

國家科學及技術委員會人文社會科學研究中心  
補助學術研究群暨經典研讀班結案報告

網路言論型態與傳散之研究：厚描數據與資訊素養的嘗試  
學術研究群

計畫編號：MOST 110-2420-H-002-003-MY3-GB11102

執行期間：111 年 1 月 1 日至 111 年 12 月 31 日

計畫召集人：劉慧雯

執行機構及系所：政治大學新聞系

中 華 民 國 111 年 12 月 31 日

國家科學及技術委員會人文社會科學研究中心  
補助學術研究群暨經典研讀班結案報告

網路言論型態與傳散之研究：厚描數據與資訊素養的嘗試

類別： 學術研究群  經典研讀班

計畫編號：MOST 110-2420-H-002-003-MY3-GB11102

執行期間：111 年 1 月 1 日至 111 年 12 月 31 日

執行機構及系所：政治大學新聞系

計畫召集人：劉慧雯

計畫成員：劉慧雯、鍾曉芳、紀明德、江玥慧、邱淑怡、傅文成、謝承諭、段人鳳

兼任助理：呂貞妮

中華民國 111 年 12 月 31 日

## 補助學術研究群暨經典研讀班成果自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否具有政策應用參考價值及具影響公共利益之重大發現）或其他有關價值等，作一綜合評估。

### 1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

本案申請規劃辦理 26 場會議，實際執行 16 場會議

達成目標

未達成目標（請說明）

說明：

受到 111 年 4 月底起突然加劇的疫情影響（每日破萬人確診，持續至年底），包括教育部、大學在內各種防疫規範進入經常性變動的狀態。雖然這波疫情對人的影響較小（輕症者佔 99% 以上），然而面對疫情迅速變化、難以預估個人在其中可能受到傷害的大前提影響下，邀約學者移動到他校，並且在相對封閉的教室內進行演講，對大家來說還是一個相對有壓力的狀態。因此，邀約講者並不容易。再加上補助單位並不接受線上演講、線上參與，使得社群成員也難以以多元方式參與。

### 2. 研究成果在學術期刊發表或申請專利等情形（請於其他欄註明專利及技轉之證號、合約、申請及洽談等詳細資訊）

論文： 已發表  尚未發表之文稿  撰寫中  無

專書： 已出版  尚未出版之書稿  撰寫中  無

其他：

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（敘述成果所代表之意義、價值、影響或進一步發展之可能性）。

說明：

(1)學術成就

本學術社群以厚描數據為關切對象。本年度邀請的講者中，大多數都是有豐富資料處理經驗的研究者，研究範疇廣及社群文本（如：YouTube 頻道留言、PTT 推噓文等）、影像與鏡面（如：直播頻道）、古代宗教典籍（如：佛教），以及族群（如：客家語言）等議題與型態。數據研究在每個領域都有投入的工作者，每個人處理的資料型態也相當不同。本年度透過學術社群演講，將領域頂尖人物聚集於一處，有助於社群成員/參與者掌握各領域的發展。

透過本學術社群舉辦的演講活動，成員之間的溝通交流更為順暢頻繁。本年度亦集中火力發展出至少五篇研究論文投稿到期刊，其中有一篇合作著作已經刊出。本學術社群的五位成員，本年度亦向國科會提出整合型計畫，並且均獲通過多年期計畫。該整合型計畫研究主題如下表：

計畫項目	主持人	服務機構、職稱	計畫名稱
總計畫	劉慧雯	政治大學新聞系教授	疫意不同：Covid-19 疫情中後的社群閱聽人語言實踐
子計畫一	劉慧雯	政治大學新聞系教授	不經意科學：社群媒體使用者的媒介化疫情資訊之旅 MOST111-2410-H-004-129-MY2
子計畫二	鍾曉芳	政治大學英國語文學系教授	疫情中之隱喻及隱喻框架：以語料庫為基礎的研究 MOST111-2410-H-143-130-MY2
子計畫三	謝承諭	國家教育研究院語文教育及編譯研究中心助理研究員	(後)疫情時代線上健康言談的科普修辭與閱聽人參與：台灣社群媒體醫學知識、立場與認同的多模態建構 MOST111-2410-H-656-013-MY2
子計畫四	段人鳳	臺東大學通識中心助理教授	作為疆界物件的疫苗 MOST111-2410-H-143-013-MY2
子計畫五	邱淑怡	政治大學資訊科學系助理教授級約用教學人員	運用深度學習探究疫情間社群媒體閱聽人之言論 MOST111-2410-H-143-131-MY2

## (2)技術創新

本學術社群主要成員來自三個領域：傳播學、語言學與資訊科學，絕大多數學術討論落在如何促發人文社會領域學者使用資訊科技，以及如何使資訊科學研究主旨緊扣社會發展等兩個面向。因此，本研究社群主要工作都在應用或改善既有技術。

本學術社群成員閱讀、使用自然語言處理相關文獻，並且產生一定數量研究論文後，在兩個方面有些技術進度。首先，社群成員開發合適社會人文領域學者上手使用的自然語言處理軟體，並布置視窗介面，協助不同類資料的斷詞、TF-IDF 詞彙向量產生、LDA 語言分析模型等分析工作在相對友善的架構中。這有助於非資訊科學領域學者自行初步處理所蒐集之資料。

另一方面，本學術社群成員在處理社群資料時，亦注意到表情符號 (emoji) 在表達、改變意義上的關鍵地位，因此開始發展處理表情符號的可能步驟程序。目前進入實作、修正的階段。

以上兩個技術活動在社群成員、參與者接下來的研究與教學活動中，都有可能成為導引人文社會領域學子進入跨領域素材處理的範疇。有如統計活動在套裝軟體化之後，成為容易親近的研究方法，如此，資料科學亦有相當潛力成為易於使用的研究工具。

