

常用字彙、詞彙與國語教學*

周 法 高

中國語中通常用“字”作單位，寫在紙上就是漢字。一個漢字通常代表一個“詞素”，詞素是語言中有意義的最小單位，也有人叫做“語位”，相當於英文的morpheme。語言中比詞素大一點的單位叫做“詞”，英文叫做word。一個詞可能由一個或一個以上詞素所構成，例如：“人”是一個詞素，也是一個詞；“桌”和“子”都是詞素，“桌子”是一個詞。漢語中詞素和詞素之間的界限非常分明，通常都寫成一個個的漢字；可是詞和詞之間的界限就很不明確了。英語中詞和詞之間的界限非常清楚，在詞和詞之間都用空白來隔開；但是詞素和詞素之間的界限就沒有漢語那麼清楚了。

中國歷代累積下來的漢字，多達五萬左右，可是我們日常應用的漢字不過四五千。如何把使用次數最多的那些字找出來，編成常用字彙，那就是語文教育專家的工作。特別是編製國民學校國語教科書時，需要參考。過去有若干學者根據日常讀物的用字加以統計，得出每一個常用字在若干萬字的讀物中出現的次數。從1921年起，六十年間從事常用字彙統計的，不下十餘家。直到1967年國立編譯館主編的《國民學校常用字彙研究》（以下簡稱《國校字彙》）由中華書局出版了。該書從報章、國民學校課本、兒童作品、課外讀物、廣播稿和民衆讀物等六種資料共753,940字當中，經分析統計而得4,864個字。其中擬定爲常用的字有3,861個，次常用字574個，備用字429個^①。前有《國民學校常用字彙研究報告》50頁，立論公允；後有部首索引，甚便檢查。這可以算是常用字彙的一部有分量的作品，可以拿來和聯合報社在該項研究完成的時候所出版的《聯合報中文自動鑄排機常用字彙表》（以下簡稱《聯合報字彙》）參看，該表以聯合報抽樣作主要的對象，整理出2,356個字來，其所根據的資料，顯然有異於國民學校兒童讀物，但在互相對照之下，只有25字沒有在國校字彙中出現，其差異尚不及百分之一，可見國校字彙具有相當的代表性。（《國校字彙》頁39-40）

* 本文是作者於一九八一年八月十七日在“七十年暑期中日中國語言教學研討會”的講稿。

① 《國校字彙》前500個最常用字共出現568,384次，佔總數75%強；前1000個最常用字共668,903次，佔94%。

1972年國立交通大學工學院計算與控制學系出版了林樹編著的《中文電腦基本用字研究》(以下簡稱《電腦用字》)“綜合現存已有的七種字彙研究結果,參酌現代社會通用的四種字彙資料,加以分析統計,核對排比,最後共得8,532個單字,然後按各字出現頻率的多寡排列等第,並按其使用情形分為四級,計‘最常用字’1,857字,‘次常用字’2,068字,‘間用字’2,182字,‘罕用字’2,425字,編訂而成《中文電腦基本用字表》。”(《電腦用字》頁21)

《電腦用字》一書純係綜合各家的統計,加以平均,不能算是一種原始性的研究,不過也可供參考之用。《國校字彙》一書編製的經過很為審慎,方法也相當嚴密,根據的讀物的字數也夠多。不過因為其目的是求得國民學校的常用字彙,其結果和一般社會上的字彙可能在頻率上以及收字上有點出入,例如《國校字彙》和《聯合報字彙》比較的結果,後者就有二十五字沒有在《國校字彙》出現;後者有三十二字在《國校字彙》中列為“次常用字”,有五字在《國校字彙》中列為“備用字”。在女部有“妓”“娼”“嫖”,貝部有“賄”“賂”“贓”,言部有“誹”“謗”“諭”等字。(以上見《國校字彙》頁40)。

