

# 用方法說故事：探析電腦輔助文本分析工具 在框架研究之應用\*

盧安邦\*\*

國立政治大學傳播學院博士候選人

鄭宇君

玄奘大學大眾傳播學系副教授

## 摘要

人文社會領域「向運算轉」後，「讓資料說話」成為一股新的風潮，然而資料如何能說故事？本文以電腦輔助處理大量社群媒體文本資料為例，說明在資料蒐集、處理及分析過程中，須經各種文本分析工具的內在邏輯及研究者介入，資料的意義方得浮現。本文使用詞頻分析、詞語共現、情緒分析工具及質性框架分析法等文本分析工具，對同一組社群媒體文本資料集進行研究，結果發現個別分析工具確實為研究者帶來不同洞見，卻也各自存在侷限。

因而，本文認為電腦輔助的文本分析方法在資料篩選、分析、詮釋的過程中，皆需研究者介入進行調整，結合多重方法的研究設計則可協助研究者從文本資料中找出意義框架，說出一則好故事。

**關鍵詞：**社群運算、詞頻分析、詞語共現、情緒分析、框架、議題設定

\* 本文為科技部專題研究計畫《探索社交媒體之民意形成過程》(MOST 103-2410-H-364 -004 -MY2) 之部分研究成果，作者感謝國立政治大學水火計畫研究團隊成員對本文的建議，並感謝審查委員對本文給予的寶貴意見。

\*\*Email: donkeydumb326@gmail.com

投稿日期：2015 年 7 月 11 日

接受日期：2016 年 1 月 27 日

## 壹、前言

現在，該讓巨量資料說話了。（Mayer-Schönberger & Cukier, 2013  
／林俊宏譯，2013：6）

近年來電腦運算工具的興起使巨量資料研究逐漸成為顯學，一時之間「讓資料說話」（Mayer-Schönberger & Cukier, 2013／林俊宏譯，2013）似乎成為新的學術時尚。然而，甚麼叫做讓資料說話？資料怎麼說話？「巨量資料」的威力下又隱含了甚麼問題？本文認為在巨量資料的風潮下，這些重要問題尚未獲得足夠的關注。實際上，「讓資料說話」仰賴電腦運算工具的介入，然電腦運算工具的使用與操作，又與研究者的主觀介入息息相關，本文試圖由此出發，討論人文社會領域的巨量資料研究中，資料、運算工具與研究者之間的複雜關係。

利用電腦運算工具進行的巨量資料分析可分為許多類型，一般而言，我們可以將社群媒體資料分為數量資料、文字資料以及後設資料，數量資料包括回應數、瀏覽量、分享量等，而文字資料則為發文內容、評論內容、標籤內容等；後設資料則指文章中引用的連結、地理資訊（常以座標表現）、使用者 IP 等，這些資料常須透過進一步資訊處理（如將座標化為地理位置、辨識引用連結網站的類型或區域等）方能使用。處理不同資料仰賴不同分析工具，操作邏輯也各異（Adedoyin-Olowe, Gaber, & Stahl, 2014），難以簡單一概而論。為助聚焦，本文因此將視野鎖定在文本資料的處理，在議題設定到框架研究這條研究軸線中討論電腦運算工具在甚麼樣的脈絡下進入此研究領域、具有甚麼特性、造成甚麼影響、隱含了甚麼問題。

議題設定是傳播研究長久以來的重要概念，主要用來瞭解媒體設定的議題與公眾心中的議題是否相關，而媒體議題又是由新聞報導的框架所構成，分析新聞框架不免涉及到新聞報導的文本內容。然而，電腦輔助的文本分析，主要處理的資料單位為「字」（words），在中文語境裡，電腦分析的文本最小單位則是由一個到多個字組成的有意義「詞彙」。因此，本文探討的主題是從電腦運算工具輔助的詞彙分析到框架分析，進而形成議題的過程中，可能衍生的相關問題。

議題設定源起於 1968 年，McCombs 與 Shaw 針對在 Chapel Hill 的美國總統大選，以一百名尚未決定投票意向的選民為樣本進行研究，試圖回答兩個問題：一、選舉期間以來，什麼是你最關心的議題？二、不考慮政治人物的說法，試舉

出二到三件你認為政府應該著力之事？在這篇經驗研究中，McCombs 與 Shaw 發現媒體對特定議題報導的「數量多寡」及「版面大小」會對公眾認知識題的重要性產生影響，這就是它們提出的「議題設定」（agenda-setting）理論，也因為這個研究，帶動後續研究者對媒體議題的持續關注（McCombs & Shaw, 1972）。

這條研究取徑歷經數十年持續驗證與發展，一方面逐漸將焦點從研究媒體報導議題拓展至閱聽人心理以及民眾的日常討論公眾議題（public agenda），另一方面則從「議題」拓展至「議題框架」，與 Goffman 提出之框架概念結合，開啟了媒體報導框架研究之路（McCombs, 2005）。

時至今日，線上新聞媒體與社群網站的發達一方面讓媒體議題設定可以快速在網路上獲知公眾反應，迅速調整議題設定方向以迎合公眾喜好，使得媒體議題與公眾議題的相互影響程度更甚於過往；另一方面則因媒體議題與公眾討論大量出現於社群媒體，使得研究者得以更為方便地蒐集公眾圍繞特定議題的相關討論，從中分析公眾意見。過去如 McCombs & Shaw (1972) 透過問卷調查，藉由受訪者回溯性的自我報告捕捉公眾意見，Entman & Rojecki (1993) 以質性框架分析法分析新聞報導，研究者如今得以直接面對人們的發言內容，甚至利用資訊技術工具分析大量資料，從中找出媒體議題以及公眾議題，甚至捕捉議題發展的趨向。相關應用如以詞頻分析研究不同媒體間關注政治議題之差異（Burgess & Bruns, 2012; Neuman, Guggenheim, Jang, & Bae, 2014），或研究人們如何談論特定主題、人物與概念（Graves, McDonald, & Goggins, 2014; Groshek & Al-Rawi, 2013; Yuan, Feng, & Danowski, 2013）。

近年越來越多研究嘗試使用電腦運算工具進行大量內容的文本分析，同時在公私部門也出現許多資料分析工具，幫助企業、政治人物、研究者找到意見聲量、情緒、詞語共現關係等。此趨勢一方面呼應了「讓資料說話」風潮，另一方面卻也讓研究者更為好奇這些資料如何能夠說話？換句話說，散布在各處充滿雜訊的「資料」究竟如何化成有意義的「資訊」，甚至藉由圖表呈現出一個值得關注的「故事」？

本文認為，從資料、資訊到故事的過程中，「運算工具的處理邏輯」以及「研究者的介入」扮演關鍵角色，不同分析工具有其內在邏輯與特性，對分析結果常有影響；而資料事實上不會自己說話，藉由研究者以及研究工具的中介，資料才得化為資訊以及故事。那麼，如何才能看見「運算工具的處理邏輯」以及「研究者的介入」在研究中扮演的角色呢？運算工具的處理邏輯必須在不同工具的比較中方能突顯，而研究者在分析過程中的角色則常隱藏在實際操作的細節中。由此

出發，本文的重點有二：一、藉由應用不同方法分析相同資料上，比較分析結果差異來說明不同工具／方法之特性以及處理邏輯；二、反省利用電腦輔助分析的巨量資料方法操作細節中隱而未顯的研究者主觀性。

在此主旨上，本文首先回顧議題設定以及框架分析之理論以及經驗研究文獻，藉此說明電腦運算工具出現的背景。接著，本文援用常用的電腦輔助的文本分析工具與方法針對同一批資料進行分析，探索工具處理邏輯以及研究者在操作過程中扮演的角色，討論資訊技術的介入如何可能（或不能）協助研究者在大量內容中回答其研究問題。最後，本文試圖探討研究者如何藉而不同研究方法的融合，分析大量新聞內容中浮現的媒體議題以及社群媒體文本中浮現的公眾議題；兼而思考未來議題設定研究的發展方向。

## 貳、從議題設定到框架研究

1968年 McCombs & Shaw 提出的議題設定理論告訴我們媒體藉由突顯某個議題，能夠告訴閱聽人甚麼是重要的，因此可以指示人們應該要想些甚麼（what to think about）；經過了眾多後續研究後，McCombs (2005) 指出其實媒體能夠帶給閱聽人的影響不只是能告訴人們應該想甚麼，還能夠影響人們對議題的看法（what to think）。因為媒體再現一個議題時，時常會與其他元素一起出現，這些元素成為了該議題的屬性，進一步影響閱聽人對議題的理解。舉例而言，日前新聞媒體大量報導復興客機空難，告訴人們該事件值得重視，然在報告空難新聞時，內容可能伴隨罹難者家屬的畫面、國際對該事件的評價或是復興航空曾經發生過的意外事件，這些內容都會成為議題的屬性，影響人們對此事件的理解。

談到議題設定研究，人們往往會聯想到框架研究，率先結合議題設定與框架研究的乃是 Entman，他的做法帶動後來許多相關研究。Entman 曾舉例說明，當一個問卷中愛滋病檢測與公民自由權連結在一起時，大多數人會支持愛滋病患者的人權；但當此議題被框架以強調公共健康時，則大多數人會支持愛滋病強制檢測 (Sniderman, Brody, & Tetlock, 1991, 轉引自 Entman, 1993)。從這個例子中，可明顯看出文本如何藉由賦予議題不同屬性的框架而影響人們對議題之理解與態度。

事實上，框架 (frame) 概念乃由社會學家 Goffman 所提出，他指出框架乃是一種人們的詮釋基模 (schemata)，以現存文化信仰與世界觀為基礎幫助個體與他人進行溝通協商。將框架運用在解釋新聞論述中者則是 Gamson，他以此概

念研究框架如何被組織起來並傳遞給閱聽人 (Nisbet, 2009)。Nisbet (同上引) 指出, 當媒體再現特定事件時會有其訊息框架, 閱聽人暴露在媒體傳送的訊息框架下, 在心中連接起他們對於事物的詮釋框架, 自然會結合兩種概念、論述, 如此一來, 便達到訊息建構者的目的。