## (3)社會影響

本學術社群關切社群媒體上的語言使用，計畫執行期間，正值全球遭遇 COVID-19 重擊。台灣剛走過最嚴峻考驗的三級警戒，又迎來傳染力高，感染人數指數上升的變種病毒。在人心惶惶又需要維持社交距離的時刻，社群媒體成為人們交換資訊、瞭解風險暴露，乃至於評估行政決策的重要場域。本研究在此時對社群媒體資訊的關切，正好與此一公共衛生議題扣連在一起。

透過不同研究主題（如：校正回歸、疫苗）的資料挖掘與分析，研究方法（如：自然語言處理、情緒分析）的嘗試，以及與不同領域學者（如：族群、宗教、典籍）的對話與相互理解，本研究社群的成員可以將在此所蒐集到的學術內涵導引回各自的教學研究領域，提供新一代學子掌握社會氛圍的工具與作法。

除此之外，本研究社群部分成員亦從中掌握了社群媒體使用活動中的素養議題，發現溝通實非如前人所想，乃是公共領域、公共論壇的良善之所。相反地，或出自於恐懼，或起因於無知，社群媒體經常是錯誤資訊、惡意或仇恨資訊的集

散地，當人們習而不察，就有可能形成全社會集體的誤解與恐慌。

### 補助學術研究群暨經典研讀班成果彙整表

計畫召集人：劉慧雯		計畫編號：MOST 110-2420-H-002-003-MY3-GB11102			
計畫名稱：網路言論型態與傳散之研究：厚描數據與資訊素養的嘗試					
成果項目		量化	單位	質化 (說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等)	
國內	學術性論文	期刊論文	7	篇	註記於本檔案第 6-7 頁。
		研討會論文	3		註記於本檔案第 6-7 頁。
		專書	0	本	
		專書論文	0	章	
		其他	0	篇	
國外	學術性論文	期刊論文	4	篇	註記於本檔案第 6-7 頁。
		研討會論文	2		註記於本檔案第 6-7 頁。
		專書	0	本	
		專書論文	0	章	
		其他	3	篇	註記於本檔案第 6-7 頁。
本國籍	教授	5		劉慧雯、鍾曉芳、劉吉軒、高照明、傅文成	

參與計畫人力		副教授	5	人次	江敏華、馬偉雲、蕭季樺、呂佳蓉、紀明德
		助理教授	5		段人鳳、謝承諭、王昱鈞、江玥慧、邱淑怡
		博士後研究員	0		
		兼任助理	1		呂貞妮
		其他	3		蔡依霖、王孝成、蘇冠綸
	非本國籍	教授	0		
		副教授	0		
		助理教授	0		
		博士後研究員	0		
		兼任助理	0		
		其他	0		
	其他成果				
(無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)					

	成果項目	質化
1	國內期刊論文	<p>1. Duann, R.-F., Chiu, S.-I., &amp; Liu, H.-W. (2022). Constructing a Deep Learning Model Using Language in Social Media: The Case Study of ‘Retrospective Adjustment’. <i>International Journal of Computational Linguistics &amp; Chinese Language Processing</i>, 27(1): 153-180.</p> <p>2. Chung, S.F. (2022) "A Morphosyntactic and Morphosemantic Analysis of the Malay FEAR Lexeme." <i>Concentric: Studies in Linguistics</i>. 48(2). 147-173. (THCI-Level 1, ESCI, Scopus). <a href="https://doi.org/10.1075/consl.22015.chu">https://doi.org/10.1075/consl.22015.chu</a> [110-2410-H-004-085-]</p> <p>3. Hsieh, C.-Y. C. &amp; Chang, Y.-Y. (2022). An N-gram approach to identifying the Chinese linguistic signals for the Problem-Solution pattern in annotated online health news. <i>International Journal of Computational Linguistics and Chinese Language Processing</i> 27 (1), 75-110. (THCI)</p> <p>4. 楊惟婷、謝承諭、鍾曉芳(2022)。〈應用文步分析探究言語行為—以公共政策網路參與平臺提案文類為例〉。《中文計算語言學期刊》，第27卷，第1期，頁53-74。(THCI)</p> <p>5. 劉大華、傅文成(2022 Accept)。〈從Twitter戰狼外交策略檢驗中共戰略文化的轉變〉，《中國大陸研究》(TSSCI)。國科會計畫111-2511-H-606-001。</p> <p>6. 鍾曉芳、曾瑋庭(2022)。「類別隱喻構式研究」。《中國語文通訊》(Current Research in Chinese Linguistics)。第101卷第2期。193—212頁。(人文學核心期刊-第三級 THCI-Tier 3 Journal)。DOI: 10.29499/CrCL.202207_101(2).0009 [109-2410-H-004-163-] [108-2410-H-004-095-]</p> <p>7. 顏瑞宏、傅文成(2022)。「外交新常態?以主題及網絡建模技術探索中共Twitter外交的戰狼溝通策略」，《資訊社會研究》，43，67-114。科技部計畫110-2511-H-606-001。</p>
2	國內研討會論文	<p>1. Chung, S.-F. (2022) “Meeting AntConc and the Creator: Corpus Linguistics in the Years to Come.” <i>Continuous Evolution of Corpora — Looking Back at 20 Years of Corpus Development (一日千里的語料庫：語料庫近20年的發展)</i>. December 9, National Chengchi University.</p> <p>2. 江玥慧(2022年06月)。運用合作學習於非資訊領域學生程式設計課程之教學設計。第18屆台灣軟體工程研討會(TCSE2022)，台北(線上)。本人為第一作者、通訊作者。</p>