在日本,對於常用漢字的選定,也是根據統計的結果的。在1956年出版的90種雜誌上大約280,000個漢字中,使用了3,328個不同的漢字;在1956年出版的報紙上大約1,000,000個漢字中,使用了3,713個不同的漢字,在1975至76年出版的70種書籍、雜誌和辭典中,表示出有4,520個不同的漢字在日本出版的報章雜誌和書籍中使用着。其中最常用的1,000個漢字構成了上述材料的90%,最常用的1,800個漢字構成了上述材料的98%—99%。(以上據第12期國語審議會審議結果報告《新漢字表試案》頁1,昭和52年,即1977年,國語審議會印行)

中文常用字彙,是根據字彙在若干萬字的讀物中出現的頻率而編訂的;對於國民學校國語課本的編訂,是相當重要的。此外,對於初級字典的編訂,也具有參考的價值。不過一兩千個常用字的組合,可以構或上萬個詞彙;哪些詞具有較高的頻率,就可以構成常用詞彙。因此,詞彙出現的頻率,就值得研究了。

早在1944年,桑戴克(E.L. Thorndike)和勞奇(Irving Lorge)就出版The Teachers Word Book of 30,000 Words,把一百萬個英文單字出現的次數表列了出來,並將其和另數種研究結果並列比較。魏斯特(Michael West)在1953年出版了A General Service List of English Words一書,依照每字各種意義應用的頻率,詳作分析。(據《國校字彙》P.2)

1969年蕭約翰(John Robert Shaw)出版了《新境界最新階梯英語字典》(The New Horizon Ladder Dictionary of the English Language),次年由蕭約翰和蕭珍妮(Janet Shaw)修訂出版;1976年由楊景邁教授主編中文譯本。據蕭氏原序說:

本字典是根據一般英文文章裡最常用的5,000個英文字所編撰。該5,000字係美國新聞總署參考多種常用字彙表,英語教學字彙表,以及收集多位英語教師意見所得,歷經十二年使用,測驗及修改以編纂各種深淺程度不一的英文書籍供給不同年齡和不同興趣的讀者們之用。

爲了教學使用方便，特將此 5,000 字依其常用度分成五個階梯，每階梯約爲 1,000 字。每字後括弧之數目字代表該字所屬之階梯。(1) 代表該字爲最常用之一千字之一，(2) 代表該字爲最常用之第二個一千字之一；餘(3)，(4)，(5) 類推。

最近顏元叔教授主編的一部英漢字典，也在常用的五千字後面註明(1) — (5) 的數字，雖然所收的字數遠超過五千。我覺得同樣的方法可以適用於漢語教學，可以使學習者增加學習的效率，特別是爲外國人編製的課本和詞典，更應當採用這種方法。

《國校字彙》頁41說：“國民學校詞彙分析工作，也已著手進行。”可見常用詞彙的重要性，已經受到注意了。香港中文大學教育學院有簡國銓講師曾於1967年進行詞彙頻率的研究達二年左右，在該院院刊《學記》第一期(1968年6月)曾發表《漢語字彙詞彙研究計劃及第一期工作報告》一文(頁73—92)。其辦法是將香港各種小學課本中的總字數452,466個，詞數261,437個，根據電報號碼輸入電腦，加以統計。可惜簡氏後來在加州大學得到教育學院博士後，此項工作即行停頓，據說卡片已經廢棄，真是可惜得很。

在歐美詞彙頻率的研究較爲普遍，而詞素頻率的研究却沒有聽說進行；但在中國，字(即詞素)彙頻率的研究較爲普遍，而較大規模的詞彙頻率的研究却還沒聽說。其原因是：中文中詞和詞之間沒有顯著的界限，而漢字却絕大多數是代表漢語中的詞素的，只有“玻璃”、“枇杷”等少數詞素不是單音的。至於“詞”在中文中就有時不易確定了。艾偉在《漢字問題》一書中(頁104—107)列出27條原則，簡氏在前文中列出20項原則，可供參考。不過由於二家都不是研究語言學的，其原則有若干條是不足憑信的②。

