

與鄭錦全院士一起來劃清詞語 和詞彙的區別

時間：民國 106 年 7 月 4 日（二）14:00-16:30

地點：國立臺北教育大學 篤行樓 6 樓 603 未來教室

主講人：鄭錦全（中央研究院院士、國立臺灣師範大學華語文教學系暨研究所講座教授）

主持人：陳純音（國立臺灣師範大學英語學系教授）

對談學者：王淳瑩（臺北城市科技大學應用外語系助理教授）

邱力環（國立臺灣大學華語教學碩士學位學程助理教授）

洪嘉翻（國立臺灣師範大學應用華語文學系助理教授）

胡佳音（國立臺灣師範大學英語學系專案助理教授）

張羣（國立臺中科技大學應用中文系助理教授）

許展嘉（逢甲大學外語教學中心助理教授）

黃淑齡（國家教育研究院編譯發展中心博士）

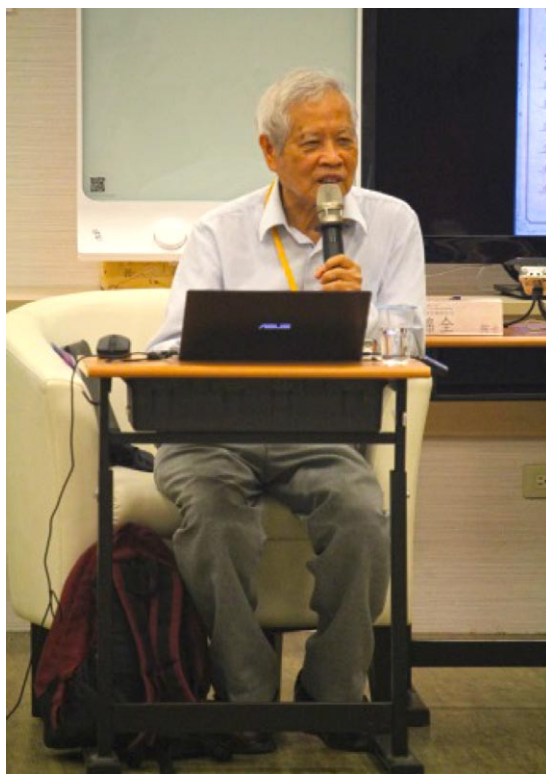
黃雅英（國立臺灣海洋大學華語中心助理教授）

盧欣宜（國立臺北教育大學語文與創作學系助理教授）

記錄：張澄清（科技部人文社會科學研究中心博士後研究員）

鄭錦全院士長期擔任《華語文教學研究》的諮詢委員，並參與審稿工作，發現學界普遍將「詞語」和「詞彙」視為同義詞，造成語言學研究的亂象。因此，鄭院士認為詞語是個別詞，而詞彙是詞語的集合體，若能嘗試劃分詞語及詞彙，或許可以各自開展出不同的研究。

當前詞語研究的主要進展，是將本體論（ontology）應用於語言學研究，但本體論的發展是由生物學的分類而來，這樣的上下層分類難以直接應用在語言學研究，倘若能適當運用此一上下階層及網絡的概念，或有助於分析文學意象。有鑑於特定的詞語在文章中會給予閱讀者某些特定的意象，且多數人認為文學意象只能領會但不可言，也就是只能意會不能言傳，進一步深入探討文學意象為何則會發現，意象是源自人們對詞語所代表的事物所產生的情感與知覺，這些情感與知覺是來自於詞語在文本中出現時，與其共同出現的其他詞



圖一：主講人鄭錦全院士

語。換言之，詞語的意象是透過感性的共現詞來形成，而共現詞是在早期文學作品中經常出現並已形成傳統。這大概就是當我們不瞭解文學意象時，老師會要我們多讀書的原因，透過大量閱讀就會發現特定事物的共現詞，進而發覺詞語的意象。如此，意象便可以言傳，而且可以利用共現詞分析來探討意象。

詞語的語義具有上位詞與下位詞的層次，且詞語所形成的意象同樣也具有上下位的概念，故可依據詞彙語義資料庫為基礎，建構意象辭典或意象資料庫。這類資料庫雖同樣具有上下位概念或上下位詞，但與本體論不同。例如「秋天」的上位詞依序是四季、時間、時間與空間；而「秋」所形成的意象，其上位詞依序可以是蕭條、繁榮與蕭條、境況、特徵等，這是語言學的分析。此外，「秋」也有瀟瀟的意象，是屬於現象與狀態之下，或是收穫的意象，同時把「秋」的意象放在不同條目（entry）之下並加以連結，就可進一步建立關於「秋」的意象網絡。

至於詞彙作為詞語的集合體，則可依據古今漢語文本來探討使用層面的差異。文本是詞語的使用或語言的表達，是一種動態的現象，若依據大量的古今

漢語文本，來研究古今詞彙使用的差異，則是捕捉古今文本的動態差異。因此鄭院士收集了中央研究院的上古漢語（十幾萬個詞語）、現代漢語平衡語料庫前後期（共一千多萬個詞語）、近代漢語標記語料庫（兩百多萬個詞語），以及人民日報、中央通訊社新聞稿、宋詞三百首、唐詩三百首、古漢語的《論語》及《孟子》等。在捨棄人名、地名及年代等詞語後，觀察各類文本出現詞頻最高的前 15 個詞語，計算累積占比，並將其視為詞彙集中度指標。

