

詞頻表在語言教學上的應用

【語文教育及編譯研究中心副研究員 白明弘】

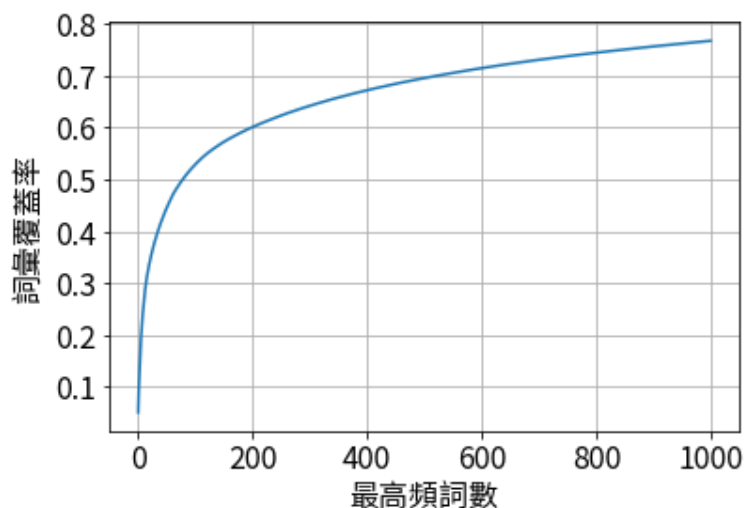
詞彙覆蓋率是閱讀能力的基礎

閱讀能力是接收新知最重要的管道，更是學生學習的關鍵基礎。根據語言教育學者的研究，要能輕鬆閱讀理解一篇文章，至少要熟悉其中 95%-98%的語詞（此百分比稱為**詞彙覆蓋率**）。當詞彙覆蓋率越高時，代表閱讀過程中遇到的生詞越少，閱讀理解的困難度也就越低。為儘早培養學生自主閱讀的能力，提高詞彙覆蓋率已成為現代語言教學的重要目標。

詞頻表和詞彙覆蓋率的關係

儘管如此，大量記憶詞彙未必能有效提高詞彙覆蓋率。我們以**美國當代英語語料庫**（Corpus of Contemporary American English，簡稱 COCA）統計的詞頻與覆蓋率的關係來觀察，當學生學會了英語中最高頻的 100 詞時，詞彙覆蓋率大約可達 50%。但後續學會排名 101-200 的詞時，覆蓋率卻增加不到 10%，而且排名越後面的詞，對覆蓋率的貢獻越少（參考圖 1）。這代表如果一個英語初學者專挑冷僻的詞彙學習時，詞彙覆蓋率的累積將發生停滯的現象。這就是為什麼現代語言教材的編輯非常重視詞頻的原因。由高頻詞開始學習可以在最短的時間內達到最高的詞彙覆蓋率，這是基於語言學習效率的考量，可以幫助學生儘早達成自主閱讀的目標。

圖 1、統計 COCA 語料庫中，詞數與詞彙覆蓋率的關係



有些人提出，只要大量閱讀文章就能提升語言能力，無須特別考慮詞彙的學習順序。此見解和詞彙覆蓋率的觀點並不衝突，但也存在部分盲點。當我們透過大量閱讀來學習語言時，越高頻的詞出現的機率也越高。所以即使沒有刻意安排，高頻詞仍會優先學到，符合高頻詞優先學習的要領。但若所選擇的文章超出學習者的詞彙能力太多時，不但閱讀過程中因生詞太多而充滿挫折感，而且對詞彙覆蓋率的提升也十分有限。這就是為什麼近年來推廣的「廣泛閱讀」(Extensive Reading)強調讀本必須分級，以簡單有趣的讀本，提升閱讀的速度與流暢度。因為當生詞太多時，就會產生眼球移動變慢、理解力下降等問題¹。

