



語象方法答客問

一，語象觀察是一種什麼方法？

這是香港大學新聞及傳媒中心中國傳媒研究計劃使用的一種實證研究方法。它是傳統話語分析在新條件下的發展，是一種基於互聯網平臺的關鍵詞檢索分析方法。“語象”是比喻。語象如同天象和氣象，可以觀察。我們用這個方法分析社會語言的產生和傳播，從傳媒研究，擴展到政治研究，透過傳媒上的語言，觀察社會演變。經由語象，觀察其背後的動力機制——政治。



二，透過語言觀察政治，可能嗎？

語言反映現實，也能建構現實。語言學家馮·戴伊克曾說：“社會政治權勢和意識可以通過話語來維持、認可和再生”（《話語 心理 社會》頁231，中華書局，北京）。中共建政六十多年來，中國成為口號大國。壟斷話語權的執政者，建立了獨有的話語編碼體系。這套體系有相當大的穩定性，其歷史演變的脈絡清晰可循，透過語象觀察政治是可能的。

2001年，筆者曾檢索研究《解放軍報》，論證文化大革命從醞釀到爆發，恰是“階級鬥爭”等紅色口號的勃興過程；2003年，筆者應香港大學新聞及傳媒研究中心總監陳婉瑩教授的邀請，到港大從事中國傳媒研究。筆者檢索研究了“輿論導向”、“輿論監督”、“新聞改革”和“新聞自由”四個關鍵詞的此消彼長，指出中國新聞改革依然受到嚴格控制，爭取新聞自由的道路艱難坎坷；2006年是唐山地震三十周年，筆者對紀念唐山地震的報導做了話語分析，發現中國傳媒正在出現“靜悄悄的裂變”，新聞界奮力突破管制，呼喚人性和公民權利的民間話語在生長。

此後，筆者聚焦政治改革。在中共十七大、十八大前後，筆者用語象方法分析政治關鍵詞的表現，對時局的發展作出判斷，事實證明基本正確。

三，語象方法是怎麼產生的？

傳媒話語分析是西方學者開創的。他們開始從事新聞話語研究的時候，尚無互聯網，甚至還沒有個人電腦。1992年，國內已有人在用電腦分析整本《紅樓夢》的語言特色。當時中國已開始研究“機器翻譯”課題，和“信息檢索”技術。在香港中文大學，著名學者金觀濤和劉青峰夫婦，建立了中國近現代思想史研究數據庫，首創使用電腦檢索系統的“計量歷史學”方法。

這一時期美國出現了搜尋引擎。1990年出現了搜尋引擎的鼻祖 Archie，此後，WebCrawler、Lycos、Infoseek、Alta Vista、Inktomi、Northernlight……，一個個互聯網搜尋引擎連翩而至，直到 Google、Baidu 風靡世界和中國。數碼技術帶來了海量搜索和模式識別，語象方法正是這一技術革命的產物。

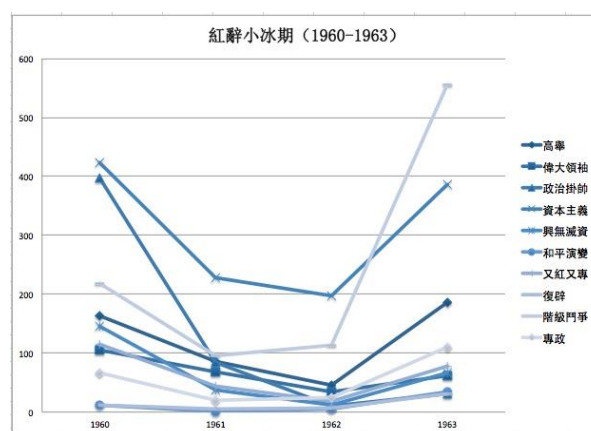
四，“語象”方法和傳統話語分析方法有何異同？

在電腦與互聯網降生之前，學者使用社會科學的各種方法，對包括媒體話語在內的社會語言作過深入研究，積累了豐富的知識。數碼科技的崛起，大數據時代的到來，使研究越上了新的平臺。