		3. 劉慧雯 (2022)。〈社群媒體語用研究新方向：功能詞角度的探索〉，發表於「中華傳播學會 2022 年年會」。2022 年 6 月 18-19 日。線上會議。
3	國外期刊論文	1. Chiang, Y.V.*, Cheng, Y.W. & Chen, N.S. (2022, Jul). Improving Language Learning Activity Design through Identifying Learning Difficulties in a Platform Using Educational Robots and IoT-based Tangible Objects. Educational Technology & Society. (Accepted). (SSCI). MOST 109-2511-H-004-004.
		2. Chung, S.-F. (2022) “Emotion Terms in Malay: Patterns in [meN-], [meN- - i], [meN- -kan], and [-kan]-only Forms.” Language and Linguistics. 23(2). pp. 147–190. (SSCI, Scopus, THCI)
		3. Lin, Y.-C. & Chung, S.-F (2022) "A Corpus-based Study of Native Speakers' and Taiwanese EFL Learners' Use of the Adverb JUST". Corpora 17 Supplement. 99-117. Edinburgh University Press. <a href="https://doi.org/10.3366/cor.2022.0249">https://doi.org/10.3366/cor.2022.0249</a> (Scopus)
		4. 吳文欽、孫懋嘉、傅文成 (2022)。 (Un)welcome guests: How Taiwanese Respond to Chinese and US Military Presence in the Taiwan Strait, China Quarterly (SSCI)
4	國外研討會論文	1. Chung, S.-F. (2022). Metaphorical Mechanisms of the Suffix -an in Malaysian Standard Malay. To be presented the <a href="#">15th Researching and Applying Metaphor Conference</a> . University of Bialystok, Poland. September 21-24.
		2. Chiang, Y.V.*, Lin, Y.R. & Chen, N.S. (2022, Jul). Using deep learning models to predict student performance in introductory computer programming courses. The 22th IEEE International Conference on Advanced Learning Technologies (ICALT 2022), Bucharest, Romania & online. MOST 110-2511-H-004-003.
5	國外其他	1. Duann, R.-F. (Under Revision). Folding or Traffic Congestion? On the Behavior of Agent and Object Metaphors for ‘Retrospective Adjustment’. International Review of Pragmatics.
		2. Hsieh, C-Y. C. (Under Revision). “It sounds a little cool”: The multimodality and multifunctionality of evaluation in the English-teaching videos of Taiwan-based YouTubers.
		3. Liu, H.-W. (Under Revision) ‘In fact’ on PPT Gossip: A Social Media Pragmatics Research on Functional Word. International Review of Pragmatics.

## 摘要及關鍵字

本社群計畫名稱為「網路言論型態與傳散之研究：厚描數據與資訊素養的嘗試」，為一跨領域的研習與討論，其中涵蓋傳播學、語言學與資訊科學。透過跨領域的研討，本計畫一方面希望結合語言學和資訊科學在自然語言處理方面的領域知識，能夠快速掌握大量網路社群言論的型態與傳播特徵；另一方面，則希望對社群言論的討論可以關照到具體的社會情境（例如：疫情、選舉等），深化對於數據的厚描解讀，並期待進一步為「媒體與資訊素養」的領域有所貢獻。

透過本社群舉辦的 16 場演講活動，成員們分別在學術成就、技術創新、社會影響方面皆有豐碩的成果。學術成就方面，跨領域的成員間彼此頻繁的交流，因此，造就成員在本年度至少投稿了五篇研討論文到期刊，其中一篇合著著作已經刊出。技術創新方面，舉例而言，透過社會人文領域的學者與資訊科學學者之間的合作，社群成員開發出適合社會人文領域學者操作使用的自然語言處理軟體，並佈置友善的視窗介面，使得社會人文領域學者相對容易上手。社會影響方面，本社群進行時正逢 COVID-19 對於全球產生重擊的時刻，與公共衛生議題的扣連使得部分社群成員，從中掌握了社群媒體使用活動中的素養議題，發現溝通事實上並非前人所想，乃是公共領域、良善的公共論壇。反之，可能是出自恐懼，又或是起因於無知，社群媒體常常是錯誤資訊、仇恨或惡意的資訊集散地。因此，當人們習而不察，就有可能形成全體社會集體的誤解與恐慌。

關鍵字：網路言論型態與傳散、厚描數據、媒體與資訊素養

## 目錄

摘要及關鍵字.....	3
壹、報告內容.....	5
第一次研究群會議記錄.....	5
第二次研究群會議記錄.....	6
第三次研究群會議記錄.....	7
第四次研究群會議記錄.....	8
第五次研究群會議記錄.....	9
第六次研究群會議記錄.....	10
第七次研究群會議記錄.....	11
第八次研究群會議記錄.....	12
第九次研究群會議記錄.....	13
第十次研究群會議記錄.....	14
第十一次研究群會議紀錄.....	15
第十二次研究群會議記錄.....	16
第十三次研究群會議記錄.....	17
第十四次研究群會議記錄.....	18
第十五次研究群會議記錄.....	19
第十六次研究群會議記錄.....	20
貳、計畫執行統整心得.....	21

## 壹、報告內容

### 第一次研究群會議記錄

- 會議時間：111 年 1 月 24 日（星期一）12:00-14:00
- 主講人：臺東大學段人鳳助理教授
- 講題：以 COVID-19 疫情為例的社群資料分析經驗
- 演講內容摘要：