Entman (1993: 52) 認為「框架本質上牽涉選擇與突顯, 去框架某事便是在感知到的真實中選擇某些面向, 並讓他在傳播文本中更加突顯出來, 藉此促成某種特定的問題定義、因果詮釋、道德評價以及/或者解決方案建議。」他將框架概念從個體身上擴展到社會, 認為框架會在四個位置出現: 傳播者、文本、接收者、文化。但無論在哪個位置, 都包含了選擇與突顯, 也就是突顯特定元素來建構一套關於問題為何、因果關係、道德評估並提供解方的說法。

所謂突顯 (salience) 指的是「讓資訊的某部分特別引人注意、有意義或者容易被閱聽人記住」(Entman, 1993: 53)。但 Entman 也指出, 「突顯 (salience) 是文本和接收者互動後的產品」(同上引), 也就是說, 框架 (frame) 與基模 (schemata) 的對應方能帶來效果, 因此研究者即使在文本中發現框架, 也不必然保證那樣的框架會影響閱聽人對事件的想法,

文本雖然能夠藉由配置 (placement)、重複 (repetition) 或者與其他熟悉的文化符號連結 (association) 來突顯某些資訊, 然而, 即使一個描述出現在文本中不被強調的位置, 但只要他與接收者信仰系統中的既有基模 (schemata) 一致, 這個訊息仍會獲得突顯, 反之亦然。  
(同上引: 53)

那麼, 為什麼對媒體框架進行研究是重要的呢? 前述 McCombs (2005) 已然指出議題框架會影響人們對議題的認識與態度, Entman & Rojecki (1993) 則進一步認為媒體對社會運動的反應會影響事件擴散的程度、影響大眾對社會運動的參與意願, 並影響菁英決定是否需要出面支持運動。這就是說, 媒體框架能夠影響人們對事件的理解、態度, 進而影響人們的行動, 並對事件後續發展產生效應, 由此可見, 媒體框架對社會之影響不可謂之不大。

另一方面, McCombs (2005) 指出, 新聞媒體的議題設定角色, 也逐漸從關注「媒體議題」的顯著性轉換到「公眾議題」的顯著性。因為雖然媒體能夠設定公眾的討論議題, 但人們總有他們自己的理由參與或不參與媒體議題, 並可能因



為與事件的相關性、事件的不確定性、個人涉入程度與取得訊息付出的努力而對閱聽人有著不同影響。

隨著網路與社群媒體的蓬勃發展，公眾議題也成為可以觀察、研究的文本，於是越來越多研究者對大眾媒體與社群媒體之間的議題與框架互動產生興趣，並由此展開系列研究。

## 參、眾聲喧嘩的研究方法：從工人智慧到人工智慧

同樣隨著科技與技術的發展，更有趣的進展在於研究者所使用的研究工具，Entman (1993) 所提出的框架研究方法為質性分析，仰賴研究者對文本之理解與判斷，但網路的普及大大降低了媒體的成立門檻，同時使得人人都能藉由網路發言與討論。「媒體議題」與「公眾議題」的文本數量迅速增加，一方面提供研究者大量有價值、可供分析的文本，另一方面卻也為傳統質性框架分析方法帶來困難，研究者這時面臨選擇：要縮小研究範圍或用抽樣的方式減少納入分析之文本，或者使用資訊科學分析工具，以電腦取代「工人智慧」來幫助處理大量資料？

### 一、傳統框架分析方法介紹：從論述分析到量化內容分析

Entman (1993: 53) 提出的框架概念指出媒體文本會「藉由配置 (placement)、重複 (repetition) 或者與其他熟悉的文化符號連結 (association) 來突顯某些資訊」，因此框架分析不只是文字的表層意義分析，更牽涉了概念與概念之間的關係，甚至是概念與當地文化脈絡之間的關聯。因此，Entman 的框架分析乃是試圖深入文本意義的質性分析方法。

舉例而言，Entman & Rojecki (1993) 利用 NEXIS Database 蒐集了 1980 ~ 1983 年《紐約時報》跟《時報雜誌》所有與凍結核武運動相關之報導文章，同時使用質化與量化方法來研究報導之主題與框架，並計算不同人物在報導中出現的次數。此研究分為兩部分，一為計算報導中引用之消息來源背景，這偏向消息來源研究；另則為分析報導內容之框架。在分析報導框架時，Entman 深入報導內文，指出時報雜誌某篇報導藉由將事件描述為「情緒驅動的」、「大眾奇觀」以強調運動缺乏力量、藉由將遊行報導為「參與人數不及警方預估數量之一半」以及「紐約有史以來最大遊行」而非「全美有史以來最大遊行」，以弱化運動獲得的公眾支持。由此可見，Entman 的框架分析試圖深入文本的深層意義，分析

時並且仰賴研究者對文化脈絡之理解與敏感度。

李紅豔、劉碧瑤 (2014) 研究媒體對「上海普陀事件」與「廣州增城事件」的報導時也使用類似方法，她們蒐集《新民網》、《中青網》、《大洋網》、《南方網》、《東方早報網》對兩事件的相關報導共 14 則，分析事件中被打小販、編制外治安人員、滋事群眾、目擊者、政府、警方等涉事各方及其行為和態度在報導中如何被描述。此外，Anstead & O'Loughlin (2014) 以質性主題分析法 (thematic analysis) 分析提及 2010 年普選期間社群媒體的媒體內容 (287 則)，或也可視為此類。

Song (2007) 則研究韓國屬於保守派的《朝鮮日報》(Chosun)、南韓《中央日報》(JoongAng)，激進派的《民族日報》(Hankyoreh)，和兩個激進派網路新聞 Ohmynews 公民新聞網與 PRESSian 對美軍裝甲車壓死女學生事件之報導共 1,063 篇，分析不同報紙對事件之框架差異。這篇文章的框架分類方式採用由 Glaser and Strauss 提出的「持續比較法」(constant comparison) 質化分析方式，將新聞框架分成抗爭、韓政府、美國、SOFA、意外調查、反美情緒等六個主題框架，並分析主題框架如何被媒體描繪。他發現在網路媒體和激進的報紙中「抗爭」框架是第一名，主要內容為反對美方的審判結果，並傾向將社會抗爭視為一種「人民的力量」。

除了如上述直接進入文本，從文本中逐漸歸納出框架的分析取向外，另一類研究方法為先參考文獻、理論或是藉由先驅研究 (pilot study) 建立類目後再將文本歸入既有類目。

以理論作為基本類目者如陳韻如 (2011) 分析新聞媒體報導婦女運動公共議題的框架及其背後的文化意涵時，選取了祕魯婦女運動議題的報導以及評論文章共 151 篇進行質性框架分析，以 Entman (1993) 提出的架構為基礎，探究與議題相關之框架中之問題定義、問題來源、責任追究、道德評估與解決方案，並分析新聞利用甚麼樣的論據、隱喻或是刻板印象來包裹議題。

Zhou & Moy (2007) 研究被稱為「BMW」事件的社會爭議事件中，中國最大的 BBS 論壇「強國論壇」(206 篇) 和新聞媒體報導 (114 篇) 間報導／討論框架之互動關係時，也以 Entman 提出的「定義問題」、「診斷原因」、「道德評判」以及「建議解決方案」框架為基本架構，分析文章內容透過甚麼樣的明確框架來達成此功能。

Entman 提出的是框架研究之架構，然而，研究者也常根據自身研究問題而

自行設計其所欲聚焦的框架類型。如陳憶寧（2002）探討民進黨執政後十個月內的核四新聞框架時，則根據其研究問題先將框架分為「遊戲框架」與「議題框架」，進而檢視遊戲框架與議題發展過程的關聯。

謝君蔚、徐美苓（2011）在研究媒體再現基因改造食品的新聞框架時，則結合歸納法與演繹法，先從既有文獻中建立分析生物科技新聞所常見的框架，再用內容分析將所蒐集到的 393 則新聞資料歸入既定類目。

前述提及的文獻無論是先以理論或是既有文獻建立類目，在對文本進行分析時皆仰賴研究者判斷文本之深層意義，而非僅由文字表層進行分析。然而，此類分析方法所遭遇之困難主要有二，一為信度問題，另一則為可處理之資料量有限。因為越要分析文本之深層意涵、越要考量文本之社會文化脈絡相關，則越仰賴編碼者的領域知識與其敏感度、編碼者間信度越難達標、能夠處理的資料量也就越有限。

因此，一般而言若想處理大量資料同時提高分析信度，便需在分析方法上有所取捨，放棄對深層意義的詮釋，而朝量化內容分析靠攏。如 Himelboim, Sweetser, Tinkham, Cameron, Danelo, & West (2014) 在分析 Twitter 使用者關於賓拉登、經濟、能源、同性婚姻、健康、移民、醫療照護、石油天然氣、女性、墮胎等十個議題的討論內容時，在所有蒐集到的資料中隨機抽樣 15%，共 7,090 則作為內容分析文本，並使用 Berelson-Type 量化內容分析法，在編碼的過程中，由訓練過的六名編碼員組成的編碼團隊針對外顯的內容（manifest content）進行編碼，而不去討論深層意義。

然而，即使是採用量化內容分析，在面對社群媒體帶來的大量資料時仍需藉由抽樣將資料量縮小到人工可操作的程度。巨量資料方法則改變了如此狀況，為框架分析帶來了嶄新發展。

## 二、巨量資料（Big Data）帶來的改變

Mahrt & Scharrow (2013) 指出巨量資料泛指數量訊息龐大到無法以一般儲存空間或處理設備處理的大規模資料。boyd & Crawford (2012) 認為巨量資料改變了知識的定義，因應大量數據的出現，各種方法、工具和研究過程也相應發展，對研究者的思維與研究設計皆產生很大影響。其中重要影響之一，便在於因為資訊技術的發展，研究者分析大量資料時不再需要對蒐集到的資料進行抽樣，而能



將所有能夠蒐集到的資料皆納入分析（鄭宇君，2014）<sup>1</sup>。

巨量資料方法至少可以分為兩個層面：資料蒐集以及資料分析。當研究者要分析網路上大量資料時，首先會碰到資料蒐集的問題，這時常見的方法為設計網路爬蟲（web crawler）工具自動抓取網頁上的資訊；或者利用網路服務的應用程式介面（application programming interface，簡稱 API）下載網站的資料及後設資料（meta data）。當研究者將網路上大量資料抓下來之後，完成資料蒐集之後，實際上另一個令研究者頭大的問題才真正浮現：面對眼前龐大的資料，研究者應該從何下手進行分析？

