

中文可讀性公式芻議

陳世敏

一、什麼是可讀性

一則新聞，凡是一般讀者願意一讀，讀了能瞭解新聞說些什麼而又能夠記得住的，就具有可讀性（readability）。「美聯社寫作委員會」（AP Writing Committee）顧問傅萊區博士（Dr. Rudolf Flesch）把上述意義再加擴充，他認為新聞寫作只有在下列四種條件下，才具有最高的可讀性（註一）

- 一、清楚：意義要清楚，句子結構要簡單。
- 二、簡短：句子愈短，衝擊力愈大。
- 三、完整：要回答一般讀者所要問的問題。
- 四、有趣：使讀者感覺到自己與所發生的事，在某方面有關聯，他才會加以注意，讀下去。

很多人談可讀性，總會下意識強調「人情趣味」，因此要從法院新聞中發掘不爲人所知的人情部份，成爲法院記者應有的技巧（註二）。這些人解釋可讀性，多偏重於趣味的要求，忽略了它在講求精確上的效果（註三）。

隨著報紙的發展，新聞寫作的要求愈來愈高。報紙也像其他的商品一樣，在自由的競爭市場中，第一要務就是引起讀者的注意，爭取讀者。對報紙而言，發行的競爭也就是生存的競爭（註四）。通常的

辦法是使讀者對這一份報紙感興趣。「現代新聞記者手冊」的作者瓊斯(John Paul Jones)說過，記者未來要掙一口飯吃，必須知道怎樣把句子寫得短寫得清楚，並且有愉悅而具有「人情」的風格(註五)。「專業的新聞記者」一書作者賀亨柏(John Hohenberg)更進一步說明報紙不僅要使它所報導的新聞，能够被讀者瞭解，而且更要使讀者不致誤解(註六)。

上述意見無非是要使報紙可讀(readable)。傅萊區博士在「美聯社寫作委員會」上對可讀性作了下面的結論：

「我們首要的任務，在於告訴一般讀者一眼能抓住的是什麼。每一段導言——以及隨後的每一段文字——都應清晰、銳利而有趣，使他不得不繼續下去(註七)。

綜合說起來，可讀性應有兩種涵意，一是簡明動人，二是值得一讀。

不僅報紙新聞應力求可讀，就是其他通俗性的、專門性的書籍文字，也應力求可讀。

二、英文可讀性公式的影響

美國傅萊區博士是研究可讀性公式最有成就的一位學者。傅萊區首先於一九四三年，以博士論文「可讀體裁之標準」(Marks of Readable Style)，提出計算可讀性的公式，迄今已二十六年。可讀性公式的研究，早在一九二〇年代就有人做過了。一九二一年凱遜(H. D. Kitson)以音節和句子的長度，來測量兩家報紙和兩家雜誌內容是否易讀。一九二三年里佛利(Bertha A. Lively)和普雷西(S. L. Pressey)以難字的百分比作標準，他們所說的「難字」，是指不在桑代克(Thorndike)常用字量表

以外的字（註八）。

美國可讀性的公式很多，常用的有三種：

第一種是甘寧（Robert Gunning）的公式，他的標準有二。第一是每句的平均字數，第二是每百字中超過二個音節的字數。兩者相加後乘以零點四，即得含霧指數（Fog Index）。

第二種是戴爾和柴爾（Edgar Dale & Jeanne Chall）的公式，標準是：一、每句的平均字數，二、超過戴爾三千常用字彙表以外的字數。

第三種是傅萊區（Rudolf Flesch）的公式。他的公式歷經三度修正，初期公式着重句子的長短，字眼的難易和人稱字眼的多少（註十一）。修改後的公式計算平均每句幾個字和平均每字幾個音節，並另有測量人情趣味的公式（註十二）。最新的可讀性公式含有兩要素：一、傳播力（Communication energy），指句子生動的程度；二是具體性（Realism），是計算具體字眼的多寡（註十一）。