新聞媒介的語言資料保存完整，易於檢索計量。我們現在可以方便地對一份報紙的全部歷史資料進行檢索，也可以方便地對幾百種報紙同時進行檢索。不僅可以在瞬間搜索到所需要的資訊，而且可以限定時間、空間（報紙、網站等資訊提供者），設定各種邏輯關係，有針對性地搜索和精確計量；不僅可以搜索、計量文字資訊，而且可以搜索、計量圖片、聲音和視頻信息；不僅可以搜索、計量傳統意義的“大內容”（Macro-contents），而且可以搜索、計量隨著 Web 2.0 而興盛的“微內容”（Micro-contents）。

今天我們所能擁有的語料庫，其規模之巨大前人無法想像，數碼科技亦提供了更多分析手段，它讓我們用數據支持、修正或否定印象性觀察、經驗性認知和邏輯推斷，讓我們透過數據，在社會與政治的“顯事實”背後發現“隱事實”。

例如，筆者曾使用《解放軍報》檢索系統，發現在“大躍進”、“大飢荒”時期，中共一系列紅色口號在1960年達到高峰，1961年至1962年上半年卻全線下滑，到1963年再度強勁攀高。這一U字形，反映毛澤東一度退縮，到1962年秋八屆十全會後，又強勁反擊。



用一系列數據支持的這個“隱事實”，在沒有電腦和互聯網的時代是無法窺見的。但這個“隱事實”，電腦不會直接告訴你，它是具有歷史和新聞知識的研究者判讀數據所得。搜尋引擎尚在發展。目前條件下，史學、語言學、傳播學、政治學，仍是不可或缺的知識基礎，而數碼化檢索和計量是一種輔助手段。

這種手段也擴展了以往傳媒研究的內容分析方法。2003年，SARS疫情爆發。中國媒體曾在短暫時間內對SARS進行過大量報道，但很快就被宣傳部壓制。在公眾需要公共衛生資訊的緊要關頭，中國媒體集體噤聲。當時筆者剛到香港大學，我們使用“慧科搜索”數據庫搜索數百種中國內地報紙，畫出了SARS 報道大起大落的準確軌跡圖。幾乎是同樣的一幅圖，5年後再次出現。這次是汶川地震中校舍倒塌的報道。又是先大量報道，後突然被禁。

“語象”方法，融合了定量（量化）研究和定性（質化）研究兩種方法。詞頻數據不難獲取，憑藉這些數據展開分析得到結論，則需要傳統學術研究的功力。

五，你們使用哪些數據庫？如何對傳播進行計量？

首先是一批報紙的圖文檢索系統，如《人民日報》、《解放軍報》，它們提供了這些老報紙自創刊以來的所有文字和版面。其二是收錄了多種媒體的綜合性數據庫，如香港的“慧科搜索”，它集納了近十多年來的數百種報紙，以及一些網站的內容。其三是網絡搜尋引擎，如穀歌、百度，這些搜索器提供的高級檢索功能，可以限定時段、數據來源，可對語詞作搭配檢索。

計量的基本元素是詞頻。在上述數據庫進行關鍵詞檢索，都能獲得詞頻數據。所謂詞頻，可以表現為“全文使用篇數”，也可以表現為“標題使用篇數”。對單一文本（一篇文章，或以一個檔形式出現的一組文章乃至一本書），可以使用Word的搜索功能，統計某一關鍵詞的絕對頻率。詞頻極其寶貴，如同血壓、體溫、氣溫、雨量、水位、股價等不同的測量數據。傳媒大數據是研究者的富礦。