一般而言，常見的文本分析工具為關鍵詞分析（keyword analysis）、詞頻分析（frequency analysis）、詞語共現分析（co-occurrence analysis）以及情緒分析（sentiment analysis）等（陳百齡、鄭宇君，2014；Liu, 2012; Stubbs, 2001）。其中關鍵詞分析試圖找出網路使用者針對某個議題所較常使用的詞彙；詞頻分析則可看出特定議題相關討論中所有詞彙之出現頻率；詞語共現分析可找出研究者關心之詞彙與其他詞彙之間的關係；情緒分析則可幫助研究者看出人們談論特定議題時之態度為何。

舉例而言，Neuman et al. (2014) 研究不同媒體間關注議題之差異時便使用了詞頻分析，他們首先從經濟與福利、政府、國外事務、公共秩序、社會議題、環境等六類議題中選出 29 個政治議題，並針對每個議題設定 4 到 9 個關鍵字，接著蒐集 Twitter、部落格、網路論壇以及傳統新聞印刷和廣播媒體之報導內容，研究那些關鍵字在不同媒體中的出現頻率。Burgess & Bruns (2012) 則透過觀察 2010 澳洲總理選舉事件研究傳統主流媒體與 Twitter 內容的關係，他們首先蒐集含有 #ausvotes 的推文，找出討論中最常使用的字詞片語建立詞庫，依據競選活動的主題進行分類，再透過詞庫進行主題分析（thematic analysis）。建立詞庫與主題後，Burgess & Bruns 則進一步比對五個議題在 Twitter 上的討論次數與官方報導（主流媒體和政黨）的消長情況，發現雖然 Twitter 的討論大致可反映兩大黨與主流媒體的競選活動內容，但兩者對於不同議題的重要性認知和討論熱度卻有很大差異。

使用詞語共現分析者則如 Yuan et al. (2013) 分析新浪微博上與「隱私」相關

---

1 事實上，雖然電腦分析工具擴大了可處理的資料量，但在大數據研究中抽樣的問題仍舊存在，從確認問題、選擇案例、蒐集資料、清洗資料、分析資料到詮釋資料，皆可見研究者依其理論視野或是領域知識介入，從眾多的可能性中挑選出其認為最適合的選項。從後文操作過程中可見研究者在不同階段扮演的角色，並將於文末「研究者角色」相關討論中進一步處理。

之討論內容。他們搜集了新浪微博上 18,000 則包含了「隱私」關鍵字的微博，並在斷詞後找出與「隱私」一詞共同出現在三個詞間距（three-word window）以內詞彙，最後根據共現頻率繪製語義網路圖。

情感分析則可視為詞語共現分析的延伸應用，如 Groshek et al. (2013) 希望瞭解人們在社群媒體上描繪候選人的用字遣詞，因此分析 2012 年 1 月 1 日至 11 月 7 日投票日後一天 Barack Obama 與 Mitt Romney 在其臉書粉絲專頁以及 Twitter 上的貼文與回應共 1,427,207 則中常共同出現的字詞，藉此找出常與批評性字眼一起出現的詞彙為何。

也就是說，在研究者確認了其研究問題後，他便面臨研究方法與工具的選擇，若研究者對媒體議題或公眾議題框架有興趣，那他該選擇質化的框架分析、量化內容分析、或是利用電腦運算工具幫助分析？若面臨大量資料因而須要電腦運算工具幫助處理，那該使用詞頻、詞語共現或是情緒分析工具？不同工具特性為何？資料處理的邏輯為何？實際操作時又會碰到哪些困難？

總體而言，議題設定與框架研究歷經數十年，使用傳統研究方法者已累積相當文獻可供後續研究者參考，然而，使用電腦運算工具進行文本研究則為新興趨勢，文獻數量相對較少，更少文獻探討不同的新興分析工具在進行框架研究上之優劣何在，以及從詞彙分析到框架分析的聚合過程。有鑑於此，本文將使用詞頻分析、詞語共現分析、情緒分析以及傳統框架分析等不同方法對同樣一組資料進行分析，一方面討論不同方法之特性與其所帶來的洞見，另一方面則思考採用這幾種不同方法進行框架研究之可能架構。

## 肆、研究方法與資料描述

反核運動從 1980 年代至今，已歷時 30 年之久，而反核運動從早期由反核團體與民進黨合作，到政黨輪替後民進黨路線轉保守，反核運動因而逐漸轉向大眾。從 2005 年綠盟舉辦的反核紀錄片巡迴，2009 年及 2010 年綠盟在貢寮舉辦諾努客音樂會、影展與在地農民市集，皆吸引了許多年輕人關心核能問題，並投身於反核運動中（崔愷欣，2011 年 9 月 20 日）。2011 年 3 月 11 日，日本發生福島核災，核能問題益加引發社會大眾關心，反核團體發起反核遊行，配合社群媒體的普及，許多與核能相關的臉書粉絲專頁及社團紛紛成立。在眾多專頁中，包括發起許多快閃及創意行動的「我是人，我反核」、匯集大量核能相關訊息的「媽媽監督核電廠聯盟」、聚集許多核工背景人士，試圖以科學角度反駁反核論

述的「核能流言終結者」、以及許多隨著反核實體活動而出現，在階段性任務結束後便沒落下來的粉絲專頁。

由此可見，在核能事件的發展上，無論是群眾動員、資訊彙整或是論述形塑，皆隨著新媒體科技的發展而日益仰賴網路，其中尤其臉書扮演著人們發聲、對話的重要平臺。

2014 年 4 月 17 日，林義雄宣布將以絕食方式反對核四運轉，4 月 26 日，總統馬英九宣布核四公投。林義雄的行動帶起來公民社會以及新聞媒體的大量討論，並且在短時間內對政治場域產生影響。在這短短 10 天的時間，各式各樣的意見在不同場域中大量交流、辯論，最終引發政治系統的回應，本文認為這段時間的核能相關討論在時間以及影響力上皆有其顯著性，因此成為本研究選擇之個案。

## 一、資料蒐集

本研究使用意藍科技提供之工具 Opview Insight 搜尋 2014 年 4 月 17 日至 26 日之間與核能相關之臉書貼文，Opview Insight 蒐集網路上公開網站的資料，在臉書方面，將使用者分為「關鍵意見領袖」、「臉書粉絲專頁」與「公眾訊息」三者<sup>2</sup>，在此我們將焦點鎖定在一般使用者所發表之「公眾訊息」，研究與核能相關之「公眾議題」。

鎖定資料搜索時間區段後，便以「反核」、「核能」、「核四」為關鍵字，蒐集臉書上之「公眾訊息」貼文共 2,157 則，本研究鎖定按讚數前 500 名之貼文進行框架分析。之所以篩選出按讚數前 500 名之文章，是為了找出在社群中較為流傳的討論，因為社群媒體是人人都能說話的平臺，一則訊息的閱讀次數越多，代表其在民意市場中扮演越重要的角色，按讚數在此並不代表人們贊同此訊息，而僅作為被關注程度的指標。

需要進一步說明的是，本文焦點並非對本文選擇的核能爭議事件進行透徹的研究，而旨在藉由個案的分析操演討論方法適用性以及操作過程中浮現之方法論議題。在此研究目的下，為便於應用多種方法對同樣資料進行操作，故僅取少量資料作為本文分析文本。

2 意藍科技將臉書使用者所生產出的內容區分為「關鍵意見領袖」與「公眾訊息」兩者，所謂「公眾訊息」指的應是「關鍵意見領袖」之外所有臉書使用者的貼文內容，然意藍科技並無明確告知「關鍵意見領袖」之挑選標準。使用商營資料庫時大多無法獲知資料庫的資料篩選方法、分析時的演算法等內部資訊，這是使用商營資料庫做為分析資料的限制，也帶來另外的方法論問題，相關討論可參見 Tiessen, McKelvey, & Simcoe (2015)。

## 二、分析方法

本文使用質性框架分析、詞頻分析、詞語共現分析以及情緒分析等不同方法對前述按讚數量前 500 名的臉書「公眾訊息」貼文進行分析。在質性框架分析上，先由研究者瀏覽 500 筆資料後提出基本框架類目，再自所有資料中抽取 30 筆進行先驅研究 (pilot study)，由兩名編碼員先就已設定好的基本框架類目對 30 筆內容進行分類，無法納入既有類目者則經編碼員討論後建立新類目。最終將發言類型分為「意見」、「資訊」、「動員」以及「其他」四類；將核能態度分為「反核能」、「反核四」、「中立」、「支持核能」四類；問題框架則分為「核能安全」、「核廢料」、「替代能源」、「臺灣能源發展方向」、「能源監督機制」、「單純對能源表態」、「政府專業可信任程度」、「政府真誠可信任程度」、「反對者專業可信任程度」、「反對者真誠可信任程度」、「策略正當性」以及「其他」等共 12 類。

最終框架類目完成後，則再次從 500 筆資料中抽取 30 筆進行信度檢驗，兩名編碼員對同樣 30 筆資料進行編碼，編碼時「發言類型」與「核能態度」皆為單選，結「問題框架」則以最明顯的框架為主，若有其他同樣明顯之框架存在，則一則訊息最多可選三則框架。信度計算採用王石番 (1990) 所提出之計算方式，編碼結果在發言類型上兩人編碼完全一致；核能態度相互同意度為 0.83，信度為 0.9；主題框架相互同意度為 0.8，信度為 0.88，信度檢驗皆達可接受程度。

在電腦運算工具方面，本文使用自然語言處理工具進行斷詞，且因收集文本為繁體中文而採用中研院的 CKIP 斷詞系統<sup>3</sup>，斷詞結果為本文詞頻分析、詞語共現分析以及情緒分析的基礎。詞頻分析時以中研院斷詞工具為根據，計算斷詞後的詞頻；在進行詞語共現分析以及情緒分析時，我們則先以「核」做為觀測詞組的起點，並尋找文本中詞組包括「核」字的詞彙，共計 46 個 (見表 1)。接著我們依照 Pointwise Mutual Information (簡稱 PMI) (Church & Hanks, 1990) 的共現計算原則分析與包括「核」字的詞彙經常出現的詞組<sup>4</sup>，共計 909 字。