此外，目前通行的公式，還有三、四十種。一九三〇年李維倫茲（Alfred S. Lewerenz）發明了一種很有趣的可讀性公式。他計算以W、H、B開頭的字彙字數以及以I、E開頭的字數，前者是易字，後者是難字。這一年他又另創一種公式，計算源於盎格魯撒克遜的字和源於希臘、羅馬的技術性字彙，前者是易字，後者是難字（註十四）。

第二次世界大戰以來，報紙新聞的最大改變，在於專家們呼籲使報紙更爲可讀。一般的趨勢是使用短字、短句、短段落，使受過九年教育的人易讀、易記、易瞭解。

美聯社在傅萊區介紹可讀性公式後，該社新聞稿，句子的長度由二十七字減至二十三字。導言每個

字由一點七四音節降爲一點五五音節。合衆社（United Press）在接受甘寧的建議後，把句子簡化成適合於受過十一點七年教育的人，以前合衆社的新聞，則需受過十六點七年教育的人才看得懂（註十五）。

傅萊區對可讀性的貢獻，可用美聯社執行編輯高德（Alan J. Gould）的話來說明：「傅萊區的觀念，在破除寫作習慣的舊格式方面，有重大貢獻。」（註十六）

甘寧和其他人的公式，功效亦同。這些公式，主要在抖散五個W一個H必須包含在導言中這種規則。現在報紙的寫作，趨向簡短，導言僅容納最重要的若干部份（通常是Who和What），省略其餘，或安排在後面各段落中（註十七）。

可讀性公式究竟是否有助於良好的寫作，看法各有不同（註十八）。新聞界詰難傅萊區等人的理由是：爲何在動手寫新聞之際，要受到短字、短句的限制？爲何要迎合低水準的讀者？如果新聞中的五個W一個H附屬於人情味之下，讀者如何吸收新聞中的主要事實（註十九）？

可讀性公式只是測度現有文字的閱讀難易程度，並非是一種增加可讀性的工具；譬如食譜，抱著食譜的人不一定就能燒出色香味俱全的好菜來。傅萊區曾說：

「我只能重複我在「說白話的藝術」的序言中所說的話，我担心若干讀者會希望有個寫好文章的神奇公式，會對我這種簡單的標尺感到失望。其他渴望求得語言正確性的人，會沉迷在這些簡單的規則和計算中，無視於白話文的原則。我所希望的是，讀者不要把這個公式看得太認真。除了把它當做粗略的估計外，不要希望太多。」（註二十）

好的寫作，不是機械的，不能以數學測出。對於文字所談到的觀念或背景一無所知的人，即無以瞭解。易讀，不僅有賴於句子短，也有賴於思想。紐約時報（*New York Times*）編輯馬柯爾（*Lester Markel*）即認為，通俗雜誌不僅字短、句短，思想也簡單。因此，真正的困難在於如何就眼前的事，給予讀者必需的消息，讓他們能瞭解。每一行業的專門字彙，均非隔行的人所能瞭解。記者是寫給大眾看的，他們長久以來即遵守這樣的規則：不要高估讀者所知，不要低估他們的智力。今天此一規則仍然適用（註二十一）。這是可讀性公式的持平之論。

可讀性公式的功過，似可以麥基（*Laine K. McKee*）的話作結論：

一、對需要改進寫作的人，可讀性公式是有用的工具，但不可一意盲從。短句就不一定比長句好，例如佩因（*Tomas Paine*）的名句：“*These are times that try men's souls*”，如改成“*These times try men's souls*”則節奏與韻味全失。

二、較長的句子，讀者需有較高的教育程度才能瞭解，這有統計數字為證。美國四年級學生的造句平均長度是十一點一字，九年級是十七點三字，大學畢業班是二十一點五字。

三、寫作者一旦瞭解字、句複雜所造成的影響，可讀性公式的任務即已達成，寫作者自不需再亦步亦趨，受公式的拘束了（註二十二）。

三、中文可讀性公式的困難

目前未見有中文可讀性公式，少數研究中文報刊可讀性的論文，均係直接借用英文可讀性公式。中

英文句子結構和文字構造，不盡相同，直接借用英文可讀性公式來測量中文，其間所遭遇到的削足適履的狼狽情形，不言可喻。截至目前，大規模的中文可讀性研究，首推于宗先先生「台灣報紙可讀性之研究」（註二十三）。

