

測量詞彙的透明度！AI 幫你解開中文詞彙的祕密

【語文教育及編譯研究中心助理研究員 黃淑齡】

想像你正在觀察玻璃，有些玻璃非常透明，讓你能清楚看見另一側的景象；有些則較為模糊，需要仔細辨認；還有些完全不透光，阻隔了你的視線。語義透明度就像這樣，反映了一個詞的整體意義能在多大程度上從其組成部分推測出來。

讓我們來看看一些有趣的例子，透明度高的詞就像清澈的玻璃，如「腳踏車」（用腳踏的車）、「握手」（握著別人的手）、「天亮」（天空變亮），這些詞只要將詞彙組合就能輕易得知意義。相反的，透明度低的詞則像磨砂玻璃，像是「江湖」（不是真的江與湖）、「放心」（不是把心放在某處）、「白眼」（不是白色的眼睛），這些詞沒辦法望文生義，沒學過之前不會自動進入我們的心理詞彙庫。

智慧助手來解謎

研究團隊巧妙運用最新的人工智慧技術，開發出能像人類一般判斷詞彙透明度的自動評估系統。這就像給電腦裝上了一副特殊的眼鏡，能看穿詞彙的層層面紗。系統主要採用三種技術或模型，其中，**BERT 語言模型**就像一位精通上下文的語言專家；**GPT 技術**則如同一位能理解詞義關係的解謎高手；最後，**大規模語料庫分析**則猶如一座龐大的語言資料寶庫。結合這三者，我們期待這套系統能為每個詞彙給出一個語義透明度的分數。

重大發現

研究過程中，團隊發現了一些令人振奮的結果，也就是約 **80%**的中文常用複合詞都具有較高的語義透明度。這個發現為中文學習帶來重大啟發：學生只要掌握基本詞素（例如「天」和「亮」），就能輕鬆理解大部分的詞彙（如「天亮」）。換句話說，在龐大的中文詞彙系統中，真正需要特別記憶的，其實只有那 **20%**語義較不透明的複合詞。

過去的研究發現，英文詞表只要背 11,000 個詞就能掌握 95%的閱讀用語，中文卻要記 40,000 個。現在這個發現徹底顛覆了我們對中文詞彙學習的傳統認知，讓原本看似艱鉅的詞彙學習任務變得更有策略和信心。

具挑戰性的是，語義透明度並非一成不變。就像同一個場景，玻璃在不同光線下會呈現不同的清晰度，**詞彙的透明度也會隨著使用者的年齡和語言程度而改變**。以「計程車」為例，對成年人來說，「計程」反映了這種交通工具的計費方式，透明度較高，是個簡單的詞；但對兒童而言，可能需要特別解釋才能理解。這個發現提醒我們，在設計教材時，必須考慮學習者的特點，特別是針對學童，要重新訂定更符合他們認知發展的語義透明度識別門檻。

創新教學的新方向

基於這些發現，研究團隊提出了一套革新性的教學建議，就像要建造一座堅固的語言大樓，我們需要從地基開始，逐層築起學習的階梯。

首先是**分級教學策略**。讓低年級學生從基本語素和容易看透的詞彙開始，像是「書包」「天亮」這樣一眼就懂的詞；等到中高年級，再逐步引入「客氣」「馬虎」這類需要更多解釋的詞彙，讓學習循序漸進。

其次要**做好詞表優化的工作**。這就像整理衣櫃，我們需要特別收藏那些難解的「珍品」（語義透明度低的詞彙），但可以適當省略一些「基本款」（透明度高的複合詞）。透過這樣的篩選，讓學習詞彙的數量維持在合理範圍，避免造成不必要的負擔。

更重要的是，研究發現**系統化理解和類推學習複合詞**是很有效的方法。例如教導學生認識詞彙結構時，可以一次介紹一組相關的詞，像是「看不起、看不上、看不慣」或「冷門、冷場、熱門」等。這樣不僅能加深理解，還能觸類旁通，舉一反三。

最後，在**編寫教材及讀物時，要特別注意內容的分級**。根據學習者的程度來調整詞彙的透明度，為較難理解的詞彙提供更豐富的上下文支持。這就像為不同程度的讀者量身打造合適的故事，讓每個學習者都能找到最適合自己的學習材料。

展望未來

這項突破性研究不僅為中文教學打開新視野，更為未來發展提供了無限可能。透過 AI 技術測量詞彙透明度，我們可以製作更精準的分級教材、提供個人化的學習建議、優化閱讀理解策略，及發展智慧型學習工具。就像有了一把能開啟語言寶庫的鑰匙，讓中文學習變得更有效率、更有趣！

資料來源

黃淑齡（2025）。國民小學各領域常用多詞單位之結構與語義透明度自動辨析研究。國家教育研究院研究計畫（NAER-2025-012-C-2-1-D1-01）。執行日期：2025-01-01 至 2025-12-31）。連結網址：
<https://www.grb.gov.tw/search/planDetail?id=17245396>