接著，我們將這些詞依照觀測的議題歸納成不同的概念，其中包括核能安

3 中研院斷詞工具網址請參考 <http://ckipsvr.iis.sinica.edu.tw/>。

4 PMI 值是後文提及 SO-PMI 的基礎，為詞彙共現狀況的計算結果。在 Church & Hanks (1990) 原文中，PMI 值乃用來計算自然語言中的常用詞 (common phrase) 或片語，也應用在介系詞適用情況 (如：\_?\_ order to)，PMI 數值高的則為常用片語 (因為經常一併出現)，它的臨介值一般都設在 3。本文則依照 PMI 將共現詞彙「議題—概念」進行排序，並且依照人工辨別，選出符合研究議題的詞彙。

核四	反核	核能	核電	核電廠	核
核災	廢核	核安	非核	核四廠	核一
核三	核二	核研所	擁核	福島核災	核工
核子	挺核	核彈	核三廠	核管	核武
核心	核一廠	核融合	核爆	核四案	建核
核工系	打核	核一	核二廠	核能廠	禁食廢核
核管處	核反應	核四商	核災正	核五	審核
日廢核	核屁	核四廠區	核能界		

表 1：分析文本「核」字相關詞彙

資料來源：本研究自行整理。

全、興建核電廠意願、替代能源、政策、經濟、聲勢、抗爭手段與象徵、國際關係、官方行動者、民間行動者等等，並且定義了適用於此文本討論內容的正／負向詞表，共計正向字 8 組，負向字 43 組。

最後再根據我們設定的詞表進行議題概念的情緒極性推估。我們依照 Semantic Orientation-Pointwise Mutual Information (簡稱 SO-PMI) 方式 (Turney, 2002) 以如下方式分析<sup>5</sup>：

$$\text{SO-PMI}(\text{"議題-概念"}) = \log_2 \left[ \frac{\text{hits}(\text{"議題-概念"} \text{ NEAR } \text{"excellent"}) \text{ hits}(\text{"poor"})}{\text{hits}(\text{"議題-概念"} \text{ NEAR } \text{"poor"}) \text{ hits}(\text{"excellent"})} \right]$$

其中「議題－概念」為前述人工分類歸納結果。計算時，我們將「議題－概念」中的詞組與「excellent」、「poor」等極性詞組集成對計算共現次數與極性詞組出現次數，並進行「議題－概念」的概念推估。

5 SO-PMI 公式是目前做情感運算上常用的公式，基本想法是假定某個不知道情感極性（正／負）的詞彙「議題-概念」，可以透過與此詞彙同時出現在句中的情感極性詞彙進行推估。假設這個詞彙經常與正極性詞彙（正向詞彙，如：很好、支持、推薦等）出現，則這個詞可推估為正向詞彙；反之，如果這個詞彙經常與負極性詞彙（負向詞彙，如：不好、反對、不建議）共同出現於同一個句子，則這個詞彙「議題-概念」就會推估為負向詞彙。詞彙「議題-概念」的正向或負向決定是依照比較而來，所以是將該詞與所有正向詞共現的結果，除於與所有負向詞共現的結果。在公式中，理論上該寫為： $\text{SO-PMI}(\text{"議題-概念"}) = \log_2 \left\{ \frac{\text{hits}(\text{"議題-概念"} \text{ NEAR } \text{"excellent"}) / \text{hits}(\text{"excellent"})}{\text{hits}(\text{"議題-概念"} \text{ NEAR } \text{"poor"}) / \text{hits}(\text{"poor"})} \right\}$ ，移項後成為內文所列公式。在 Turney (2002) 原文中的 hits 函數是指搜尋引擎「命中」(hits) 符合搜尋項目的資料集合，而其研究範圍是一般性無特定領域的研究，所以使用搜尋引擎所回傳的數值。但由於本研究的文本集合是特別收集的特定領域文本集合（僅討論核四議題），因而在本文中 hits 是指「在我們收集文本中符合特定條件的資料集」。例如 hits("核四\_促進會" NEAR "危險性") 則是搜尋所有收集文本中「核四\_促進會」的前後有出現「危險性」的情況。



在本文中，我們將正負向詞設定為支持核能和反對核能，一個概念與正向詞出現的頻率越高，則代表人們較常以支持核能的態度討論此概念，此概念獲得的正值也會越高，反之亦然。

## 伍、用方法說故事

一般而言，使用電腦運算工具分析文本，會碰到兩個主要困難，一為中文斷詞準確度，另一則為文本意義之判讀。以下分析可看出這兩類難題在實際資料中如何顯現出來。

### 一、詞頻分析 (Word Frequency Analysis)

詞頻分析的主要邏輯就是將文本斷詞之後計算詞彙出現的數量，藉由詞頻分析研究公眾議題，可看出哪些詞彙、哪些概念、哪些名詞、形容詞、動詞較常被人們討論。詞頻分析的假設在於越常被人們提及的詞彙，代表它越受人們重視，因此詞頻分析可以幫助我們掌握哪些是人們重視的議題或概念。

在表 2 中，本文整理了 500 筆臉書貼文中出現頻率高於 50 次的 94 個詞彙，從表中可看出有許多對公眾議題分析無意義的中文常用詞如「沒有、叫做、這些、現在、可以、什麼、就是」等詞彙被納入，將這些常用詞刪除，最常出現的二十個名詞為「核四、臺灣、不要、林義雄、我們、拒絕、政府、反核、核能、先生、核電、如果、發電、公投、反對、問題、大家、因為、核電廠、自己」。這樣的分析結果，顯示了在這 500 筆臉書貼文中，人們最關切核四問題的主要概念，其中「核四」、「核電」和「反核」本就是本文篩選臉書貼文時的關鍵字，這樣的結果理所當然；除了這三個詞外，「林義雄」、「政府」、「公投」等詞彙也常被人們討論，後三者剛好牽涉了這次反核事件的兩造主要行動者以及相關政策議題。

詞頻分析的操作過程看似簡單，但在資料處理時其實已經碰觸重要的方法論問題。原始文本經過斷詞後計算詞頻，結果大多雜亂且數量龐大，如本文資料在斷詞後共有 9,100 個詞彙，其中包括數字、單字、英文、標點符號、錯誤斷詞以及無關研究主題的結果。如此龐雜資料難以直接進行分析，需經過資料清洗步驟，也就是根據研究者的研究意識將斷詞後的結果進一步刪除或合併。斷詞以及資料清洗是電腦工具輔助文本分析的基礎，對詞頻、詞語共現與情緒分析結果都有影響，那麼研究者到底應該如何對資料進行挑選？應該介入到甚麼程度？這就

詞彙	次數	詞彙	次數	詞彙	次數	詞彙	次數	詞彙	次數
核四	560	現在	127	經濟	80	拜耳	64	支持	55
臺灣	458	大家	121	連結	80	希望	63	怎麼	54
不要	348	因為	115	需要	79	還是	63	決定	54
林義雄	274	什麼	113	失效	78	民主	62	門檻	53
我們	273	核電廠	112	已經	76	政治	62	廢料	53
拒絕	203	自己	105	應該	75	古根漢	61	可能	53
沒有	202	人民	105	人家	74	壟斷	60	核災	53
政府	191	服貿	99	不會	73	但是	60	便宜	53
反核	178	營收	97	團體	73	資金	60	發現	52
叫做	176	發生	95	一樣	71	馬英九	60	禁食	52
核能	166	知道	94	總統	71	一下	59	社會	52
這些	165	他們	94	所以	71	無法	59	覺得	52
先生	156	停建	93	你們	71	絕食	58	民進黨	52
核電	149	真的	92	警察	70	工作	58	很多	52
如果	147	今天	90	投資	69	這樣	57		
發電	143	人才	90	為了	68	生命	56		
公投	141	台電	88	臺灣人	66	日本	56		
可以	138	發展	84	杜邦	65	國民黨	55		
反對	137	能源	82	迪士尼	65	安全	55		
問題	129	國家	82	就是	64	不能	55		

表 2：出現頻率 50 次以上之詞彙

資料來源：本研究自行整理。

碰觸到了「研究者角色」這個長久以來的方法論問題。

本文在初步資料清洗上僅先將數字、單字、英文、標點符號刪除，保留錯誤斷詞以及無關研究主題的結果，以展示資料樣貌，從中看見需研究者介入的挑選過程。

詞頻分析可以幫助研究者分析大量資料的媒體議題以及公眾議題，針對表層文字資料的分析，與量化內容分析相比，詞頻分析除了能夠以電腦取代人工，迅速而準確地處理文本外，還能以歸納的方式讓關鍵字由下而上地從文本中浮現出來，解決以研究者為中心由上而下給定觀察詞彙所可能產生之缺漏。然而，詞頻分析僅能幫助研究者找出人們在想些甚麼，而無法知道人們如何思考、如何理解該議題，也無法知道公眾對該議題的意見與態度又是如何。

框架研究認為與議題共同出現的元素會成為議題的屬性，影響閱聽人對議題的理解 (McCombs, 2005)，若研究者希望獲知與議題共同出現的元素為何？則可使用詞語共現工具進一步分析資料。

## 二、詞語共現分析 (Co-Occurrence Analysis)

詞語共現基本計算邏輯在於找出與特定觀測詞彙在文章或句子中同時出現且相鄰的詞彙，視為一個由二個詞彙組成的詞組，計算後的詞組 PMI 值越高，代表該詞組共同出現的頻率越高。藉由詞語共現分析，研究者可以找出與觀測概念較常一起出現的詞彙，而進一步聚焦議題，或是賦予觀測概念特定屬性。如當「核能」常與「爆炸」一起出現時，「爆炸」便成為核能概念的屬性，在公眾議題中突顯了人們對議題的重視層面，在媒體議題中則影響了人們對核能的認識。而「電價」、「林義雄」、「絕食」概念也各自有其意義。

前文說明在進行詞語共現分析以及情感分析時，本研究先以「核」做為觀測詞組的起點，鎖定文本中詞組包括「核」字的 46 個詞彙（見前文表 1）。接著依照 PMI (Church & Hanks, 1990) 的共現計算原則分析與包括「核」字的詞彙經常出現的詞組，分析後共得 909 組詞組。