在六十年代和七十年代，荷蘭海牙的Mouton書店，由Alphonse Juilland主持，出版了西班牙語(1964)，羅馬尼亞語(1965)，法語(1969)和中國語諸種語言的詞彙頻率詞典。《中國語詞彙頻率詞典》(Frequency Dictionary of Chinese Words)是在1973年由

② 趙元任教授《中國語的文法》丁邦新中譯本 p. 82 註 13 說：

“把詞解釋成可以隨意停頓的最小字組，要比說成是最小的獨用語式更加好些。”見〔Rulon S. Wells 在“Immediate Constituents”一文〕115 頁，n. 58。“詞是……句子中前後可以停頓的任何一個片斷。”見 Hockett, *A Course in Modern Linguistics*, 167.

P. 101 說：

Bloomfield 的定義認爲詞是最小的自由語式(所謂自由就是指能單獨成話)，咱們已經說過這個定義太嚴格，所以加上了停頓，可能的停頓，跟常用性(或自由運用性)，作爲輔助的辨認法。咱們不妨這麼開始說，造句詞是前後可停的最小字組，但得馬上補充說明，像“啊”、“吶”、“麼”這些語助詞，雖然跟前一音節緊緊結合，但事實上是跟整個詞組或句子不可分的。因此，Bloch 認爲，要是一個語式的直接成分，有一個是兩個或兩個以上的詞構成的詞組，而另外一個是不能用停頓分開的字組，那麼後者就是一個詞。這樣就不用因爲有個語助詞，而硬把整個長句子或詞組，叫作複合詞了。例如：“英國的國王的女兒”(the king of England's daughter)，不會因爲其中有第二個“的”字，就使“英國的國王”成了一個詞，就像英文裏不會因爲 the king of England 加了一個 's 就成了複合詞一樣。

劉森 (Eric Shen Liu) 博士編撰出版的。該書採用了不超過二十五萬個詞的語料，按語料的性質分爲五類，每類大約包含五萬個詞：(1) 小說類，(2) 戲劇類，(3) 論說類，(4) 報章雜誌類，(5) 科技論文類。

劉書把每一個詞後面都注明詞類，可是既不注漢字，也不注明英文的同義詞，在他所舉出來的三千個常用詞和附錄中的大約七千個罕用詞中，有不少不容易寫出相當的漢字來，有時同一個詞類的同音詞有好幾個，就不大容易確定了。例如：劉書 R. 133 “gung¹ shr⁴” 名詞，35次，可以寫作“工事”、“公事”、“公式”、“攻勢”四種不同的漢字。加註漢字和英文釋義這兩點，都很不容易做到；可是另有一個辦法，是可以實行的，却沒有採用。如果在按英文字母排列的詞彙當中，把首字同用一個漢字的詞都歸到一起，讀者也許比較容易找出原來的漢字。可惜在該書第一部分按字母排列的詞彙當中，却沒有採用這個辦法，以致於好多詞都不能確定其漢字，這不能不算是該書的一個嚴重的缺點。

劉書除了導論而外，把 3,059 個較常見的詞在第一部分按英文字母排列，第二部分按出現次數的多少排列，分爲十組，每組三百個詞，最少的五次；每一個詞除了出現次數而外，還列出該詞在五類性質不同的讀物中分佈的係數 (coefficients of dispersion) 和用法的係數 (coefficients of usage)。關於上述二種係數的計算方法，在該書導論中有詳細的說明，可作爲今後統計詞彙頻率的參考，而這些都可用電腦順利達成任務的。