除此之外，對非母語教學來說，以大量閱讀來學習詞彙仍有不足之處。以 COCA 詞頻表為例，頻率排名第 5,000 左右的詞（如：offender, ego, compassion, triumph, verify 等），在文章中平均每 10 萬詞才出現一次。如果學生一年無法閱讀超過 10 萬詞，這些詞一年將複習不到一次，造成這些中頻詞很難進入大腦的心理詞彙庫中，使心理詞彙量進入停滯的狀態。這也就是為什麼非母語的學習特別注重詞頻的原因，即不僅要注意高頻詞優先學習，還要兼顧語言暴露量不足的情況下中頻詞學習停滯的問題。

母語教學需要考慮詞頻表嗎？

第二語言的學習特重詞頻的安排，相對而言，母語學習似乎比較忽略詞頻的順序，是否在母語學習的過程中，不需要考慮詞頻的安排呢？在國外針對母語學習者所做的閱讀與聽力的研究中發現，詞彙覆蓋率和閱讀能力，甚至和聽力的關係都是高度相關的。其中微小的差別是，母語者猜測生詞的能力略高於第二語學習者。也就是說，母語學習者的詞彙覆蓋率也是很重要的。不過因為學習者是沉浸在母語環境中，高頻詞在日常生活中就必然常出現，以致於越高頻詞就自然越早學會。而且在母語環境中語言暴露量充足，即使是中頻詞彙也能在日常生活中補足。所以在語言教學上，並不需要特別依詞頻順序安排來提升詞彙覆蓋率。

然而，這並不意味著在母語教育環境中，其他學科教科書的撰寫不需要考慮兒童的詞彙能力。事實上，我們分析美國的學科教科書發現，不同年級的詞彙使用仍然依循高頻詞優先的順序。在小學一年級的社會教科書中所使用的詞彙大約落在排名前 4,500 詞，一直到小學五年級約落在排名前 15,000 詞左右（參考圖 2）。小學一年級科學教科書中所使用的詞彙大約落在排名前 8,000 詞，一直到小學五年級約落在排名前

¹ https://erfoundation.org/guide/ERF_Guide_Trad_Chinese.pdf

25,000 詞左右（參考圖 3）。此一分布顯示，在美國小學教科書所使用的詞彙仍然考慮學生的詞彙能力。因為當教科書的詞彙超出學生的能力太多時，會嚴重影響學生的閱讀與理解，而使教科書失去應有的功能。

圖 2、各年級社會科學教科書使用詞彙頻率分布。

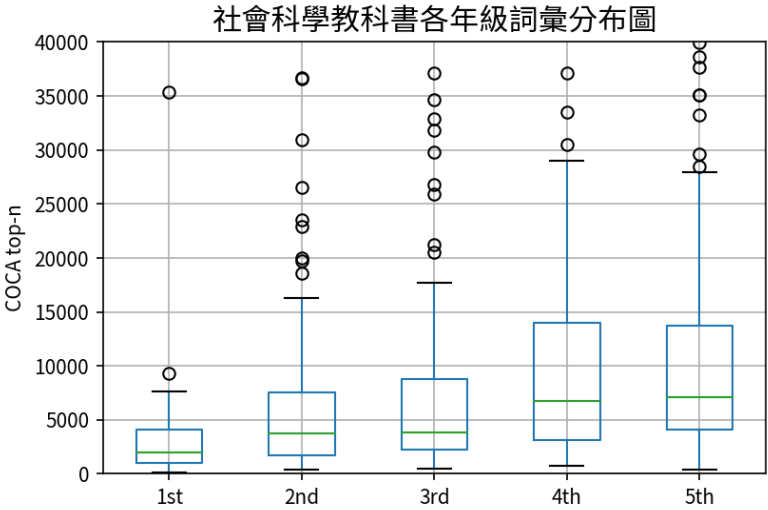
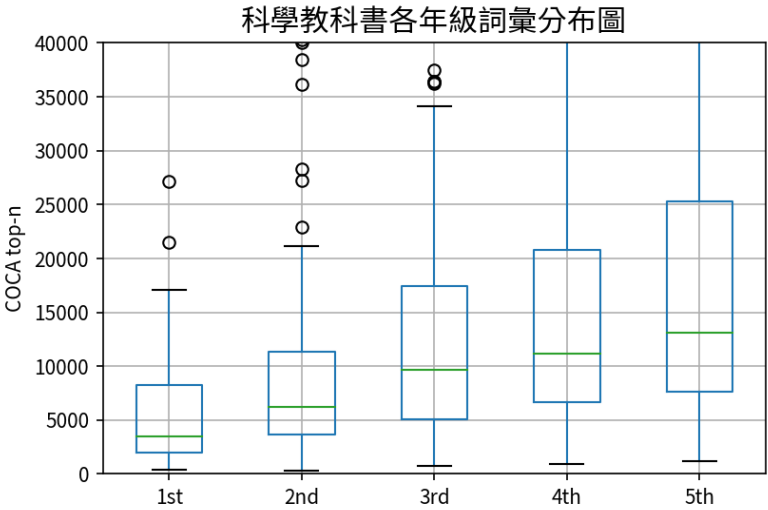