但如同詞頻分析，詞語共現分析獲得的結果中存在大量無法確定意義、難以納入分析的雜訊如「搭上、上岸、碼頭、屆時、早日、缺、來自、那邊、實在、到底、還是、一定、不會、沒有、真的」或單字「陳、食、正、應、重、公、早、來」等，在人工清理資料挑選對議題框架有意義之詞組後，共留下 195 組詞組。

表 3 整理了與核四、反核與核能三個觀測詞共現頻率最高的十組詞組，由此可發現人們在討論核能相關議題時，常將重點放在興建意願（意願、復建、停建）、核能安全（危險性、癌、傷風、災害、乾淨）、經濟效應（便宜）、兩造行動者（張俊雄、內政、部會、部隊、狗官）、反抗聲勢與象徵（旗、浪潮、信念、高漲、絲帶、大聲、吵吵鬧鬧）。此外，也可看出人們討論「核四」時較會著重在行動者與興建意願上，「反核」較常與運動聲勢連結，而當討論「核能」時，則較常將焦點放在危險性、汙染以及經濟效應。值得注意的是，扮演這段期間最重要行動者的林義雄在與核能相關詞彙的共現關係上反倒較不明顯，PMI 值皆在 1.5 以下（見表 4）。

在詞語共現分析操作過程中處處可見研究者介入挑選、處理資料，除了前述斷詞後的資料清洗外，還包括設定觀測詞組起點、設定共現單位、設定 PMI 臨界值<sup>6</sup>、從初步獲得的 909 個詞組中挑選出分析上有意義的共現詞組以及對工具分析出的詞組資料進行意義詮釋。

6 Church & Hanks (1990) 計算自然語言中的常用詞 (common phrase) 或片語時，將 PMI 臨介值設在 3，但要如何設定臨界值並無標準答案，需依照研究者研究問題而定。

核四共現詞組	PMI	反核共現詞組	PMI	核能共現詞組	PMI
核四_促進會	4.034	反核_反完服貿	6.518	核能_危險性	5.034
核四_意願	3.864	旗_反核	5.518	減碳_核能	5.034
核四_張俊雄	3.864	浪潮_反核	5.518	癌_核能	5.034
核四_吵吵鬧鬧	3.864	象徵_反核	5.518	管制_核能	4.811
裝填_核四	3.864	部隊_反核	5.518	核能_便宜	4.749
狗官_核四	3.864	高漲_反核	4.933	核能_傷風	4.619
核四_大陸	3.864	反核_一同	4.933	核能_乾淨	4.619
核四_內政	3.864	反核_信念	4.518	癌症_核能	4.619
核四_復建	3.601	絲帶_反核	4.296	部會_核能	4.297
核四_停建	3.535	大聲_反核	4.196	災害_核能	4.297

表 3：與核四、反核、核能共現 PMI 值前十詞組

資料來源：本研究自行整理。

「核」與林義雄共現詞組	PMI
義雄_核四	1.405
核四_林義雄	-0.146
林義雄_反核	1.005

表 4：「核」與林義雄共現詞組 PMI 值

資料來源：本研究自行整理。

從以上分析可看出，詞語共現可幫助研究者找出觀測概念與其他詞彙之間的關係，除了找出常與核四、核能與反核共同出現的概念外，若研究者對詞頻分析結果顯示出現頻率高的概念有興趣的話，也可進一步將這些新出現詞彙設為觀測概念，探索人們如何談論或媒體如何再現這些概念。也就是說，詞語共現分析能夠幫助研究者聚焦，並且在框架分析的角度探討與特定概念一同出現的屬性。

然而，詞語共現分析雖能幫助研究者聚焦，卻仍無法協助判斷人們對議題及特定概念的態度與喜好。研究者除了理解人們傾向談論與核能議題相關的那些面向，更希望進一步知道人們對核能議題的整體態度與立場時，那便須仰賴情緒分析工具。

### 三、情緒分析 (Sentiment Analysis)

情緒分析邏輯與詞語共現分析相通，簡單來說就是分析與觀測概念一起出現的詞彙屬於正面或負面情緒，藉此探索人們對此概念之態度。中文的情感分析資源相較於英文而言尚相當缺乏 (Wan, 2009)，目前常見作法包括翻譯英文情緒

分析資源 (Wan, 2009)、使用機器學習方法建立情緒極性字集 (Fellbaum, 1998)、直接透過詞語定義資源建立情緒極性字集 (Huang, Hsieh, Hong, Chen, Su, & Chen et al., 2010) 外, 亦可以直接對現有文本或詞語概念進行情緒極性字集標記 (Ku & Chen, 2007)。

然而, 字詞所代表的情緒正負並非恆久不變, 反倒常因主題、領域以及脈絡有所改變, 在一般概念下「害怕」會被定義為負面情緒, 但若研究者試圖分析人們對恐怖片的評價, 那麼「害怕」成為正面, 「笑」則反而可能是對恐怖片的負面評價。

本文作法為自行根據研究問題以及領域知識從詞語共現所整理出的 195 組詞組中定義正負面詞彙, 試圖藉由情緒分析看出人們對發展核能的支持與否, 因此文中所謂正面指的是「支持核電」, 負面則為「反對核電」。

由於 195 個詞彙中大多為無法明確代表支持或反對核電者, 如建造、復工、確保、裝填、反應爐、儲存、核安、安全、臺灣、燃料、技術等等, 因此將之排除在正負詞表之外。詞彙中能夠代表支持核電的詞彙不多, 本研究僅能列出 8 個「支持核電詞彙」, 「反對核電詞彙」則有 43 個 (見表 5)。

在建立正負詞表的同時, 本研究也將前述 195 個詞彙依照議題歸納成不同概念, 其中包括核能安全、興建核電廠意願、替代能源、政策、經濟、聲勢、抗爭手段與象徵、國際關係、官方行動者、民間行動者等十類 (見附錄)。並依前文提及之 SO-PMI 方式 (Turney, 2002) 進行情緒極性推估, 計算結果如表 6。

SO-PMI 值越高, 則代表議題與支持核電相關詞彙越常一起出現, 分析結果發現人們討論核能的經濟議題時特別會出現支持核電的態度, 在談及政策、替代能源以及國際關係時, 整體立場也較偏向支持核電。然而, 當談及核能安全以及興核電廠意願時, 則整體呈現出偏向反對核電之態度, 尤有甚者, 當談及民間行動者、抗爭手段與象徵以及聲勢時, 與該議題關鍵字共現之情緒極性詞彙全部都是反對核電詞彙, 也就是呈現出全面反對核電的態度。

在情緒分析的操作中可見明顯的研究者介入, 情緒詞表的選擇無須贅述, 正負極性詞的設定以及議題類型所包含的關鍵詞設定等更大大仰賴研究者的領域知識, 「焊接」之所以為負面極詞原因在於人們將核四形容為「拼裝車」, 「路跑」則為當時舉辦的「反核路跑」, 「專業」之所以為正面極詞原因在於「科學性」、「專業性」以及「經濟影響」是核能支持者的主要論述, 「回歸專業」則是常見口號。直接使用一般性的情緒詞表常會因為分析案例領域不同而產生誤差, 既有



正面詞表（支持核電）		負面詞表（反對核電）	
專業	拼裝	癌症	除役
減碳	焊接	傷風	民主
乾淨	威脅	災害	民意
煤	出事	災難	絲帶
便宜	假	福島	旗
漲價	擔心	死	路跑
吵吵鬧鬧	危險	廢料	遊行
貴	癌	核爆	抗議
	恐懼	騙人	苦行
	爆炸	核彈	上街
	斷層	引爆	絕食
	車諾比	烏克蘭	
	承受不了	事故	
	蘭嶼	終結	
	汙染	反核	
	風險	非核	

表 5：核能相關正負詞表

資料來源：本研究自行整理。

議題類型	SO-PMI 值
經濟	4.780
政策	2.466
替代能源	1.916
國際關係	1.041
官方行動者	0.774
核能安全	-0.074
興建核電廠意願	-1.518
民間行動者	只與負極性詞共現
抗爭手段與象徵	只與負極性詞共現
聲勢	只與負極性詞共現

表 6：議題類型情緒極性推估結果

資料來源：本研究自行整理。

情緒分析工具尚無法協助研究者針對個案特性自動化地分類情緒極性詞，而需由研究者扛起責任，根據其領域知識及其對現象、案例、資料的認識而決定。

此外，情緒分析研究的仍是詞組之間的關係，雖可告訴我們在各項與核電相關的次議題上人們呈現出的態度，但仍存在許多限制。一方面，這樣的分析仍

如 Lewis, Zamith, & Hermida (2013) 所述受限於詞語之表層意義，無法分辨人類使用語言過程中的細微差異與潛在意義；另一方面，詞語共現和情緒分析之分析單位為鄰近出現之詞彙，當詞彙距離超過分析所設定的詞組距離時則不會被分析工具所納入，而人類的語言使用存在反諷、隱喻等，並可能藉由句與句甚至段落與段落之間的關係表達意涵，在論述一個事件時，我們可能會先提出一個一般常見的看法，後再舉證推翻它，最後提出自己真正的意見，這些複雜的語言使用皆無法被上述電腦運算工具所捕捉；最後，正因為前述兩項限制，我們似乎無法單純根據詞頻分析與詞語共現分析結果便判斷人們對議題之態度，若想如 Entman & Rojecki (1993) 一般藉由分析文本之深層意涵捕捉文本呈現的議題框架，那仍需仰賴研究者以質性方法實際閱讀文本後做出判斷。

#### 四、質性框架分析

因此，本文用質性方式對 500 則貼文進行框架研究，以貼文為分析單位，分析文本之「發言類型」、「核能態度」與「議題框架」。對「發言類型」進行編碼原因在於過濾出屬於「意見」之貼文，並將與核能議題無關之貼文排除在分析範圍之外，扣除了與核能議題無關的非「意見」型貼文以及原始貼文已被刪除的內容後，最終鎖定了 395 則貼文進行框架分析。