此書劃分詞類有好些不妥的地方，例如：(1) “她很美麗”，(2) “她的美麗”和 (3) “美麗的她”中的“美麗”一詞，通常都歸入形容詞一類，在 (1) 例中作述語，(2) 例中作名語，(3) 例中作修飾語。而該書却分屬 (1) 述詞 (predicatives)，(2) 名詞 (nominals)，(3) 修飾詞 (modifiers) 三類^③。像這樣的情形有上百個的例子。又如：“應用”在該書中分隸 (1) 名詞 (nominals)，26次，(2) 述詞 (predicatives)，81次，(3) 修飾詞 (modifiers)，19次三類。爲了明瞭起見，我姑且舉三個例子來說明：(1) “這種原理的應用是很有價值的”，(2) “我們應用這種原理”，(3) “這是一門應用的學科”。我通常認爲“應用”在上述三例中同屬一類，不過在 (1) 例中用作名語，(2) 例中用作述語，(3) 例中用作修飾語。因爲現代漢語中名詞通常不能用作一個句子的述語，但是動詞和形容詞却可以用作名語和修飾語，所以我把“應用”一詞歸入動詞。不過也有人把“應用”一詞根據它在句中的用法分屬三類，例如林語堂的當代漢英詞典在“應用”一詞下註明：

v.t. , to apply; adj., applied (science, art)

如果再加上名詞的用法，就要解作“application”了。我總覺得在字典中這樣劃分詞類固

③ 趙書 p.90 引梅祖麟說：

凡是詞一定能填到幾個可以互相變換的架構之中，比方形容詞 (即性質動詞 quality verb) 能够填進：(A) 名 + 動式，(B) 動 + “的” + 名式，跟其他“很……”、“最……”、“比” + 名……一類的架構之中。一個相當於詞的單位如果能填進 (A) 式，就一定也能填進 (B) 式，因爲兩式可以互相轉換。

屬未嘗不可，但是在計算頻率的時候，我還要採用我的辦法把“應用”一詞的三種用法合併起來計算頻率，因為在漢語中缺少語形變化(inflection)和抽象名詞的詞尾(例如英語中的名詞詞尾-tion)等等，不像英語那樣。又如在現代漢語中名詞和動詞(加上“的”或不加)都可以作修飾語，例如“時代的青年”和“青年的時代”中的“青年”，我都認為是名詞，可是在劉書中，“青年”用作名詞149次，修飾詞14次，在當代漢英詞典中，“青年”解作：“n. & adj., youth; youthful”。像以上兩類的例子在劉書中不下幾百個，如果採用我的辦法，頻率的計算可以大為改觀。此外，我覺得該書其它還有許多不妥的地方，在這裡也不必詳細討論了。

現在我把劉書最常見的十個詞，拿來和《國校字彙》、《電腦用字》比較如下：

中國語詞彙頻率詞典	國校字彙	電腦用字
(1)“的”用作記號(marker)18,072次	(1)26,738次	(1)71,669次
(2)“是”用作述詞(predicative)5,836次	(3)8,916次	(3)28,178次
(3)“他”用作名詞(nominals)4,462次 (用作修飾詞301次)	(10)4,506次 (外加“她”792次,“牠”368次)	(8)19,698次
(4)“不”用作動詞修飾語(verb. m.)3,817次	(8)6,834次	(4)26,601次
(5)“一”用作修飾詞(modifier)3,231次 (用作動詞修飾語148次)	(2)14,147次	(2)33,940次
(6)“我”用作名詞(nominal)3,037次	(4)7,461次	(6)23,374次
(7)“了”用作助詞(particle)2,587次	(5)7,176次	(5)26,451次
(8)“有”用作述詞2,421次	(6)6,894次	(7)20,632次
(9)“在”用作介詞(preposition)2,380次	(9)6,631次	(11)16,581次
(10)“人”用作名詞1,962次	(11)5,855次	(9)19,675次

我們看了上表，可以看出出現次數最多的都是單音詞，劉書次序也和《國校字彙》、《電腦用字》差不多。“的”通常讀輕聲，附屬於它前面的或分；不過“的”和詞尾不同的地方是：它時常附屬於詞組的後面，好像英語“King of England's”的“s”附屬於“King of England”一樣。不過這種情形在英文很少見，而中文“的”和日文“の”附於詞組後，却很常見，所以趙元任先生把它叫作“助詞”(particle)，而不叫作“詞尾”(suffix)；劉書叫作“marker”，即王力所謂“記號”。④

④ 參註2所引趙書 P. 101。