



詞辨是非——謝志成反芻辨文認字

轉載自中大網頁特寫（2019年8月）

兩個或以上的漢字可以組成複合詞，產生不同意思，例如「足」和「球」組成的「足球」、「山」和「城」組成的「山城」。有研究指出，雙字複合詞佔當今漢語逾七成。我們看見「手機」，隨即知道所指何物。就在辨文認字的電光火石之間，我們受到一籃子因素影響，決定我們辨認詞語的快慢。「午飯」和「早飯」兩個詞語筆畫數相若，語意同樣簡單清楚，但我們能夠更快辨認「午飯」；「中餐」和「西餐」，我們更快辨認後者。「爸爸」和「父親」，我們更快辨認前者。本院成員中大教育心理學系謝志成教授就辨認雙字複合詞進行研究，正好解開快慢之謎。

謝教授選取大約二萬五千個詞語，招募五百九十四名母語為粵語的中大學生，把他們分成十八組，每組三十三人，要他們辨認實驗中顯示的是真有其詞，還是非字（例如「禁截」、「寒泣」）。每組在實驗中看到約一千四百個雙字複合詞和一千四百個虛構的非字。研究團隊亦記錄他們辨認詞語所需時間，再計算每個詞語的平均辨認時間，繼而分析辨認時間和詞語特徵（包括非字）之間的關係。

謝教授發現，一些詞語本身的特徵，例如詞語出現的頻率，可以影響我們辨認詞語的反應時間。一個常見的詞語，我們不假思索便能辨認出來。「午飯」二字，平均僅要 0.57 秒辨認；「早飯」就要 0.63 秒，其中一個可能的原因是，「午飯」在香港比「早飯」更為常見。

然而，我們應如何界定一個詞語是「常見」還是「不常見」？謝教授以 Google 的統計作為標準，把搜索範圍設限為香港的網頁，然後輸入有關詞語進行搜索，數算所顯示的網頁數目。包含某複合詞的網頁數目愈多，代表該詞語愈常見。

複合詞二字有前後，我們辨認時會不會被前面的字引導，還是後面的字喧賓奪主？謝教授指着圖上兩個數字，解開謎團：「其實詞語左方的字和右方的字，對我們辨認詞語的影響近乎一樣。」現在的刊物排版不少由左至右閱讀，理論上我們會「先入為主」，詞語左方的字比右方的更影響我們辨認詞語的反應。另一方面，語言學中的「中心語居右原則」指出，部分複合詞最右方的元素主導了詞語的意思和詞類，例如「火車」的「車」顯示火車是交通工具；「白飯」的「飯」指出白飯是一種糧食，理論上右方的字會更影響我們辨認詞語的反應。然而，謝教授的研究證實，我們辨認詞語時，左右平等。

複合詞的語意與其中單字的語意相近，也會讓人更快辨認出來。例如「花園」與「花」和「園」二字皆有關聯，而「花生」卻與「花」和「生」無關，理論上我們能夠更快辨認詞語「花園」。接受測試的中大學生用 0.54 秒認出「花園」，「花生」則要 0.6 秒。

要從洋洋字海中篩選出近二萬五千個複合詞，還要創造數以萬計非字，單是測試前的準備功夫，已經十分艱巨。謝教授苦笑道：「這個研究做了三年，只是篩選複合詞，我們已花了一年，而且都是人手做。」研究團隊從字典把近五萬個複合詞抄下來，經由一班大學生和研究員把一些不合適的詞語（包括幾近絕跡的詞語、專有名詞和鄙言俚語）逐一剔除，始得出近二萬五千個詞語。測試完畢，謝教授的團隊又得逐一數算詞語出現的頻率，並評估複合詞與其中單字的語意相近程度。然而，愚公移山，才造就這項研究的不一樣。有前人做過類似研究，但只是提供二十至一百個詞語作測試。

憑着這個研究，謝教授榮獲 2018-19 年度傑出研究學者獎，但這次並非他首次從事同類研究。早於攻讀博士時，謝教授已埋首研究自動理解文字的情況，即是我們看到文字的一刻，即使受到干擾，仍能即時辨認和理解文字。

謝教授這項研究只是第一步，當日參與研究的學生提供多項資料，包括文憑試中文科的成績、閱讀習慣等，可用於進一步研究。謝教授舉例：「例如對於中文程度較高的人士，『詞語出現的頻率』這個因素會不會影響較小？即使是一些不太常見的複合詞，他們也許能很快辨認出來。我們可以考證這個假設。」另外，若複合詞的語意與其中單字的意思完全無關，學生辨認的表現稍遜，從另一角度看，中小學階段可考慮加強這方面的詞彙訓練。

文字，南宋鄭樵的《通志·六書略》載道：「獨體為文，合體為字」。「山」字始於象形，為文；「城」字由土和成二字形聲合成，為字。山城，就是文字。輕輕二字，卻蘊藏無涯學問。