前文說明質性框架分析的操作過程乃先由研究者瀏覽資料建立框架類目後再對所有文本以「貼文」為單位進行編碼。這樣的研究操作過程實際上是先歸納、再演繹，也就是先以對文本進行深度閱讀，就研究者對文本的認識歸納框架類目，再以建立完成的框架類目為前提，再一次閱讀文本並將文本編入既有框架類目中。然質性框架分析與前述幾個電腦分析工具最大差別，則在於無論是歸納框架類目或是將文本編碼入既有框架，皆仰賴研究者藉由深度閱讀方式養成對於一群文本的領域知識，研究者方能跳脫詞語之表層意義而進入人類複雜的語言遊戲中捕捉意義之網。

本文的質性框架研究結果發現，在對核能的態度上反核能者共有 97 則（24%），反核四者有 128 則（32%），中立者 146 則（36%），支持核能者則僅有 24 則（6%）。整體而言，對核能與核四的反對態度占所有貼文的一半以上（56%），足見臉書上人們討論所呈現出之議題態度。

而在對議題框架上，本文將框架分為「核能安全」、「核廢料」、「替代能源」、「臺灣能源發展方向」、「能源監督機制」、「單純對能源表態」、「政府專業可信任程度」、「政府真誠可信任程度」、「反對者專業可信任程度」、

「反對者真誠可信任程度」、「策略正當性」以及「其他」等共 12 類，編碼時採取一則貼文至多三項框架的複選方式。

研究結果發現最常出現的框架為「反對者真誠可信任程度」（122 則），次為政府真誠可信任程度（77 則），其後依次為單純對能源表態（64 則），核能安全（59 則），其他（30 則），臺灣能源發展方向（23 則），核廢料（23 則），替代能源（20 則），政府專業可信任程度（18 則），策略正當性（12 則），能源的監督機制與政策（8 則）以及反對者專業可信任程度（2 則）。進一步分析可發現在這次事件中，許多人們聚焦在主要行動者林義雄的身上，強調林義雄的個人特質與人格，其中有褒有貶，但其個人特質在此事件中確實扮演了重要角色；而政府真誠可信任程度也在人們的討論中特別突顯，由此可見人們討論此議題時大多傾向把焦點放在兩造行動者身上，且主要關注行動者之真誠性而非其核能專業。

事實上，當編碼者閱讀實際文本試圖判斷貼文對核能的態度時，文本表層意義和潛在意義的差異一覽無遺。舉例而言，下引貼文若光看表面字詞或許會將之編碼為反對核能，但藉由「山林濫伐」、「土石流」、「沒事蓋了那麼多水庫」、風力發電「超高分貝的噪音」和「颱風一吹就倒的基座」等等描述，發文者實際以諷刺的方式批評反核者提出之替代能源論述，這正是語言使用之巧妙，也是目前電腦運算技術在分析文本意義時尚無法克服之困難。

既然那麼多人都討厭核能，就拆掉阿～

反正，火力發電造成的地球暖化問題，沒有砍伐原始叢林來的嚴重，天然氣又便宜又可以無限再生（？），沒有理由不用阿～

再說，臺灣沒事蓋了這麼多水庫，水力發電一定可以充分的供電給下游的人用，在加上山林濫伐的土石流，一定會讓渦輪機組更賣力阿～

而臺灣人引以為傲的風力，怎麼不可以拿來用呢，即使每天發出超高分貝的噪音，加上颱風一吹就倒的基座，也不損他純靜無汙染的地位阿！蛤，你說供電量太低？蓋個一萬支，把西部海岸填滿，不就好了嗎？

傻逼，在臺灣石虎都會走專用步道，白海豚都會自己避開工業區了，區區幾隻巨大風車，又不會破壞自然生態。你說對吧！（燦笑～）  
（臉書公眾貼文，文本編號 260）

## 陸、結論

本文整理議題設定與框架相關文獻，將重點放在對媒體議題及公眾議題之分析方法與工具上。藉由相關文獻之爬梳可看出，隨著傳播科技發展，議題設定與框架研究同時發生了量變和質變，可供分析的文本數量迅速增加，使得研究者必須發展新的方法蒐集並分析資料。而電腦運算技術的進展則為研究者提供了更多研究工具，人們必須在各種研究方法與工具中找尋最適合自身研究問題的組合。

為了探討不同工具在媒體議題與公眾議題分析中的特性以及處理邏輯並反省電腦輔助分析操作細節中隱而未顯的主觀性，本文使用詞頻分析、詞語共現分析、情緒分析工具以及質性框架分析法，對同樣一組資料進行分析。從前述內容可發現，個別分析工具確實能夠為研究者帶來不同洞見、說出不同故事，卻也各自存在限制，值得研究者注意：究竟甚麼樣的方法適合應用在甚麼樣的個案中？能夠回答甚麼樣的問題？具有甚麼特性？研究者在運用不同方法時又扮演了甚麼的角色？本文最後一部分將試圖針對上述問題提出一個初步想法，並藉此思考議題設定研究的方法設計。

### 一、工具讓資料說話——運算工具處理邏輯

#### （一）電腦分析工具鎖定議題的能力強，判斷意義的能力弱

本文認為詞頻分析能夠幫助研究者由下而上地找到被人們大量討論的重要議題；詞語共現分析可讓研究者看見欲觀測概念與其他概念之間的關係，進行概念屬性的研究；情感分析則可幫助研究者探索人們對觀測概念之態度。然而，上述以電腦運算工具進行的分析方式至今所遭遇之最大困難便在於無法判定語言之深層意義，若研究者希望深入文本探索意義，仍需仰賴質性分析方法。

#### （二）認清運算工具的視野、優缺與適用性

此外，不同的分析工具因為運算邏輯的不同也有其侷限，如詞頻分析與質性框架分析都指出了林義雄在事件中扮演關鍵角色，但在詞語共現分析中，林義雄與核能相關詞彙的關聯則相當低。若光以詞語共現工具分析本研究的經驗資料，便可能輕視了林義雄在此事件中的重要性。由上述兩點皆可看出，分析工具會影響研究者所可見的視野，在這樣的狀況下，研究者必須首先掌握所使用工具的設計與運算邏輯，理解工具本身在回答研究問題上的特性與弱點。舉例而言，詞語

共現的結果受到工具所設定的觀察單位之影響甚鉅，有的詞語分析工具將觀測單位設定為5，指的是與觀測詞彙相鄰4個詞彙之內者，才會被計算為「共現」。如圖1所示，若觀測單位設定為5，觀測詞彙為F，則BCDEGHIJ會被工具計算為與F共現之詞彙，A與K則不會被納入。

A [ B C D E F G H I J ] K

圖1：詞語共現計算邏輯

資料來源：本研究自行整理。

本研究使用之詞語共現工具設定之觀測單位是5，也是在這樣的設定條件下，林義雄與核四、反核、核能之間的關聯才會較低；若我們採用另一個工具，將觀測單位設定為10，或是以單篇文章為單位，將在同篇文章內出現的詞彙皆視為共現，或許結果就會大大改變。也就是說，研究者若想避免被工具牽著走，便須以研究問題為核心，同時掌握工具的運算邏輯，理解工具提供出的資料之所以如是，乃有其所以然。

## 二、說故事的始終是人——研究者角色

在方法論的討論中，研究者與研究對象的關係以及研究者角色一直是重要主題。17世紀，反對專制主義以及宗教神權帶來了自由主義中的理性主義與經驗主義，到了19世紀，經驗主義和理性主義則在孔德的論述中以實證主義的形象出現。實證主義採取實在論觀點，以物理學為模範堅持可觀察、可驗證的知識才是有效的、正確的知識，因此科學就是藉由科學研究方法觀察、發現、驗證事實，研究者則應成為價值中立的觀察者 (Delanty, 1997)。

海德格以降的詮釋學傳統則否認科學的客觀性，認為科學無法超越其文化文本，所有理解皆須藉由語言中介，而語言則與生活世界息息相關。在這樣的基本立場下，理解更像是對話，研究者在對話的動態過程中，主客體分野已經不再清晰，研究者與研究對象間不斷地相互建構，過程中研究者的視域不斷改變，文本的意義也隨之改變 (Iser, 1972)。因而，詮釋典範下認為研究者不應以自然科學態度看待社會科學，認為社會研究者應該（或可以）反映客觀真實，社會研究者的任務在於盡可能的貼近研究對象的生活去理解並嘗試解釋其意義結構，而不在於「反映」某種真實 (Geertz, 1973)。

那麼，在使用電腦輔助分析大量的文本資料過程中，研究者究竟扮演甚麼樣



的角色？從詞頻分析、詞語共現分析到情緒分析的操作過程中，研究者的前見、理論背景、領域知識的介入，與資料分析詮釋息息相關，研究者不再僅是客觀的工具操作者，熟練地操作著電腦分析工具一步步「發現」真實，而處在持續不斷的閱讀、詮釋過程中，在經驗資料中擴大視域並且回頭賦予資料意義。

實際上越來越多研究者對巨量資料方法進行反思，如 *boyd & Crawford (2012)* 指出雖然大部分人認為數據是客觀的，但研究者處理巨量資料仍經過詮釋過程，因此有其主觀因素，他們對於數據的想像、選擇或排除皆未必客觀中立。

另一方面，*Mahrt & Scharkow (2013)* 則提醒，在龐大的資料下，研究者興趣可能受到資料本身的引導或限制，在設定研究方向時被資料所驅動（*data-driven*）。*Mahrt & Scharkow*（同上引）認為研究者應該要懂得「巨量資料的誘惑」，避免讓資料型態或方法決定自己的研究方向，無須迷信巨量資料，而應採用適當的研究方法回應研究者設計之研究問題。