本次演講以 COVID-19 疫情為例，透過語料分析與「回歸校正」相關之臉書貼文，探討以下兩個問題：比較人工標注和模型預測之間的差異為何？人工標注和模型預測差異最大的類別中，人工標注的語言特徵為何？採取的研究方法分別是：人工情感分類、自然語言處理的預測訓練（pre-training）模型預測情感。透過 CrowdTangle 進行資料搜集。蒐集時間為期 4 日（2021 年 5 月 22-25 日），共 6,917 筆。

「校正回歸」相關貼文的重要性在於，「校正回歸」一詞在 5 月 22 日 CDC 例行記者會上首度由陳時中部長口中說出時，舉國譁然。當天之後，不論是電視上的政論性節目、社群媒體上的意見領袖，或者一般庶民，都出現了許多對於此詞彙的闡述、評論與解讀。顯示了這個詞彙雖然有其既有的學理上的定義，但對於一般民眾而言，卻呈現多元多樣的闡述意義。因此，藉由對於此詞彙的資料搜集、標記與自然語言處理，研究者可以透過在社會生活中呈現多義狀態的詞彙，一窺人工標記（解讀），與自然語言處理「解讀」同一批語料時的差異之處，進一步提出對社群媒體使用者語言實踐和（繁體中文）自然語言處理的調校方式。

## 第二次研究群會議記錄

- 會議時間：111 年 2 月 18 日（星期五）12:00-14:00
- 主講人：政治大學劉吉軒教授
- 講題：跨領域學術對話合作基礎與經驗分享
- 演講內容摘要：

身為政大資訊科學系特聘教授的主講人劉吉軒教授，從個人學術生涯回顧作為例子分享跨領域合作的實際經驗。首先，介紹資訊技術與計算觀點與人文社會科學之間的關係。並以美國電腦科學家如何透過統計分析技術與專精古文的語文學家合作破解了 Copiale cipher 這本屬於歐洲啟蒙時代的十八世紀古書為例，分享跨領域合作的重大意涵。第二，講者介紹人工智慧的發展與 21 世紀的數位世代。除了展示了大數據的計算觀點如何幫助資訊科學建立系統性的連結，亦即從資料到資訊到知識的過程，並強調作為人類語言的書寫模式，文字不僅是表達與記錄的最重要媒介，更是重要的可以提供計算的大數據資料。資訊檢索及資訊擷取、文本分析和文本探勘、自然語言處理和計算語言，以上分別是幾種演算邏輯和知識與技術之結合。而這樣的結合使得機器得以巨觀或微觀閱讀文字，可以縱觀也可以聚焦分析文字。因此講者提醒，「人類在閱讀文字這件事情上，開始有了競爭對手，或是得力助手」。然而，講者認為，創新的方法將是人類不被機器取代的核心，而跨領域的激盪將是創意與創新的來源。

### 第三次研究群會議記錄

- 會議時間：111 年 3 月 25 日（星期五）12:00-13:30
- 主講人：中央研究院江敏華研究員
- 講題：客語研究方法面面觀
- 演講內容摘要：

講者從客語的研究範圍開始介紹客家研究的多重面向。客語研究的範圍主要包含：客語的內在結構、與客語有關的外延知識、客語的跨領域應用研究。內在結構包括：書寫系統、語音、音韻、詞彙（構詞）、語義、語用、語法；外延知識包括：語言與歷史、客語的分布、語言的體系和分類、語言與族群文化、語言與文學、語言與社會。而跨領域應用研究則涵蓋：心理語言學與神經語言學、語料庫語言學、機器翻譯與人工智慧、語音辨識、語音合成、語言治療、計算語言學。日本學者 Hashimoto 將客家研究的歷史自 1957 年以前劃分為三部分：清末民初在字源方面的研究、西方傳教士的貢獻、東方學者的貢獻。1957 年以後的客語研究則受到漢方言研究傳統的影響，重要的學者為楊時逢，著有《臺灣桃園客家方言》、《美濃地區客家方言》。

在與客語有關的外延知識部分，講者特別強調了客語的歷史研究之重要性。歷史語言學是客語的歷史研究重要的參考依據，透過對語言變化的描述和解釋來探究語言變化的成因及方式。其採用的方法為：比較法、內部構擬法。講者強調，客語的歷史研究很難跳脫族群和血統的研究。因此在研究上注重「共時研究」的概念。援引 Joan Bybee 的觀點，講者說明：「所有共時形式都是歷史演變所帶來的歷時變化結果，在解釋語言結構時必須關注歷時過程。」此外，透過將客語語種標記在空間上，觀察其分布可以繪製出語言地圖，這是另外的外延知識的部分，即語言地理學。從中可以觀察到，客語研究的豐富性。

#### 第四次研究群會議記錄

- 會議時間：111 年 4 月 22 日（星期五）12:00-14:00
- 主講人：台灣大學高照明教授
- 講題：Lexical Changes and Variations in Contemporary Chinese: a Corpus-based Approach
- 演講內容摘要：

本次演講講者主要透過語料庫分析來研究漢語的變遷。利用不同的語料庫，不僅可以觀察到當代漢語的詞彙變化和變化，而且可以自動識別它們。主要使用的語料庫有三種：1. 1951 年以來的華文報刊檔案 2. 2009 年以來台灣、香港、中國報紙 3. 來自 TAUS 搜索的多語言語料庫。

此研究分析提出問題有二：一個詞頻率分佈的劇烈變化是否意味著語義變化？能否在語料庫中自動識別單詞的語義變化？

對於如何透過語料庫自動辨識，講者認為，不同的含義有不同的搭配。因此當含義改變時，它們的搭配也會改變。而我們所使用的中文依賴前後文來提取依賴關係，並根據所搜集的語義訊息，對中心名詞進行語義分組。如果一個詞的搭配在給定時間段後發生變化，則該詞可能會發生語義變化。