原作者以傅萊區最新可讀性公式，研究中文報紙的可讀性。中英文的結構不盡相同，英文公式未必每一種都適合於中文，因此，原作者選用傅萊區的最新可讀性公式。因為這個公式着重兩種因素，即（一）具體性（realism）；計算特定或具體的人、事、物。（二）傳播力（Communicative energy）；計算語勢強弱的數字（註二十四）。

不過，原作者也指出應用這項公式會有偏差：

一、在計算具體性方面，結果可能偏高。這與中文的特質有關。因為中文是表象主義，中國的語言在表現具體事實方面，是非常活潑的，而在說明抽象觀念方面，則較乏力。

二、在計算傳播力方面，可能有偏低的現象，這與英文的表達方式有關。根據傅氏可讀性原則，加強語調的暗示，也算是傳播力因素的一個單位。例如，英文「I can't...」，有強調的意思在內，中文即無法表達（註二十五）。

偏差的程度是多少，原作者並未說明。直接引用這項公式，是否適合，不無疑問。因此有必要根據常用的數種英文可讀性公式，歸納出適合中文的公式來。

傅萊區的第三種公式，既不適合我們直接借用，我們不妨從傅氏的第一種、第二種公式，再就甘寧公式、戴爾和柴爾公式，分析比較，尋出可讀性公式的幾個共同要素是什麼。這四個公式，歸納起來，

不外着重兩點：

- 一、文句的長短：上述四個公式均以句子的平均長度衡量可讀性的高低。
- 二、用字的難易：傅萊區第一種公式，以可讀程度與字首字尾數成反比；第二種公式以全篇音節愈多愈難讀。甘寧公式是求超過兩個音節以上的字數；戴爾和柴爾公式則計算超出三千常用字彙表以外的字爲難字。（註二十六）

文句長短與易讀與否有直接的關係這一點，大致不會引起爭論（註二十七）。關於以音節多寡作爲難易字的標準，不無商榷的餘地。例如，按傅萊區的公式，beautiful, candlestick等字，屬難字；而Ohm, erg, id等字屬易字。這顯然是極不合理的。倒是制度常用字彙表，以超出字彙表以外的字爲難字，較爲客觀、實際。

本文擬改良甘寧公式，作爲測量中文可讀性之用。甘寧（Robert Gunning）的含霧指數（fog index）公式如下：

$$\text{含霧指數} = (\text{每句平均字數} + \text{每句音節數}) \times 0.4 (\text{字數})$$

此項公式的第一項因素是求句子的平均長度，中文自可適用；第二項要素以二音節以上的字爲難字，不妥處已如上述。戴爾和柴爾的公式中制定三千常用字彙表，比硬性規定兩音節以上的字爲難字，較爲合理，中文可讀性公式借用過來，也無困難。這樣，剩下來的問題是：中文常用字彙有多少字？根據什麼標準決定那些字是常用字，那些是非常用字？

四、中文常用字彙表

康熙字典所收錄的漢字，達四萬七千餘字，日人諸橋轍次編的大漢和辭典中，所收字數為四八、九〇二字（註二十八）。值得注意的是：漢字的總字量雖多，一般人常用的只是其中一小部份。一般字典所收錄的，平均只有七八千字。在這七八千字當中，還是有很多罕用字，所以一般人日常閱讀的和應用的，實遠較七八千字為少。到底一般人要認識多少字才能讀書、看報和寫作呢？

由研究而得的字彙，其功用是顯而易見的。從字彙表上，我們可以知道一般人在日常生活中，究竟需要應用多少字，需要那一些字，那些字出現最多，應用最廣，在那一方面用得較多。學習的人和寫作的人，有了這種根據，使用文字，盡量不超出常用字彙表，則寫出來的文字，一定更易讀、易瞭解、易記，達到「可讀」的最高目標。