由此看來，電腦運算工具的興起並不代表研究者的主觀性便能退位，由電腦工具全權負責分析工作，就能保障研究的客觀性。在本文分析過程中，我們發現使用電腦運算工具進行分析本身實際上存在大量人為介入，從個案選擇、蒐集資料時資料庫的選擇、撈取資料時設定的關鍵字、初步分析後清洗資料的邏輯、分析時工具參數的設定到最後對資料進行詮釋，每個階段都如 *boyd & Crawford (2012)* 所述，牽涉對數據的想像、選擇或排除。尤其在清洗資料時，須仰賴研究者對議題本身與文本的認識，舉例而言，「先生」一詞在文本中常用來稱呼林義雄，但在斷詞時先生會被斷為單獨的詞彙，這時，若研究者對文本沒有一定的認識，便可能會認為「先生」為與事件無關的雜訊因而將之刪除。而若進行情緒分析時研究者不僅僅使用一般情緒極性字表，同時也欲根據領域知識建立特殊情緒極性字表，那麼便牽涉了研究者對領域知識甚至手中資料的理解與敏感度。舉例而言，本文在設定正負面詞彙時，若缺乏對臺灣反核運動的領域知識以及手中資料的基本掌握，研究者很可能會就著字面意義將「復工」直接定義為代表支持核電的負面詞彙，但實際上人們討論核電廠之復工大多採批評角度，若未對文本有著一定的熟悉度，資料的表面意義便可能會引導研究者做出錯誤結論。由此可見，無論使用甚麼樣的工具，研究者對現象脈絡的熟悉仍扮演重要角色，而若欲使用資料回應研究問題，研究者也總需扮演資料的詮釋者，而非僅為中立客觀的描述者。

也就是說，巨量資料研究並非撈取大量資料，將資料丟進某個神奇工具後便能跑出答案。從資料取得到分析結論之間，研究者須不斷地確認研究問題、根據

研究問題過濾資料、找尋適合的分析工具、根據分析工具的運算邏輯進一步處理資料，最後再將分析結果與既有理論與文獻對話，找尋最適當的詮釋方法、賦予資料意義。上述每一個階段，研究者都需要仰賴其既存知識和對資料的理解幫助其決斷，藉由一次又一次的抉擇，研究架構方得以確立。

本文認為，電腦運算工具與質性方法的結合，能幫助研究者藉由對資料的初步認識掌握領域知識、建立更為適切的研究架構。因為當我們同時使用電腦運算工具與質性文本分析方法時，不同方法所得出的結果存在相互辯證與補充關係。電腦運算工具除了幫助研究者過濾資料外，也能打破研究者囿於既有概念的預設，由下而上地提供新的洞見；而質性文本分析結果則可幫助研究者在清洗資料及分析資料時能同時留意文本的深層意義，幫助研究者正確地解讀電腦運作工具所分析出的結果。在進行真正的巨量資料分析前，藉由這樣的辯證過程一步步釐清研究架構與假設，或許能夠幫助研究者在後續分析時省下許多功夫。

### 三、新的說故事可能性——結合不同方法的研究想像

量化內容分析一般能夠達到更為理想的信度值，是因為無須仰賴研究者對文本意義的理解與判斷，僅要求編碼者根據文本之表面意義進行編碼 (Himelboim et al., 2014)，在清楚明確的編碼指示下，無論是內在信度或是外在信度皆可輕易達標。電腦運算工具可說是此分析邏輯的進一步發展，利用電腦取代人腦，只要給與電腦清楚的編碼指令，電腦便能快速而準確地處理人腦無法處理的大量資料，降低人為誤差的同時也進一步提高了信度<sup>7</sup>。

究而察之，電腦分析工具之所以能夠大幅提高信度，除了降低人為誤差外，更重要的原因在於其分析單位。質性文本分析之分析單位常為文章或是段落，而在文章、段落、句子、詞等不同分析單位中，電腦分析工具多以「詞」(word)作為基本分析單位，也就是說，為了讓電腦得以「閱讀」並「分析」文本，電腦分析工具傾向將文本切到最小意義單位，然而，此舉則可能犧牲了研究效度，因為人類的語言使用並不僅僅是個別詞的組合加總，在人類複雜的語言遊戲中，需考慮脈絡，脈絡包括文字組合、前後文以及語用情境方能更為適確地解讀意義

7 事實上，若將前述研究者的介入選擇、詮釋過程考量在內，則電腦分析工具的信度便需重新評估。若將情緒分析以及共現分析的資料清洗、條件設定都定下來，那麼重複操作的信度當然會非常高，但若不同研究者各自挑選各自熟悉的斷詞工具，並依自身領域知識以及對個案的認識進行資料清洗，將詞組刪除、合併、歸類後進行詞頻、詞語共現或是情緒分析，最終進行資料詮釋，最後是否仍能呈現出高信度？則有賴進一步研究設計進行解答。

(洪漢鼎, 1995), 如 Lewis et al. (2013) 便指出雖然電腦運算能幫助研究者快速處理大量資料, 但運算法仍無法分辨語言中的細微差異與潛在意義, 他認為利用詞頻分析或詞語共現分析大多僅能捕捉字詞表面意義, 難以掌握文本的內在意義。若研究焦點為找到重要的議題或關鍵字, 那麼電腦運算工具確實能夠兼顧信效度完成任務, 但若研究者希望進入更為深層的語言使用探索議題框架, 電腦運算工具便會遭遇難以克服的信效度問題。那麼, 有沒有一種方法設計能夠結合電腦分析工具以及質性研究的優點, 兼顧研究的信度與效度呢?

基於以上反思, 越來越多研究者傾向綜合巨量資料方法以及質性研究法, 試圖在資料規模與研究深度間找到平衡, 同時運用多種方法幫助回答研究問題。Lewis et al. (2013) 便認為應結合電腦運算與人工方法, 如此一來, 一方面可藉由電腦運算幫助分類, 提供更多線索來幫助人工編碼, 進而提升信度並以運算分析方法來看新媒體的結構特性; 同時也能在人工編碼時觀察文本所富含之社會文化情境。Small (2011) 研究討論加拿大政治議題的推特文章時便利用了 Twitter 的結構特性幫助其判斷文本類型, Small 在判斷 Twitter 貼文類型 (分享/評論/告知式評論) 時, 採用人工判斷分類的方式, 而當他試圖判斷貼文是否為對話時, 則用內文出現的「@reply」與「retweet」等 Twitter 之結構特性作為判斷指標。

Manovich (2012) 舉例說明較佳的合作方式, 他認為若要研究某國家在某一段期間內上傳的所有 YouTube 影片類型, 結果找出了 100 萬部分影片, 接下來可先以電腦運算幫助研究者自動將影片分成幾個類目, 並指出各類目最典型 (typical) 與最獨特 (unique) 的影片。這程序可幫助研究者將研究範圍從 100 萬部影片縮小為幾種類型的影片, 再從當中進行抽樣, 實際觀看、分析每類型中最典型與最獨特之影片, 來回幾次之後, 研究者便可以逐漸掌握這些影片特徵。

Graves et al. (2014) 的研究設計或能與 Manovich 相互呼應。Graves et al. (同上引) 研究 2012 年 NFL 超級盃期間討論超級盃相關內容的 Twitter 時共蒐集到 797,128 則推特貼文, 但他們有興趣的是其中與競賽 (gameplay) 相關的那些貼文, 於是利用與競賽相關的關鍵字 (gameplay words) 如球員名字、足球術語等幫助篩選資料。將資料限縮至競賽相關貼文後, Graves 等人便進一步研究貼文中的共通詞彙, 並分析這些詞彙在貼文中的位置與意義。在此例中電腦運算便扮演過濾資料的功能, 鎖定了欲分析的詞彙後, 研究者則可分析特定詞彙隨著時間的數量變化, 或是採取質性方法進一步分析特定詞彙出現的推文內容。

除了質性文本分析方法外, 以電腦運算配合訪談也是可行的搭配。Hassid

(2012) 研究傳統媒體和部落客討論公共議題時相互的影響關係時，便同時採用了內容分析、相關分析與訪談。他先利用關鍵字詞庫進行不同媒體中的詞頻分析、分析關鍵字出現頻率在不同媒體間之關聯性，並以訪談佐證數據結果。

也就是說，電腦運算工具可作為過濾龐大資訊、幫助研究者聚焦的有效方法，藉由機器學習 (machine learning)、後設資料分析 (metadata analysis)、關鍵字分析、詞頻分析、詞語共現分析等工具，研究者能夠很快地在巨量資料中找到與其研究問題最為相關的資料集，並在聚焦後搭配質性文本分析方法詮釋資料意涵，甚至藉由訪談對文本做進一步的闡釋。

隨著電腦運算技術的發展，可供研究者使用的資料分析工具越來越多，綜合使用傳統文本分析方法與電腦運算工具的例子也日漸增加，但最重要的是，我們應理解不同分析方法與工具之特性，並將 Mahrt & Scharkow (2013) 的提醒謹記在心，從研究者關心的問題出發，尋找適合的研究方法幫助回答問題。面對社群媒體時代產製出的大量文本，使用電腦運算工具輔助分析似為必須採取之策略，但若回應議題設定與框架研究之理論概念，研究者便須考量電腦文本分析工具「以詞彙為單元」以及框架分析「以意義為依歸」之不同取徑，透過研究設計結合二者之優點，方能幫助研究者更為適切地掌握當代社群媒體大量文本之意義流變脈絡。

在研究方法的操作過程以及資料詮釋過程中，本文則認為應回歸詮釋典範對研究者角色的觀點，將研究過程視為研究者與研究對象在對話中持續地擴展視域 (Iser, 1972)，著重研究者與資料對話過程中對「意義之網」的重視 (Geertz, 1973)，清楚意識研究者在研究過程中所涉及的選擇與詮釋，由此打開研究者反身性思考的空間，在交互主體的對話過程中盡可能貼近、理解研究對象的意義結構。

## 柒、研究限制與未來研究建議

本文在有限篇幅中無法探討所有運用電腦輔助文本分析工具的議題，有些議題將留待未來探討，像是使用社群資料供應商所提供的資料時，其實涉及商營資料公司對於一般公眾的母體想像為何？蒐集資料時抽樣方法為何？以及分析邏輯的演算法之不透明性等。臉書資料始終存在因隱私設定所帶來的代表性問題，而即使是較易取得的 Twitter 資料也仍面臨免費開放取得之資料量有限之問題。前

述每個問題皆牽涉重要的研究方法及方法論議題，值得專文深入探討，在本文有限篇幅中，僅優先藉由突顯操作過程細節中所存在的問題，探討「運算工具處理邏輯」及「研究者角色」兩大議題。