研究結果發現，台灣和中國不同的詞同義；同一個詞在台灣和中國有不同的意思；同一個詞的不同搭配模式。如：高階智慧型手機（台）、高端智能手機（中）。

## 第五次研究群會議記錄

- 會議時間：111年5月6日（星期五）12:00-14:00
- 主講人：中央研究院馬偉雲副研究員
- 講題：典籍資料處理經驗分享
- 演講內容摘要：

本次演講，講者主要介紹了中研院的漢籍自動化資訊擷取系統、漢籍人工標注輔助工具，同時講解漢籍的詞彙知識本體、如何應用深度學習於漢籍文本分析，並介紹相關的詞庫工具。

漢語自動化資訊擷取系統起初是為了建置「人名權威—人物傳記資料庫」。資料庫建置時所面臨的問題有：一、人工建置雖然精確，卻耗時。二、從大量語料當中自動化建置，雖然省時省力，可以一網打盡卻存在判斷錯誤。因此，提出：自動化建置+人工檢查，可能是最佳策略的假設。然而，隨之而來的挑戰是：僅具備資料庫（人名履歷），缺乏文本標記。Distant Supervision 是研究者提出的解決方案，亦即，透過標記人名、標記官名、標記人名與官名的配對（舉例而言，假設一段文字中出現了人名和官名，且剛好是資料庫的一組配對，就假設這段文字可以描述這組配對）。在此邏輯下，研究者進一步推算出新的三元組：「人名—擔任—官職」，有助於發掘多個先前已建置之人名檔中未出現過的人名，加上電腦輔助判讀有異之履歷資料，研究者發現，此系統不但可以開發新的人名、官名，且有助於補充尚未建置之基層官員資料。此系統解決了人工建置時可能面臨的僅收錄有名的大人物，而缺少小官吏的疏漏。

## 第六次研究群會議記錄

- 會議時間：111 年 6 月 17 日（星期五）12:00-14:00
- 主講人：東海大學蕭季樺副教授
- 講題：台灣直播平台溝通機制：「做自己」言談互動
- 演講內容摘要：

本次演講主要探討台灣直播平台「浪 Live」的直播主和觀眾，如何看待、討論直播主是否「做自己」，亦即，呈現「真實」的那一面給觀眾。此研究聚焦於「打賞直播」：觀眾購買虛擬禮物贈送給直播主。在直播平台中，觀眾想知道他們看的是直播主刻意創造出來的「persona」還是「真實」個性。此研究採取網路民族誌的研究方法，並應用 Andrew Tolson 的媒介言說框架(media discourse framework) 作為主要分析結構。研究者透過兩個案例分析與「做自己」有關的言說後發現，「真實的自我」可以透過不同方式呈現。在 Claire 的例子中，直播主 Claire 透過回應觀眾提出的要求（以生氣的方式說謝謝）自覺，外人眼中偏向撒嬌的語調是她做自己的真實展演。在直播主 Nida 的例子中，她做自己的方式則是展演出被外界認為是誇張的肢體語言（body language），並且是具備連貫性的。她自述，觀眾可以透過不同時間段進入直播平台欣賞她的表演來發現她展演的一致性，因此察覺出她的真實。據此，研究者的第二個發現為：與觀眾的互動是一個判斷直播主是展演真實的個性的關鍵要素。第三，相對於廣義對真實的定義，研究者發現，在「浪 Live」中呈現出一種在地意義的真實，其特徵為直播主線上的展演與線下的生活（性格）具備一致性。第四，研究者發現，兩位直播主都不需要總是迎合觀眾的需求，或者總是積極的與觀眾互動，取而代之的是直播主們忠於自己的感覺。第五，研究者發現「浪 Live」直播平台的設計是以觀眾為中心的（audience-centered）。最後，研究者認為，此研究揭示了真實性的人際關係本質，「浪 Live」的互動功能提供了一個基礎，可以將真實性轉換為社交符擔（social affordance）使直播主與觀眾輕鬆分享生活的時刻，並透過協商反思直播主的真實情況。

## 第七次研究群會議記錄

- 會議時間：111 年 7 月 1 日（星期五）12:00-14:00
- 主講人：國家教育研究院謝承諭研究員
- 講題：YouTube 英語教學網紅言談中的評價語言：多模態的途徑
- 演講內容摘要：

根據 Benson 等學者先前的研究指出：YouTube 作為語言學習與後設語言評論的媒介與平台。引起研究者好奇：英語教學 YouTuber 的流行和影響力。以及他們的言談有什麼特別的？以此為研究動機發想出以下四個研究問題：一、英語教學 YouTuber 語言教學影片中評價語言使用與分布為何？二、英語教學 YouTuber 語言教學影片中評價語言的功能為何？三、英語教學 YouTuber 語言教學影片中與評價語言共同出現的非語言資源為何？四、英語教學 YouTuber 語言教學影片中達成平價的語言與非語言資源如何互動？本研究的研究方法可分為以下五個步驟：一、挑選 YouTuber；二、挑選影片；三、轉寫 YouTuber 語料 / 下載觀眾評論；四、利用 Appraisal Theory (AT) 分析評價語言；五、分析多模態評價。研究者挑選出以下三位 YouTuber：阿滴、哥倫布、Lily Chen。利用 AT 分析，研究者初步發現：一、YouTuber 語言教學的言談的確充滿評價語言 (evaluation language)。二、這些評價語言在教學中有不同的功能 (課程內容、後設評價、言談組織)。三、評價語言以中文為主，但也會使用其他語言。四、除了語言之外，其他如音調、手勢、表情、視線、肢體動作、影片動畫都會用到。尚未完成的研究為：一、YouTuber 言談的細節分析與量化。二、多模態的分析框架。三、觀眾推文的分析。四、推文與影片之間的連結。

