

## 這個詞語有很難嗎？電腦幫您作篩選

### 語料庫科技在語文教學的應用（IV）

【語文教育及編譯研究中心副研究員 吳鑑城】

「這份教材用語我認為已經很淺顯了，學生怎麼還是看不懂？」，「學生反應教材中的詞語很多都沒見過，導致閱讀上很挫折，但這些詞語不是蠻常見的嗎？」這樣的疑問是否曾在您的教材被反應用詞過難時，浮現心中呢？

雖然教材編輯者都會考慮學習者的程度設計合宜的教材，但由於每個人的成長學習過程不同，心理詞彙庫的豐富度乃至於對於每個詞語的熟悉度，不免因人而異，也造成用詞難易度判斷上的困擾。為了使華語文教學能有可參考之標準，本院建置了「臺灣華語文語料庫」（Corpus of Contemporary Taiwanese Mandarin，簡稱 COCT），透過語料庫蘊含語言真實使用情境的特性，以 COCT 為基礎，邀請學者專家共同研發，歷時 6 年完成分為 3 等 7 級的「臺灣華語文能力基準」（Taiwan Benchmarks for the Chinese Language，簡稱 TBCL），其中更包含了具級別資訊的漢字、詞語以及語法點列表。

#### 國教院教材編輯輔助系統

TBCL 雖提供了分級詞表可讓編者參考所用詞語的難度級別，然而逐一查詢，或當發現難度過高，思考可替換的詞語以及再度確認級別，仍將耗費編者的寶貴時間。為此，本院研發了教材編輯輔助系統 (<https://coct.naer.edu.tw/sentedit/>)，本系統除了可快速分析教材內容的詞語難度等級外，更有關聯替換詞語功能來輔助編者選擇適合的詞語。

關聯替換詞語功能是應用 Google 所提出的 Word2Vec 方法，該方法透過大量文本資料，藉由轉化每個詞語於文句中的前後鄰近詞語成為向量，代表該詞的語意。當兩個詞語運用在文章中，他們的鄰近詞語群很類似時，也意味著這兩個詞語語意或是運用的情境都會是比較接近的。像是「開心」和「快樂」，抑或是「提升」與「增進」，都會有較高的向量相似性，藉此，可有效的尋找關聯（相似）度高的詞語作為推薦替換詞語。以下，就以一應用實例來為大家介紹此輔助系統：

設想當正撰寫著一個文句：「他藏匿在大門後，沒有被發現。」並且想先了解此文

句中詞語的難易度時，就可以輸入本句於國教院教材編輯輔助系統，此時，依照設定注(拼)音資訊的類型，系統已經先自動加注對應資訊，便於後續教材使用(見圖 1)。

圖 1、國教院教材編輯輔助系統示例



在送出文句進行分析後，系統首先就會提供文句中各個詞語根據 TBCL 詞語分級標準所屬之等級，以及整體等級分布情形，作為教材編輯者先行瞭解目前所撰寫文句的詞語難度(如圖 2)，標記為 X 者表示未收錄於分級詞表中。緊接著，編輯者若想調整使用詞語的難度，即可利用前述技術所研發出的關聯詞語替換選擇器，進行詞語替換。以「藏匿」為例，若編輯者設定的教材使用對象非進階學習者，藏匿一詞已屬第 7 級，對學習者可能已經過於困難，此時僅須點選該詞，系統即會提供與藏匿較為近似的詞語，且同時提供詞語等級，例如第 4 級的「躲」、「藏」，第 6 級的「躲藏」，以及同屬第 7 級的「藏身」、「隱蔽」。編輯者即可挑選等級較低之詞語降低教材難度。反之，亦可將等級較低的語詞替換為高等級來加強難度。(如圖 3)

圖 2、文句中詞語等級資訊

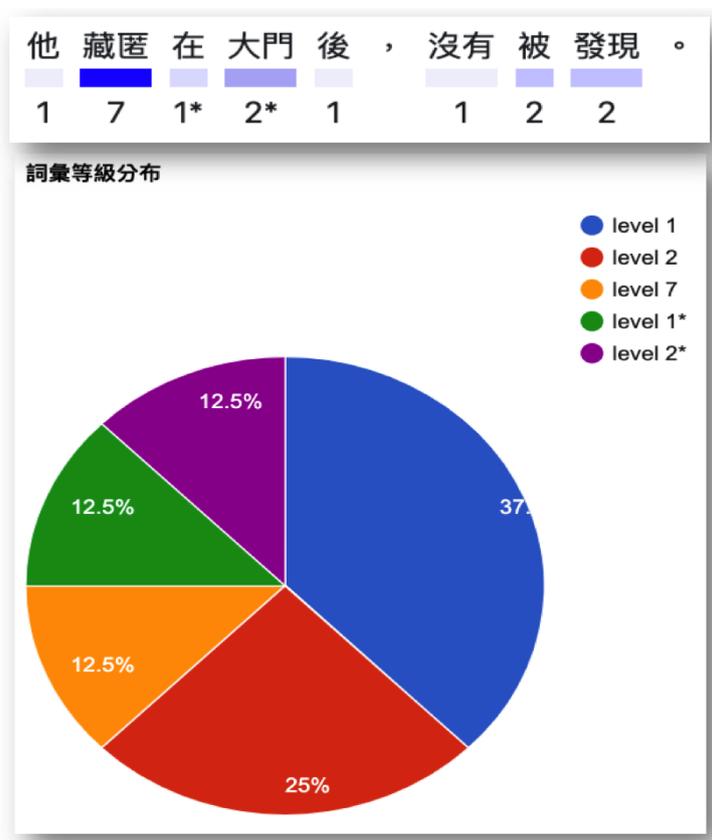


圖 3、挑選合適等級之詞語進行替換

分級標記 詞彙等級分布 詞表 關聯替換詞語選擇器

分級詞表：國教院詞語分級表

1 他	<ul style="list-style-type: none"> <li>✓ 7 藏匿</li> <li>6 躲藏 0.65</li> <li>7 藏身 0.57</li> <li>X 潛入 0.55</li> <li>6 搜查 0.53</li> <li>X 窩藏 0.53</li> <li>X 搜捕 0.53</li> <li>4 躲 0.53</li> <li>7 隱蔽 0.53</li> <li>4* 藏 0.52</li> </ul>	1* 在	2* 大門	1 後
1		1*	2*	1
X ,		2 被	2 發現	X 。
X		2	2	X

本系統已獲得許多華語文教學單位的推薦，包括國家華語測驗推動工作委員會、國立空中大學數位華語文中心等等，希冀藉由此一結合語料庫能力基準及資訊科技的應用，能減輕教材編輯者的辛勞，同時讓教材的難度更為適中，進而提升華語文教材

的品質！

#### 資料來源

林慶隆、柯華葳、吳鑑城、白明弘、陳茹玲（2019）。《建置應用語料庫及標準體系》期末研究報告。國家教育研究院研究計畫成果報告（編號：NAER-107-12-F-1-01-00-1-11）。新北市：國家教育研究院。連絡網址：  
<https://www.grb.gov.tw/search/planDetail?id=12562546>