其次，由於本文嘗試藉由運用不同分析方法與工具在同一批資料上，加上本文除了電腦分析工具外還引入了質性框架分析，因此納入分析之貼文資料筆數僅有 500 筆，但經由電腦分析工具進行斷詞與詞彙統計，獲得上千個詞彙及其詞頻，必須透過電腦運算工具才能幫助我們找出詞彙與詞彙之間的關聯性。本文旨在反思電腦輔助文本分析的操作過程研究者所需介入的種種細節，並探討這些操作對結果產生的可能影響，在此情況下，總體資料筆數的多寡並不足以影響本文之參考價值。

## 參考書目

- 王石番（1990）。《傳播內容分析法：理論與實證》。臺北，臺灣：幼獅文化。
- 李紅豔、劉碧瑤（2014）。〈地方媒體對群體性事件的解讀——以「上海普陀事件」與「廣州增城事件」的報導為例〉，《新聞學研究》，120：79-119。
- 林俊宏譯（2013）。《大數據》。臺北，臺灣：天下文化。（原書 Mayer-Schönberger, V., & Cukier, K. [2013]. *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.）
- 洪漢鼎（1995）。《語言學的轉向——當代分析哲學的發展》。臺北，臺灣：遠流。
- 崔愷欣（2011年9月20日）。〈臺灣反核運動的歷史與策略（從1980至2011）〉，《苦勞網》。取自 <http://www.coolcloud.org.tw/node/64150>
- 陳百齡、鄭宇君（2014）。〈從流通到聚合：重大災難期間浮現的資訊頻道〉，《新聞學研究》，121：89-125。
- 陳憶寧（2002）。〈公共議題之遊戲框架初探：以核四議題為例〉，《新聞學研究》，72：85-117。
- 陳韻如（2011）。〈保護誰的生命？秘魯婦女運動公共議題的新聞框架〉，《新聞學研究》，109：121-166。
- 鄭宇君（2014）。〈向運算轉：新媒體研究與資訊技術結合的契機與挑戰〉，《傳播研究與實踐》，4：67-83。
- 謝君蔚、徐美苓（2011）。〈媒體再現科技發展與風險的框架與演變：以基因改造食品新聞為例〉，《中華傳播學刊》，20：143-179。



- Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F. (2014). A survey of data mining techniques for social network analysis. *Journal of Data Mining & Digital Humanities*. Retrieved from <http://jdmhdh.episciences.org/18/pdf>
- Anstead, N., & O'Loughlin, B. (2014). Social media analysis and public opinion: The 2010 UK general election. *Journal of Computer-Mediated Communication*. doi: 10.1111/jcc4.12102
- Burgess, J., & Bruns, A. (2012). (Not) the twitter election: The dynamics of the #ausvotes conversation in relation to the Australian media ecology. *Journalism Practice*, 6, 384-402.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16, 22-29.
- Delanty, G. (1997). *Social science: Beyond constructivism and realism*. Buckingham, UK: Open University Press.
- boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15, 662-679.
- Entman, R. (1993). Framing: toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
- Entman, R., & Rojecki, A. (1993). Freezing out the public elite and media framing of the U.S. anti-nuclear movement. *Political Communication*, 10, 155-173.
- Fellbaum, C. (1998). *WordNet*. Hoboken, NJ: Blackwell.
- Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York, NY: Basic Books.
- Graves, I., McDonald, N., & Goggins, S. (2014). Sifting signal from noise: A new perspective on the meaning of tweets about the "big game." *New Media Society*. Retrieved from <http://nms.sagepub.com/content/early/2014/07/09/1461444814541783.full.pdf+html>
- Hassid, J. (2012). Safety valve or pressure cooker? Blogs in Chinese political life. *Journal of Communication*, 62, 212-230.
- Himmelboim, I., Sweetser, K. D., Tinkham, S., Cameron, K., Danelo, M., & West, K. (2014). Valence-based homophily on twitter: Network analysis of emotions and political talk in the 2012 presidential election. *New Media Society*, 1-19. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/1461444814555096>

- Huang, C., Hsieh, S., Hong, J., Chen, Y., Su, I., Chen, Y., et al. (2010). Chinese Wordnet: Design, implementation, and application of an infrastructure for cross-lingual knowledge processing. *Journal of Chinese Information Processing*, 24(2), 14-23.
- Iser, W. (1972). The reading process: A phenomenological approach. *New Literary History*, 3, 279-299.
- Groshek, J., & Al-Rawi, A. (2013). Public sentiment and critical framing in social media content during the 2012 U.S. presidential campaign. *Social Science Computer Review*, 31, 563-576.
- Ku, L., & Chen, H. (2007). Mining opinions from the web: Beyond relevance retrieval. *Journal of American Society for Information Science and Technology*, 58, 1838-1850.
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, 57, 34-52.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*. Retrieved from <http://www.morganclaypool.com/doi/abs/10.2200/S00416ED1V01Y201204HLT016?journalCode=hlt>
- Mahrt, M., & Scharrow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, 57, 20-33.
- Manovich, L. (2012). Trending: the promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 460-475). Minneapolis, MN: The University of Minnesota Press.
- McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, 6, 543-557.
- McCombs, M., & Shaw, D. (1972). The Agenda-setting function of mass media. *Public Opinion Quarterly*, 36, 176-187.
- Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention-agenda-setting theory meets big data. *Journal of Communication*, 64, 193-214.
- Nisbet, M. C. (2009). Knowledge into action: Framing the debates over climate change and poverty. In P. D'Angelo & J. Kuypers (Eds.), *Doing news framing analysis:*

*Empirical, theoretical, and normative perspectives* (pp. 46-83). New York, NY: Routledge.

- Song, Y. (2007). Internet news media and issue development: A case study on the roles of independent online news services as agenda-builders for anti-US protests in South Korea. *New media & Society*, 9, 71-92.
- Small, T. A. (2011). What the hashtag? A content analysis of Canadian politics on Twitter. *Information, Communication & Society*, 14, 872-895.
- Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1991). *Reasoning and choice: Explorations in political psychology*. Cambridge, UK: Cambridge University Press.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford, UK: Blackwell.
- Tiessen, M., McKelvey, F., & Simcoe, L. (2015). A consensual hallucination no more? The Internet as simulation machine. *European Journal of Cultural Studies*, 18, 577-594.
- Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Association for Computational Linguistics (Ed.), *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 417-424). Stroudsburg, PA: Association for Computational Linguistics.
- Wan, X. (2009). Co-training for cross-lingual sentiment classification. In Association for Computational Linguistics (Ed.), *Proceedings of the Joint Conference of the 47th annual meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (Vol. 1, pp. 235-243). Stroudsburg, PA: Association for Computational Linguistics.
- Yuan, E. J., Feng, M., & Danowski, J. A. (2013). "Privacy" in semantic networks on Chinese social media: The case of Sina Weibo. *Journal of Communication*, 63, 1011-1031.
- Zhou, Y., & Moy, P. (2007). Parsing framing processes: The interplay between online public opinion and media coverage. *Journal of Communication*, 57, 79-98.

## 附錄：情緒分析主題與主題詞彙 (Snippets)

核能安全	確保	危險	核安	控制	廢料	高階	孩子
	裝填	出事	家園	燃料	核爆	處置	核彈
	拼裝	儲存	臺灣	發電廠	指數	承受不了	引爆
	安檢	工程	減碳	災難	沿海	貢寮	烏克蘭
	焊接	假	癌	煤	耐震	貯存場	事故
	信心	專業	癌症	技術	恐懼	蘭嶼	
	威脅	機組	傷風	嚴重	爆炸	汙染	
	工地	平安	乾淨	福島	斷層	風險	
	反應爐	擔心	災害	問題	負責	騙人	
	燃料棒	安全	發電機	死	車諾比	把戲	
興建核能意願	復建	完工	反	繼續	停掉	逐步	
	停建	建	同意	建設	終結		
	復工	運轉	停止	完成	懷疑		
	續建	廢除	禁	反對	支持		
	去留	興建	營運	發電	建造		
	廢	推動	停工	反核	除役		
	發動	落實	支持	非核	封存		
	終結	贊成	要不要	停止	蓋		
	試	商轉	挺	需要	關閉		
	成立	蓋	通過	反對	暫緩		
國際關係	大陸						
	日本						
	美國						
	美金						
	中國						
官方行動者	張俊雄	當權者	內政	國民黨			
	資政	委員會	任內	行政院			
	總統府	部會	官員	主席			
	馬英九	台電	立院	總統			
	狗官	政府	核研所				
民間行動者	林義雄	專家	學生				
	義雄	李國鼎	團體				
	李登輝	蘇貞昌					
	人員	民進黨					
	民眾	人士					

替代能源	天然氣 煤 火力 再生 能源	
政策	條例 聯署 公投 連署書 連署	門檻 管制 政策 政治
經濟	股市 經費 追加 預算 便宜	漲價 成本 貴 電價 錢 電費 度 經濟
聲勢	沸沸 揚揚 吵吵鬧鬧 爭議 吵	壓力 激烈 亂 堅持 高漲 力量 浪潮 象徵 信念 決心 行動



# Telling Stories with Different Methods: Exploring the Use of Computational Analysis Tools on Framing Research

Lu, An-Pang\*

Doctoral Candidate, College of Communications, National Chengchi University

Cheng, Yu-Chung

Associate Professor, Department of Mass Communication, Hsuan Chuang University

## Abstract

As the slogan “let the data speak for itself” gains currency, it becomes important to reconsider the role of the researcher. This study tries to explore the intertwined relationships among data, tools and researchers. In this study, we review the literature on agenda-setting research and framing research, with a particular focus on research methods and analysis tools. To investigate the use of different analysis tools to study media/public issues, this article introduces three analysis tools, word-frequency analysis, co-occurrence analysis, and sentiment analysis, in addition to the qualitative frame analysis method. We find that different analysis tools tell different stories and researcher choices are always important across different research periods. We think that if researchers consider the results obtained using different tools, they will be able to devise a more comprehensive research structure. Therefore, in this article we suggest that researchers apply multiple tools to their data during the pilot study phase.

**Keywords:** social computing, word-frequency analysis, co-occurrence analysis, sentiment analysis, framing, agenda setting

---

\* Email: donkeydumb326@gmail.com

Received: 2015.7.11

Accepted: 2016.1.27