## 第八次研究群會議記錄

- 會議時間：111 年 7 月 12 日（星期二）12:00-14:00
- 主講人：政治大學劉慧雯教授
- 講題：很多的「事實上」
- 演講內容摘要：

本場演講者旨在探詢社群媒體使用者在不同的情境、語境、特定事件等，研究者該如何透過語料庫分析來找出相對應的關鍵字。

在每一個社群平台（Facebook、PTT、Twitter）皆有其獨特的特殊語言文化之下，有無可能從社群媒體介面設置查知輿論風向變化？且有無可能不指定事件，仍查知重要議題？

所謂「功能詞」是為是一種功能特徵（functional property）。當一個說話者因為特定目的決定要以一個「殼-內容」的集合狀態（shell-content complex）來使用某個名詞時，這個名詞就成為空殼名詞。而此「殼」常用來作為後面較為複雜細節資訊的整體概括詞。

講者擷取在三級警戒期間，八卦板上推噓文超過 100 之貼文共計 266 則，這些貼文下方推噓文（120,663 則）中，有「事實上」一詞的推噓文計 867 則。分類出三類：複述貼文，製造客觀效果、轉折語氣以反對貼文內容、表達相反意見，不考慮「迴避知識責任」。

## 第九次研究群會議記錄

- 會議時間：111 年 7 月 29 日（星期五）12:00-14:00
- 主講人：政治大學鍾曉芳教授
- 講題：類別隱喻構式研究
- 演講內容摘要：

隱喻的形式非常多元，常常能透過一些關鍵字來搜索隱喻，當隱喻結構為「是一種」結構時，本研究稱之為「類別隱喻」。本研究以語料庫作為研究工具，透過其他關鍵字，觀察除了「是一種」外，更豐富的構式，也將進一步分析類別隱喻中，出現頻率極高、用法固化，且隱喻成分降低的詞彙。期望透過統整類別隱喻構式及固化詞彙幫助未來類別隱喻研究發展。

國科會人文社會科學研究中心

## 第十次研究群會議記錄

- 會議時間：111 年 8 月 8 日（星期一）12:00-14:00
- 主講人：法鼓文理學院王昱鈞助理教授
- 講題：自然語言處理於佛學研究之應用
- 演講內容摘要：

本次演講主要介紹：自然語言處理與數位人文；漢文典音譯詞辨識與音譯詞對擷取；漢文佛典分詞；基於詞嵌入之 CBETA 詞語關係；佛教引用具推薦。

講者認為，自然語言處理對於佛學研究與許多數位人文的問題皆有發揮之處。舉例而言，利用現成的 NLP 技術與工具，能幫助我們解決許多問題。另外，對許多數位人文研究問題而言，不見得要让機器解決全部的研究課題，有時只要能大幅節省人力投入即具相當之貢獻。講者認為，未來需要更多資訊領域與人文領域結合的學者，有以下四點原因。一、數位人文可以為人文學研究帶來不同的視野與機會。二、許多 NLP 模型或方法都需要合適的高品質語料集。三、在模型的設計與資料處理上都需要 Domain Knowledge。四、對於語言的深度理解，還有很長的路要走。

## 第十一次研究群會議紀錄

- 會議時間：111 年 8 月 26 日（星期五）12:00-14:00
- 主講人：臺灣大學呂佳蓉副教授
- 講題：社群上的語言活動：認知語言觀點
- 演講內容摘要：

本次演講主題為社群上的語言活動：認知語言學的觀點。主要探討語言如何轉變以及如何傳播，並以大數據調查輿論呈現的盲點。

社群媒體上的語言基本上在模因理論的基礎之上進行傳播，而演變則是在基模、構式等語言模式下所建構（例如：.jpg）。模因論是解釋文化進化規律的新理論。模因仿照基因而得名，它指一些思想或觀念通過人類文化加以散播，並一代一代地相傳下來。模因實際上指文化基因，它靠模仿傳播而生存，語言是它的載體之一。模因有利於語言的發展，而模因本身也靠語言得以複製和傳播。從模因論的角度看，語言模因揭示了話語流傳和語言傳播的規律。

在模因論中，模因往往被描述為「病毒」(viruses)，它可以感染 (infect) 其他人的大腦或者傳染到其他人的大腦中。一個網路模因存在的形式可以是圖像、超連結、影片、圖片、音樂和網站等。

網路模因可能會保持不變，亦可能會隨著時間推移變化。網路模因會因評論、戲仿、或新聞報導而變化。網路模因可以以很快的速度發展和傳播，有時更加可以在幾天之內散播到世界各地。(黑人問號、被字句)

而由於迷因在現今會被不同的平台近乎永久的保存下來，因此相較過去更為適合透過大數據進行語言研究。

最後結論包含，網路言論型態與傳散之研究需注意常用語言模因 (meme/ 迷因 哏)，因為其帶有渲染力傳播快，也容易成為語言改變的契機。在變動快速、傳播迅速的資訊 (爆炸) 時代，語言改變速度比以前快，schema 以及 construction (構式) 的彈性較容易解釋其被模仿、重製、而改變的現象。

## 第十二次研究群會議記錄

- 會議時間：111 年 9 月 16 日（星期五）12:00-14:00
- 主講人：政治大學紀明德副教授
- 講題：社群媒體視覺化探索工具
- 演講內容摘要：

本次演講講者主要分享社交媒體上的視覺化圖像應用。透過深度學習，能夠讓機器模型針對不同資料進行處理，轉譯成可視化的圖像資訊。

講者也分享如何以視覺化系統探索社群媒體中的使用者行為，包含：如何找出特定議題中的意見領袖、觀察是否有抬轎行為、只回覆特定意見領袖或特定議題的文章的特定個人/團體。