分析結果顯示，現代漢語平衡語料庫前期詞頻最高的詞為「的」，約占 5.8%；其次為「是」，約占 1.7%，累加前 15 個高頻詞的占比，約為 17%。現代漢語平衡語料庫後期的詞語集中度約為 15%，前後期加起來約為 16%；上古漢語的詞語集中度約為 28%；唐詩三百首的集中度僅約為 3.7%；宋詞三百首約為 3.3%（由於詩詞避免使用重複的詞語，所以詞語集中度較低）；近代漢語的集中度約為 18%；人民日報 13.8%；中央通訊社 12.3%。

把詞語集中度當成重要的指標，觀察詞彙的動態變化，並可藉此區分古代漢語、唐詩、宋詞、現代漢語平衡文本、近代漢語和現代漢語新聞體文本，則詞彙的研究就有自己獨立的方法、觀點和計量表徵。

根據本次研究結果，近代漢語與現代漢語的詞彙集中度很接近，或許可以說明為何在未學習文言文之前，就能夠閱讀《紅樓夢》這類的近代漢語小說。此外，鄭院士又把近代漢語依據不同的書籍進行區分，則詞語的集中度從 13% 到 22% 都有，呈現了相當大的分布範圍，這解釋了近代漢語文本中，因為作者個人的寫作風格，不同比例的使用文言文，造成詞彙集中度的差異。

在鄭院士完成詞語和詞彙的劃分，以及說明詞語和詞彙各自可開展的研究後，現場開放提問。提問內容彙整為以下三個面向：

一、詞語及語義的演變

與會學者提到「泛華語共時語料庫」的相關研究，並針對不同地區之現代詞語的歷時變化，就教於鄭院士。另有學者問及要如何看待新出現的詞語？如何研究？如何教學？鄭院士認為，不同地區民眾針對相同詞語，在語義上有著不同的瞭解和差異，以及某詞語在短時間大量出現，或是慢慢演變，皆屬於語言發展的自然現象。若要進行相關研究，則是發現這個問題之後，要進一步問什麼問題？要用什麼樣的理論架構來解釋？至於面對新詞語的教學，鄭院士則十分贊同與會學者提出的方法，雖然不直接教授這些短時間內大量出現的流行用詞，但是可以考慮教授這些新詞語的同義詞。



圖二：鄭錦全院士(中)和與會者對談互動

二、語言學研究成果在華語文教學之應用

根據文學及語言學的研究成果，是否有具體可行的方法來協助外籍學生進行華語教學？鄭院士認為，一詞多義是華語教學的主要挑戰，鼓勵學生針對特定詞語進行廣泛閱讀，是具體可行的方法。因為相同詞語會出現在不同的地方，而造成語義上的些微差異，透過泛讀就可以捕捉這些差異。或是某些詞語的固定意義，也可以經由泛讀來掌握。此外，語言是很活潑的東西，當語言違反一般的使用情形，有時被理解為幽默。但是跨過了一般的規範太多，則無法被人們理解。

至於，依據高頻詞來進行華語教學是否可行？以及教材編輯是否有考慮高詞頻的問題？鄭院士認為這些是資料科學誤導了語言教學。依據詞頻來決定華語教學的次序固然合理，但語言中詞語的出現有其語境，因此也需要考慮共現詞，而不是只看某詞語頻率。

三、語料庫建構的問題

關於語料庫建構，與會學者提問在進行中文斷詞時，有時會有不一致的問題，應當如何避免或進行答辯？鄭院士舉當前中研院的上古漢語語料庫、現代漢語平衡語料庫、近代漢語標記語料庫等，其斷詞的結果是經過多年來的錘鍊，雖然偶爾會發現一些錯誤，但仍有很高的正確率。斷詞的見解不同可能導

因於依據不同語言學理論系統、學思或是語言學分析方法，甚至是離合詞的問題。

由於語義會有歷時變化，且不同時代有不同的詞彙集，所以有學者提出在編輯意象辭典時，是否要限定在特定時間範圍？或編制不同時間範圍的意象辭典？鄭院士對此則持不同的看法，認為詞語所形成的意象，是有很深厚的文學傳統，不是短期變動的。因為都是依據相同的中國傳統文學或文本，因此對於特定詞語所形成的意象是一致的，可以跨越時代。

此外，透過詞彙集中度可以劃分古今漢語文本，所以有學者提問前 15 個高頻詞之內涵的不同是否也有其意義？若能透過電腦程式及先前所提到的意象網絡，我們是否還需要像傳統的詞彙語義學，針對每個詞語，寫一個詞條，給予一個定義？鄭院士認為，詞語的定義仍是需要的，出現頻率高的詞語有其語言的背景，因此語言的意義和結構的意義是有其意義的，可以進一步深入探討。

最後，關於如何開創前瞻性語言學研究的問題。鄭院士觀察到近年有許多漢語研究，追隨 George Lakoff、Mark Johnson 等人的理念，大談漢語的隱喻，此舉仍然是在別人的理論架構下觀察漢語，不是以漢語使用者的觀點出發，針對我們最熟悉的漢語，進行深入的研究。現在的年輕老師學術生涯十分辛苦，一方面要維持自己的研究，也要負擔很重的教學工作，又要開創自己新的研究領域，在取得學位後持續進行深入研究，建立自己的領域成為一個專家。至於要怎樣深入自己的研究？鄭院士建議不要只想著國際接軌或跟著國際的腳步，這樣永遠只是追尋者，沒辦法超越。

若企圖做出最深刻的研究，提出自己最先進的理論，則不要把其他學者建構的理論，套在中文或漢語研究上，應該嘗試走在研究的前沿，才是真正的研究。國際接軌不是國際領先的途徑，應當要從自己的語言背景，深入地、堅實地提出自己的分析，才是有意義的貢獻。也就是不要做別人做過的，要有開創性，並試圖超越自己。