

字碼、字型、輸入法的概念

廖偉成

電腦裡的文字是透過字型來呈現。字型或叫做字體，是一套向量圖庫，一個字就是一張圖。字型要依照字的編碼來製作字的圖片。編碼就像是每個字的編號，例如「𨮑」的 Unicode 編碼是 U+2028E。我們用輸入法打字，打出來的不直接是字，而是編碼，編碼再對應字型圖庫，而呈現出具體的文字樣貌。

漢字的編碼，不同系統的收字量與排序方式不一。如「大五碼」系統共收 13,060 個漢字，但不包含客家語的「𨮑、𨮑」。另外有 Unicode 編碼系統，在 1.0 版時收錄 20,902 個漢字，但亦不包含「𨮑、𨮑」；在 3.0 版時增收 6582 個漢字，才有「𨮑」的編碼；在 3.1 版時再增收 42,711 個漢字，才有「𨮑」的編碼。雖然有了編碼，不表示螢幕上能看得到這個字，仍要有字型依照編碼，製作「𨮑」這張圖才能顯示。如手機等行動設備所預裝字體，若支援的字圖不足，則「𨮑」這個字仍無法呈現。

在漢字編碼不多或字圖設計數量不足的資訊發展階段，為了解決客家語電腦「缺字」問題，或有以「造字」來處理。文字編碼系統預留一個私人使用的造字區，使用者透過造字軟體畫出自己所需的缺字。編碼位置雖相同，但每個人所畫的字圖並不一樣。以至於同一份文件以不同的電腦開啟，顯示出不同的字，或被稱為亂碼，其實編碼沒有亂，只是字圖不一樣。

字型設計者依照共同的編碼位置，為各別字型設計獨特的樣貌，而形成不同的字體。如「標楷體」是楷書風格，「新細明體」是模仿古代印刷字風格，其他如硬朗的黑體、可愛的娃娃體等，都讓同一個編碼的漢字有不同的變裝秀。不過，即便 Unicode 11.0 版已收錄 87,888 個漢字，但字體設計者並不會為每個字體設計這麼多字。如客家語的「𨮑」雖能以「新細明體」呈現，但卻無法以「標楷體」或「娃娃體」呈現。另外，某些包含拼音或注音的字體，是在設計字圖時，便同時畫上標音，所以只要一更換字體，就能同時顯示出漢字與標音。

關於聲調，客家語拼音的聲調標示於字母的右上方，約占半個漢字的空間。但部分字體如「標楷體」，聲調卻是設計為占一個漢字的空間，如「ngai ˇ」。要標示得正確又美觀，若透過軟體「上標」功能，既繁瑣又費事，此時只要改換一套字體，如 Calibri，則能完美標示成「ngaiˇ」，因該字體已將

聲調符號設計為上標樣貌。

當前，若能在現有字體中補足客家語所需漢字，則客家語的漢字才可能以楷體、黑體或娃娃體等多樣貌呈現。雖然如此，若這些新製作字體並非預裝於電腦系統，其他設備無這字體，亦無法正確顯示客家語漢字樣貌。此時，我們可在其他設備安裝所需字體，或是將文件轉成 PDF 格式，或是透過軟體的內嵌字體功能來達到文件傳遞後不漏字。另外，在網頁上也可透過內嵌字體技術，即便瀏覽者的手機沒有安裝該字體，仍可以正確顯示客家語漢字。

因此，客家語的漢字與拼音，要在資訊設備中暢通無阻，除了要有編碼，還要有字型設計對應的圖，另外還要預裝於資訊設備中，才能實現客家語資訊溝通的無障礙環境。