例如，在分析 PTT 發文者帳號和內容上，圖像化的概念包含作者與文章關係以樹圖（Treemap）呈現，並可透過不同的形狀（矩形 or 圓形）表示作者，面積則與其影響力分數成正比，也能透過點擊欲探索的文章以顯示回覆的使用者。

明確案例以觀察使用者帳號「hahaha」在 PTT 八卦版發表的三篇文章為例（2019/4/15 - 2019/5/2）。主要原因是上站次數少，但發表文章少但三篇皆是爆文。而最後將社群媒體上的資料圖像化所帶來的影響，包含能以多視圖的視覺化系統探索社群媒體中的使用者行為，並同時以多個使用者為中心探索使用者活動；而基於熱度圖的視覺化觀察使用者回覆文章的相似度以及回覆時間，且基於二分關係的視覺化亦可以觀察不同社群在議題上的交集程度。

## 第十三次研究群會議記錄

- 會議時間：10月14日（星期五）12:00-13:30
- 主講人：中研院資料所黃瀚萱助理研究員
- 講題：隱喻、情緒判斷與標題生產
- 演講內容摘要：

本次演講主要介紹：自然語言處理（Natural Language Processing）如何被運用在言說分析，以及自然語言處理如何生產隱喻或具有情緒的句子。講者首先介紹自然語言處理的運作邏輯以及基於這個邏輯而帶來的限制，如，處理長篇文章依然具有挑戰性。因此，自然語言通常被運用在短文或文章結構整齊的文體，如，病歷。聚焦於短句，講者表示其適合於自然語言處理的關鍵在於，可以將短句視為序列。若此，以序列對序列的機械化邏輯則可以訓練出，如，google translate 這類的服務。講者介紹了 GPT、GPT3、BART Model、T5（Text-to-Text Transfer Transformer）Model 等語言模型。雖然單一語言模型無法生產出符合人類真實語言脈絡的句子，然而，講者提到，若混用兩種語言模型，讓不同特色的語言模型功能對應句子生產內容需求的展現，則有機會讓句子更貼近真實語言脈絡。舉例而言，可以將擅長閱讀的 BART 結合擅長寫作的 GPT。然而，講者在結語中表示，自然語言生產譬喻句子的最大障礙是，目前仍然得透過人工的方式來檢視生產出來的句子是否符合人類文化語境，尚不存在可以透過標準答案來檢核生產出的句子是否有效。

## 第十四次研究群會議記錄

- 會議時間：11 月 11 日（星期五）12:00-14:00
- 主講人：江玥慧助理教授
- 講題：數位時代的資訊教育與資訊素養
- 演講內容摘要：

江老師從兩個與台灣與資訊相關的教育變革中切入，分別是：「高中端」與「大學端」。在高中端的部分，新制的 108 課綱正式將資訊科技納入國高中必修課程。在大學端，則希望透過程式設計教學普及，來推廣培育科技素養。以此為背景，江老師分享她在教育現場觀察到的大一新生——作為 108 課綱實施後的第一屆大一——於課堂中學習程式語言的情況。江老師發現，雖然 108 課綱制定了許多必修主題，立意良善的希望幫助高中老師教導學生資訊科技，然而，可能伴隨而來的問題是，學生並沒有辦法在有限的時間內全部吸收。因為很多概念在本科系的大學教育裡甚至要用一學年來學習，更何況是一般的高中生。江老師透過以下四種方式：課程定位（為非資訊領域同學設計）、探究式學習、小組任務導向、多元評量方式，以達到幫助非資科相關學系的大學生提升其資訊科技素養。江老師觀察到，非本科系的同學其實在學習基本的資訊科技觀念後，便可以發想出許多本科系同學可能無法想像到的資訊科技題目。舉例而言，土耳其語文學系的同學可以將學習到的程式語言應用在土耳其語的特地語法結構中進行分析。江老師發現，透過範例程式來引導學生完成實作，可以較具效率地達成學習目的。然而，學生脫離範例程式後的統整運用能力，是有待加強的。這也成為大學如何幫助學生提升這方面的資訊科技素養的一個課題。

## 第十五次研究群會議記錄

- 會議時間：12月9日（星期五）12:00-14:00
- 主講人：邱淑怡助理教授
- 講題：資料探勘及深度學習於社群媒體之應用——以臉書為例
- 演講內容摘要：

此次演講內容主要介紹資料探勘及深度學習如何應用於文字資料的分析。透過案例分享，講述如何對社群媒體的貼文資料進行語意分析，利用政治活動及選情期間進行貼文的語意分析，並找出重要的社群媒體的使用者，並透過指標評估建立的 AI 模型。並以「社群媒體語言建構深度學習模型：以『校正回歸』為例」。

模型的演算法主要先將資料進行清洗和分類，先透過中文斷詞系統 CKIP 來找出設定的關鍵詞後，再將之進行標記。延續上述模型分類，並以訊息傳散做為評分標準，從臉書上一則貼文所獲得的按讚、留言、分享等互動行為，來計算出一則貼文的訊息傳播的速度。

目前正訓練社群媒體語言建構深度學習模型。進行中的研究主要先搜集不同的貼文內容之表情符號 emoji，並將 emoji 所代表的意義轉換成文字意思，藉此訓練模型在爬臉書貼文時能否正確預測貼文的語意表達（正面/負面），並且針對更深一層的諷刺語意進行訓練和預測。