在英文常用字彙的研究中，以桑代克（E. L. Thorndike）、洛奇（Irving Lorge）、韋斯特（Michal West）等人的貢獻，最受人注意。桑代克和洛奇，把一百萬個英文單字出現的次數，用表列出（註二十九）。韋斯特則把最常出現的兩千字，依照每字各種意義應用的頻率，詳作分析，編成「一般應用英文字表」（A General Service List of English Words），對語文教學者的幫助很大。

民國五十二年以前，國人對常用字彙的研究，至少有七種（註三十）。其中較值得注意的是王文新的「小學分級字彙研究」，他分析兒童作文和國語課本，經過多次淘汰，共得到三、七九九字，分配到小學的各年級裏去。此外，教育部於民國二十四年頒布「小學分級暫用字彙」，計二、七一一字。

這些研究，有些已經佚散，有些不合時宜，只有參考而無實用的價值。最新、最大規模的字彙研究工作，是國立編譯館「國民學校常用字釐訂委員會」於民國五十六年完成的「國民學校常用字彙研究」（註三十一）。他們從報章、國民學校課本、兒童作品、課外讀物、廣播資料、民衆讀物以及憲法、國歌總共七五三、九四〇字當中，經分析、統計而得四、八六四個字。其中以三、八六一個字爲常用字，五七四個爲次常用字，四二九個字爲備用字。

第二種是「聯合報中文自動鑄排機常用字彙表」。聯合報爲了選擇這些單字，曾按報紙各版所刊文字，除副刊和廣告以外，分別逐字統計其出現的頻率，依次選出二千五百常用字，再參斟過去各家對於常用字彙研究的資料，最後加上二十個標點符號，挑選所得之常用字計二千三百七十六個，以部首檢字法編成「中文自動鑄排機常用字彙」（註三十二）。

第三種字彙表是「寫真植字機研究所自動照相排字機中文字彙表」，把常用字分爲三級，包括二十四個標點符號和括號，共計二千五百五十八個字（註三十三）。

上述三種，後兩種的功用和性質很相似。以第一種和第二種常用字彙表對照起來，在聯合報的二、三、七、六常用字彙中，只有二十五個字不會在「國民學校常用字彙表」中出現；聯合報另有三十二個字在後者被列爲「次常用字」，有五個字被列爲「備用字」。細查聯合報這二十五個，完全不會在「國民學校常用字彙表」上出現的字，在女部有「妓」「娼」「嫖」，貝部有「賄」「賂」「贓」，言部有「誹」「謗」「諭」等字，都是在報紙上常用而在兒童讀物中少見的字（註三十四）。

以這三種字彙表作測量可讀性難易字的標準，大致可行。現在我們常見的字，未必是三、四十年前

也是常用字，這三種字彙表，時效上有很大的優點，與現代的生活關係較密切，是可以斷言的。

分開來說，如果是測量報紙的可讀程度，常用字標準，自以「聯合報中文全自動鑄排機常用字彙表」和「寫真植字機研究所自動照相排字機中文字彙表」較宜；如果用來測度一般書籍、雜誌，則以「國民學校常用字彙表」較適合，這樣可避免若干個新聞用語，在一般人所使用的字彙中變成「罕用字」。

本研究擬就「國民學校常用字彙表」中，取二千三百二十五個字為常用字，超出這些範圍為難字，作為中文可讀性公式常用字彙數的標準。說明如下：

一、兒童的字彙是否就是成人的字彙？目前台灣省兒童的就學比率，達百分之九十五以上，成人的用語雖不就是兒童用語，但成人學過兒童在國民學校中所學過的「字彙」，是顯而易見的，不同的只是「詞彙」不同。本研究既以「字」為單位而非以「詞」為單位，選擇「國民學校常用字彙表」作標準（註三十五），當不致有明顯的偏差。聯合報字彙未在本字彙中出現的，不及百分之一。

二、「國民學校常用字彙表」上，列為常用字的有三千八百六十一個字，何以只選取二千三百二十五個字作常用字？聯合報的字彙表，僅二千三百七十六個常用字（其中有二十個是標點符號），根據該報的報告（註三十六），缺字並不嚴重，這些常用字如應用得當，已屬够用。年來使用結果，新聞稿件的缺字率僅百分之三左右。換言之，如以聯合報字彙表為標準，一個讀報的人，平均一百字中可碰見三個「非常用字」，這樣的數字，對一般僅受過國民教育的讀者，是很正常的現象。

