language）。關於後設語言、數位化文章的性質和結構等，請參考本專欄〈數位化的文章〉一文，95年4月。