## 第十六次研究群會議記錄

- 會議時間：12月16日（星期五）12:00-14:00
- 主講人：國防大學新聞系傅文成教授兼主任
- 講題：資料探勘及深度學習於社群媒體之應用-以臉書為例
- 演講內容摘要：

從當前資料科學方法運用在政治傳播與外交研究概況，並以網路行為模式檢驗威權國家外交策略的可能性。透過 LDA BERT TF-IDF 與 SNA 如何交互應用，來獲得新資料探勘輔助系統簡介。

分析「俄羅斯在 Twitter 的認知作戰策略」，總共搜集了 9 百多萬篇含有 Ukraine 的 post，並把語言翻譯成英文之後，比較不同推文之間的回推文相似性，結果發現有五千多個帳號在一兩分鐘內，並持續三個月內同時發布同樣的 retweet，以此來判斷是否為網軍。而透過 Twitter 的 API 工具，發現在兩萬多個疑似網軍的帳號裡，有 17,676 的帳號位置是為在俄羅斯政府機構，進而計算時間能夠發現 52% 的網軍帳號創立近五個月內。過程中發現的挑戰是，如何在機器學習的過程中，例如不同帳號表述「種族屠殺」的用詞跟角度是否不一樣。

在比較微信和微博的訊息操弄上，微博的帳號具有特別的 pattern。例如在發文角色分析上，一般 KOL 視為紅點，諸多疑似假帳號則為綠點進而回文，引以一般民眾灰點回覆後，紅點和綠點便不會再發文。

PTT 上的認知操作，則藉由探查是否為 VPN 進來發文，並藉由 Phi-phi coefficient 調查共進共出的指標，以及是否在 15 分鐘內以同一個 IP 登入為研究指標。

## 貳、計畫執行統整心得

透過本次科技部補助，使得這個成形中的研究社群得以有固定的聚會時間，並且在聆聽相關研究者的研究報告、心得的過程中，可以將自己獨自做研究時遭遇到的困難、發展中的想法透過討論慢慢形成具體的研究問題。尤其，由於本社群是一個跨領域的社群，主要成員來自傳播學、語言學與資料科學三個領域，演講者也是從這三個領域找到研究成果豐碩的研究者，對於大家打開視野，並且持續練習跨領域對話，是非常重要的經驗。

不僅如此，由於對於單一領域研究者來說，本社群提供至少兩個不同領域的思維，因此社群成員需要持續保持開放的態度與心情，並且嘗試將自己所思所想轉化為其他領域研究者能理解的話語、陳述出來，這對於訓練社群成員向不同領域學者表達意見，是非常有挑戰也非常有建設性的活動。

本學術社群提案的過程中，透過不同領域成員提供該領域重要研究者的歷程，也可使得跨領域研究者掌握最重要的研究議題與研究人才，這對於個別成員擴大學術同儕，有明顯的幫助。不僅如此，由於成員也已開始合作合寫文章，因此在每次報告分享中，都能看到各自在不同研究中的思維變化，有助於大家各自形成對跨領域研究的不同想像。同時，本學術社群中的五位成員，從去年成立學術社群起，即建立較為密切的討論關係，因此也在去年年底提出了科技部（國科會）整合研究計畫「疫意不同：COVID-19 疫情中後的社群閱聽人語言實踐」，五位成員各自的子計畫都獲得兩年期補助，這對於社群來說，是莫大的鼓勵。也因此得以在學術社群活動結束後，持續在整合計畫的操作過程中，繼續以聚會討論的方式分享資料、互相厚實各自的研究成果。

有關社群的執行，召集人有三點建議：

第一、由於學術社群的行政工作頗為繁複，因此召集人花費頗多心力在配合行政作業方面，雖然助理已經完成絕大部分工作，但從學術的角度來看，中心實在不需要以過度的「防弊」思維來看待所有活動。舉例來說，本社群執行期間，受到 111 年 5 月份 COVID-19 疫情突然確診大爆發的影響，多位原本答應來演講的講者，或者成員自身因為家中有老人小孩等易受影響的成員，而對於實體活動有很多顧慮。我想不會有任何一位公共衛生學者、醫師會說這樣的顧慮是錯誤的。中心實在應該體察這種非咎責之因，給予線上參與的便利，而不是一味堅持非要實體到場不可。這不但有違疫情公共衛生的指引，也實在不利於利用傳播科技來達成實質社群溝通活動。國科會既然是國家「科學與技術」的最高指導委員會，如果因為防弊思維而昧於科技便利的現實，實在可能貽笑大方。

第二、一樣是防弊思維，中心嚴格限制召集人的演講或分享不能領取講座鐘點費，也沒有主持費，還要代墊誤餐費等支出，說起來，等於是既不獎勵鼓勵學者擔任召集人，甚至還要用代墊金懲罰他，誰叫他自願召集社群呢？更有甚者，

中心對於助理的防弊心態，更讓召集人為辛勤付出的助理感到不捨，居然還要助理出示召集人交辦工作的信件或通訊記錄。我們不是天生小偷，中心對學術工作者與初入學術領域的學生施以這種父家長式的監督監管監視，令人不敢恭維。

第三、所有學術社群的活動嚴格限制餐費份數，導致某些明明很難得邀請到的學者也無法做較為公開的演講，實在是很浪費資源。我個人在好幾次重要演講中，邀請博士生參與，因為人數超過限制，於是數次自己墊付餐費；明顯又是一個懲罰召集人的惡例。我沒想到一個學者能被一個學術社群的活動懲罰至此，是不是中心暗示我們不要太雞婆、不要想帶領年輕學者、不要想要跟同儕交流、不要把學術資源分享給博士生呢？真心建議中心未來能善待學者，並且避免以行政便利性引導學術活動的思維。

國科會人文社會科學研究中心