三、根據研究，三民主義全書共有十六萬三千二百九十六個字，其中所用的單字不過二千一百三十五個。再如一部禮記不過二千三百六十七字；莊子三千二百多字；墨子二千六百多字；荀子二千七百多

字。因此，選取二千三百二十五個字作為常用字（註三十七）作標準，似無不妥。

另一項顧慮是，常用字標準過寬，在運用公式時，各類材料中每百字僅一、二難字，相差不大，則失去了以難字多寡解釋可讀性高低的原意。好在常用字字數的標準，可鬆可嚴，視需要變通，這裏不過是爲了研究方便而已。

五、中文可讀性公式

茲參考上述三種公式，試擬中文可讀性公式如下：

（一）求句子的平均長度：在文章中以任意選擇（random sample）方式，選出一段或數段，每段以一百字爲準，可多幾字或少幾字，但必須以句點結束。再計算這一段中有幾個中斷語氣的標點符號——包括句號、逗號、冒號、分號、破折號、驚嘆號和疑問號，引號則合併在冒號或句號之中，不另計算——然後拿這數字來除總字數，即求得每句的平均字數。

（二）求難字數：計算這一段中，超出「國民學校常用字彙表」內前二千三百二十五個字的字數，就是難字數。

（三）每句平均長度加上難字數之後，乘以零點七，即得：「含霧指數」。如果「含霧指數」是六，表示這段文字一般受過六年教育的讀者，可以看懂。餘類推。

爲求精確，一篇文章可選取數段作樣本，再平均其結果。抽樣時盡量避免選取文章中顯然十分特殊的段落；此外宜避免以第一段作樣本，根據學者的研究，第一段往往比較難懂。

以下就 $(\text{詞字長字總數} + \text{難字總數}) \times 0.7$ 的公式，求得街坊常見的一般性雜誌的可讀程度：

東方雜誌	十點一
傳記文學	九點一
春秋	八點三
讀者文摘	七點八
皇冠	七點七
婦女雜誌	七點二
模範	七
王子	六點五
連環圖畫	六以下（註三十八）

事實上，刊物的內容，並不一定與句字的長度和難字數成正比。「看得懂」，僅是就文字而言，至於內容與含霧指數的關係，應作如下的補充：

含霧指數	適合閱讀的年級
五以下	四年級以下
五至六	五至七年級
六至七	七至九年級
七至八	九至十一年級

八至九	十一至十三年級
九至十	十三至十五年級
十以上	十六年級以上

「可讀」的另一項要素是「人情味」。要知道一篇文章的「人情味」程度，可參用傅萊區的第三種公式——即求具體性和傳播力的公式；惟因中英文性質不盡相同，此公式應用於中文，仍有缺陷，前已述及。除非再加修正，否則目前宜作參考之用。

六、中文可讀性公式的檢討

不論中英文，不論使用那一種公式，可讀性公式有兩點力所不逮之處：

一、測量不出思想。

二、測量不出文字風格。

依賴可讀性公式，不能寫出好文章，理由至為淺顯；因為寫作不是機械的，好文章無法以公式得之。但可讀性公式仍有其實用價值，例如，你若想向美國婦女雜誌投稿，文章的「含霧指數」超過七的，利用的可能性便比較低了。從事大眾傳播或公共關係的人，也可以藉此公式，提醒自己的文稿要保持「易讀」，達到老嫗能解的程度，以接近更多的讀者。

中文使用可讀性公式，最大而又難以解決的困難是：中文不以「字」為思想單位，而以「詞」為思想單位。隨便以「等閒」一詞作例子，雖然這兩個子都是國民學校的字彙，但初中學生不一定懂得這一