六，如何選擇、觀察關鍵字？

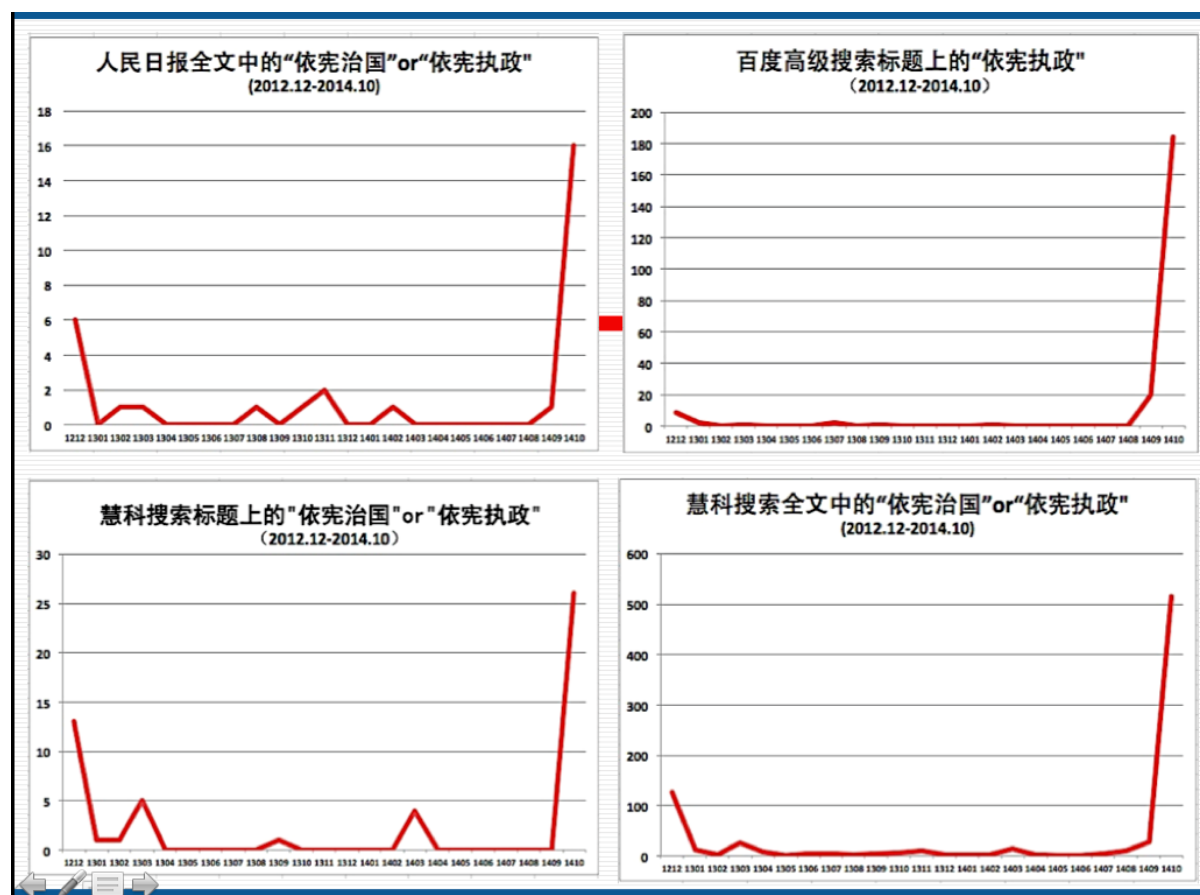
在一項研究中，選擇什麼樣的關鍵詞做數碼檢析，實有賴研究者對媒體和文獻的長期觀察和閱讀經驗。例如，筆者自從上世紀八十年代以來一直關注中國的政治體制改革。“政治體制改革”顯然是值得觀察的關鍵詞語，同時，筆者在2007年便提出，還須關注若干“政治體制改革”的同義語。在中共體制內，它們是：“政治文明”、“民主政治”、“黨內民主”；在民間，它們是：“憲政”、“公民權利”和“公民社會”。通過數碼檢析，筆者持續觀察這兩組詞語的交織與衝突，獲得寶貴的認識。正是這種較為清晰的認識，當2013年“七不講”傳來，筆者即意識到中國大陸“話語戰爭”已經開始，此後，發表多文，分析政治改革正在倒退。

對於關鍵詞，我們觀察“語源”：一個關鍵詞是何時出生的？在什麼背景下出現？我們觀察“語流”：它的傳播，在流量上發生什麼樣的變化？我們觀察“語變”：它的衰落，它的消亡。我們觀察：潛藏巨大能量的口號，是怎麼降臨的？在語言的“華爾街”，不同語詞走出了什麼樣的升降曲線？傳媒如何參與製造了話語的波瀾起伏——又被這種起伏所裹挾漂流？我們更注重觀察“語境”。任何語象都在特定語境下出現，語境不明，語象無解。傳統的研究方法，和基於互聯網平臺的檢索手段結合，使我們可以勾勒出更宏觀也更準確的圖像：傳媒在說什麼？什麼人在說？不同人（不同傳媒）所說的有何異同？在不同的時間，所說的又有什麼不一樣？從紛繁的語言現象中，經比較，看出語言“棋盤”裏各個“棋子”的關係，看出錯綜的關係如何位移，識別偌大棋局的總體態勢。比較，是語象觀察的基本方法。