圖 3、各年級科學教科書使用詞彙頻率分布。



另外，雖然沉浸在母語環境中的母語學習者，可以很自然的形成越高頻的詞就越早學會的情況，似乎不必重視詞彙的教學編排。然而，此一情形在中文裡並不全然適

用。英文是一種表音文字，大部分語詞在學生掌握發音原則後就能念出。所以在閱讀過程中，可以透過詞語的發音和母語產生連結。但中文不是表音文字，文字和發音間沒有系統性的規則，即便是形聲字也是如此。例如：「抱、袍、雹」三個字的現代發音都不同，沒有發音規則可遵循。中文字形的辨識與書寫是獨立於母語學習的，必須在學校中花幾年的時間才能完成。因此，中文的閱讀在先天上比表音文字困難。儘管國語屬於母語，但中文字的教學卻和第二語言教學有著類似的困境，學生必須儘早學會足夠的高頻中文字，才能達到自主閱讀的目標。所以，國語教材的編排仍須參考字頻安排，優先學習常見中文字。

中國古代的啟蒙教學特重道德教育，教材多以《三字經》、《弟子規》、《二十四孝》等德育教材為主。南朝梁周興嗣作《千字文》一篇，由一千個不重複的漢字組成，是少數專門為中文字識寫所設計的教材。《千字文》對中文字的學習的影響極大，甚至傳到日本、韓國成為中文字學習教材。然而，若以詞彙覆蓋率的角度來看，這些啟蒙教材在文字學習效率上都非常低，即使是《千字文》也沒有考慮到文字常用性的問題，所以從學習效率的角度來看，傳統的教材並不適合當作識字的基礎教材。

結論

一般以為詞彙量是閱讀能力的重要基礎，但更確切的說，詞彙覆蓋率才是影響閱讀能力的真正關鍵。因為詞彙量的累積未必能有效推升詞彙覆蓋率，初學者若把學習時間花在冷僻的語詞對閱讀能力的提升並沒有太大的幫助。詞頻表的用途主要是提供語言教學在詞彙編排上作為參考，將高頻詞優先安排進教材中，使學生能夠在較短的時間內達成足夠的詞彙覆蓋率，及早達成自主閱讀的能力。同時在非語言教科書的編寫上，也必須搭配各年級學生的語言能力，透過詞頻表可以提供詞彙的常用性訊息，才能避免閱讀教科書的額外負擔。

另一方面，由於中文屬於非表音文字，文字和發音沒有系統性的規則，所以中文字形的辨識與書寫是獨立於母語學習的，必須在學校中花幾年的時間才能完成。這是中文閱讀在先天上比表音文字困難的原因。亦即，儘管中文在臺灣屬於母語，但中文字的教學卻和第二語言教學面臨類似的困境，學生必須盡快學會足夠的高頻中文字，才能達到自主閱讀的目標。

資料來源

白明弘 (2021)。華語文教學通用詞頻表和華語文搭配詞研究。國家教育研究院個別型計畫案成果報告 (NAER-2019-029-C-1-1-B5-01)。新北市：國家教育研究院。連結網址：<https://rh.naer.edu.tw/handle/dd2y7>