詞彙的意思。成語所遭遇的困難亦同。中文的「生字」有限，但「生詞」則可因人創造，不斷出現。

「字」的研究，可以說已告一段落，我們希望有人研究「詞」，那麼可讀性公式以「常用詞」作標準，更能符合中文的特性。

註 解

- 註 一··Writing for the AP, The Second AP Writing Handbook, P. 3.
- 註 二··John Paul Jones, Making the Reader Understand, The Modern Reporter's Handbook, P. 23.
- 註 三··潘家慶，「從語意學看寫作」，報學三卷十期，第三十二頁。
- 註 四··Robert E. Park, "The National History of the Newspaper", Interpretations of Journalism, Frank Luthermott & Ralph D. Casey, ed., P. 125.
- 註 五··見註二。
- 註 六··John Hohenberg, The Professional Journalist, P. 67.
- 註 七··見註一。
- 註 八··Blaine K. McKee, "Readability Formulas Can Aid Good Writing", Public Relations Journal, July 1967, P. 8.
- 註 九··Robert Gunning, The Technique of Clear Writing, New York, McGraw-hill, 1952.
- 註 十··Edgar Dale & J. S. Chall, A Formula For Predicting Readability, Educational Research Bulletin, Ohio State University, 1948.
- 註 十一··Rudolf Flesch, The Art of Plain Talk, Harper & Brothers, New York, 1946.
- 註 十二··Rodolf Flesch, The Art of Readable Writing, Harper & Brothers, New York, 1949.
- 註 十三··Rudolf Flesch, How to Make Sense, Harper & Brothers, New York, 1954.

註十四·See note 8.

註十五·Curtis D. Mac Dougall, Interpretative Reporting, Third Edition, PP. 87-88.

註十六·Rudolf Flesch, The Art of Readable Writing, Forward.

註十七·See note 15, P.90.

註十八·(一) See note 15, PP. 90-92.

(二) See note 8, PP. 8-10.

註十九·See note 16.

註二十·See note 16, Preface.

註廿一·See note 15.

註廿二·See note 8.

註廿三·見于宗先「台灣報紙可讀性之研究」，報學二卷六期，十八頁。該書是政大新聞研究所碩士論文，國立政治大學社會資料中心藏。

註廿四·同上註，二十四頁。

註廿五·同上註。

註廿六·這三千字是百分之八十的四年級學生認得的。

註廿七·例如，「我參加了在台北市中山堂舉行的第九屆工程師年會」，就不比「我參加了第九屆工程師年會，

該會在台北市中山堂舉行」難讀。

註廿八·國立編譯館主編，「國民學校常用字彙研究」，台灣中華書局出版，頁一。

註廿九·E. L. Thorndike & I. Lorge, The Teacher's Word Book of 20,000 Words, 1944.

註三十·同註二十八，頁五—十。

註卅一·同註二十八，頁十七—三十。

註卅二·王惕吾，「聯合報式中文全自動鑄排機的創造和運用」，聯合報出版，頁三。

註卅三：羅敬典，「新聞紙印刷應走的路線」，報學三卷八期，頁一二六。

註卅四：同註二十八，頁四十。

註卅五：「國民學校常用字彙表」的材料來源，包括報章、書籍和廣播材料等成人所使用的字彙和詞彙。見該書頁十一—二十一。

註卅六：同註三十二，頁五一—六。

註卅七：這些字是「國民學校常用字彙表」上，出現頻率在二十次以上者。根據王清波「國民學校現行國語課本國字初現課次及重現次數之分析研究」（高雄市政府出版）一書指出，五十年小學暫用課本十二冊中，有生字二千二千九百三十個。

註卅八：（一）「今日世界」的含霧指數是九點七七，艱難程度僅次於「東方雜誌」，究其原因，是句子太長的緣故，而句子長，與該雜誌多半是翻譯作品有關。「今日世界」句子長度，平均每句是十三點二個字，「東方雜誌」的句長雖為文中所提各種雜誌之冠，也只不過是每句十二點六個字。

（二）根據甘寧的研究（見註九），美國各大雜誌的含霧指數是：

- 大西洋月刊……………十二
- 哈潑雜誌……………十一
- 時代雜誌……………十
- 讀者文摘……………九
- 婦女家庭雜誌……………八
- 真實自白月刊……………七
- 連環圖畫……………六以下