語象背後的“隱事實”，須通過比較發現。任何一個關鍵詞（語）的單一詞頻數據，若不進入比較，則無意義，無信息。所謂比較，分縱向比較與橫向比較。縱向比較，是詞語傳播的歷史觀察。如“政治體制改革”，可以檢索統計自1980年至今在《人民日報》上每年的使用篇數，由一組數據構成傳播量變動曲線。縱向比較也可作更細微觀察。例如在2003年春非典型肺炎爆發期間，筆者對“非典”一詞在報紙上的傳播作過逐日統計，其變動曲線，顯示了當局對疫情報道從嚴控到放開的過程。橫向比較，可以是一個詞與另一個詞的比較，可以是一組詞和另一組詞的比較（包括傳播總量和歷史曲線），也可以是相同詞語在不同媒體上的傳播實況的比較。

七，如何通過語象方法提升研究水準？

此前，研究課題（聚焦點）大多來自研究者的經驗性判斷。如今，通過使用數碼技術的語象方法，研究者有可能發現更多新的探究點。例如，中共十八大後，筆者一直關注習近平提出的各種口號，其中，對“依憲治國”、“依憲執政”在黨報上的出現頻率作了持續觀察。2013年，發現這兩個口號條起忽落，2014年，《習近平總書記系列重要講話讀本》出版。筆者對這本發行逾千萬的超級大書進行了語象分析，發現上述二詞竟被刪去，即撰文《“依憲治國”為什麼不見了》，指出這是批判憲政所致。不久，在中共十八屆四中全會，這兩個詞語“復活”，筆者遂發表《過山車上的“依憲治國”》，作進一步剖析。可以說，這個研究題目，完全得自語象觀察。語象的“異動”，往往正是研究的題目。基於語象觀察的研究，必須聚焦異動，解釋異動，從異動，追蹤政治變遷。



前面談及，詞頻數據不難獲取，分析研究則需學術功力。但筆者也在嘗試，讓數碼技術在分析研究中扮演更重要角色，例如布爾邏輯的運用。搜尋引擎問世之初，就有布爾邏輯功能（即“邏輯是”、“邏輯非”、“邏輯與”、“邏輯或”）。筆者在做“政治體制改革”相關詞語研究時，將體制內話語“政治文明”、“黨內民主”、“民主政治”編為A組，將民間話語“憲政”、“公民權利”、“公民社會”編為B組，將兩者融合部分編為C組，分別檢索“有A無B”、“有B無A”、“有A有B”三種結果。統計發現，在中國傳媒的傳播總量上，A大於B；而AB同時出現的情況稀少。數據證明：關於政改，存在十分明顯的話語分野。筆者還使用布爾邏輯方法，分析過語義和語族的變遷。

八，語象方法有多大適用性？有何局限？

我們從事語象觀察的實踐表明，以此方法分析華文媒體，如中國大陸媒體，特別是中共黨媒，較為有效。中共的政治口號，稱為“提法”，是權力話語，它的提出和變動，均受控於權力，有較強規範性，這對研究是一種便利。

但是中共的話語，有“命脈語”、“穴位語”、“旗幟語”、“皮毛語”、“脂粉語”之分別。如“四項基本原則”這樣的命脈語、“依憲治國”這樣的穴位語、“毛澤東思想”這樣的旗幟語，須長期觀測。而幾十年如一日的高調，如“人民”、“大好形勢”等皮毛語脂粉語，既難檢析，對研究也無大價值。

通過數碼技術分析民間語言，也是可能的，主要數據在市場化媒體和網站。但對互聯網的檢索，須注意到數據庫的不穩定。目前，語象方法適用於對百度新聞搜索（高級）和主要門戶網站進行較短時段的檢索。對社交媒體的關鍵詞檢析，筆者做過嘗試（如對新浪微博），然而情形較為複雜，尚在探索。

九，語象方法的發展前景如何？

未來數年，數碼技術在儲存容量、傳輸速度、搜索能力等方面將發生革命性飛躍。換言之，我們有可能在更大範圍、更高效地進行新聞話語檢索計量；借助電腦和網絡，分析水準也會大大提高，甚或發生飛躍，有理由樂觀期待。

錢鋼

香港大學新聞及傳媒研究中心 中國傳媒研究計劃主任

7.2015