

科技部人文社會科學研究中心  
補助學術研究群暨經典研讀班成果報告

社群媒體訊息處理與分析—  
語言科技與資訊流動學術研究群

計畫編號：MOST 107-2420-H-002-007-MY3-SG10912  
MOST 110-2420-H-002-003-MY3-SG11005

執行期間：109年7月1日至110年6月30日

計畫召集人：鍾曉芳教授

執行機構及系所：國立政治大學英國語文學系

中 華 民 國 110 年 7 月 20 日

# 補助學術研究群暨經典研讀班結案報告

## 社群媒體訊息處理與分析— 語言科技與資訊流動學術研究群

計畫編號：MOST 107-2420-H-002-007-MY3-SG10912  
MOST 110-2420-H-002-003-MY3-SG11005

執行期間：109年7月1日至110年6月30日

執行機構及系所：國立政治大學英國語文學系

計畫召集人：鍾曉芳教授

計畫成員：鍾曉芳 英國語文學系教授

劉慧雯 教授兼傳播學院研究部副院長兼新聞系主任

紀明德 資訊科學系副教授

江玥慧 資訊科學系助理教授

許展嘉 通識教育中心副教授

段人鳳 通識教育中心助理教授

施孟賢 通識教育中心兼任助理教授

兼任助理：李季陵、楊惟婷

研讀書籍：（研究群可不必填寫）

中 華 民 國 110 年 7 月 20 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

達成目標

未達成目標（請說明）

說明：

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  未發表之文稿  撰寫中  無

專書： 已出版  尚未出版  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）。

本計畫邀請傳播領域、資訊科技領域與語言學領域學者與研究生共同參與，每月一次召集成員，討論相關研究進程，相互交流研究心得，並且定期將學術成果發表於國內外期刊與會議。本研究群為一跨領域的研究群，因為社群媒體中資料巨量，本研究群深信若是能借助資訊科技的技術，透過數據科學來取得社群媒體上的龐雜訊息，進一步視覺化處理，再透過傳播學領域與語言學領域分析，能夠更深入地研究當代社群媒體的傳播與語言現象；同時整理出重大議題在社群媒體中的傳播途徑，以及「社群媒體」上對於重大議題的論述。「語料庫語言學」研究方法，便是希冀透過大量語言資料的分析，描繪出更全面、完整的語言現象。

身處於網路時代的現代，網路世界裡虛擬「社群」不僅提供人們自由發表言論的平台，也反映社會脈動、影響現實生活，透過研究社群媒體，有助於更佳掌握資訊流動與社會脈動。然而網路媒體即時、大量且複雜，並不如一般語料庫，只需透過簡單的檢索功能便能取得所需語料，因此更凸顯了資訊科技輔助的重要性。透過自動化的資訊探勘，不但能更快速全面地蒐集網路語料，甚至能透過歷時性的角度，能夠觀察分析語言流變之情況。

## 補助學術研究群暨經典研讀班成果彙整表

計畫主持人：鍾曉芳		計畫編號：MOST 110-2420-H-002-003-MY3-SG11005			
計畫名稱：社群媒體訊息處理與分析—語言科技與資訊流動 學術研究群					
		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	4	篇	細項說明於 p.3
		研討會論文	6		細項說明於 p.3
		專書		本	
		專書論文		章	
		其他		篇	
國外	學術性論文	期刊論文	5	篇	細項說明於 p.3
		研討會論文	7		細項說明於 p.4
		專書	3	本	細項說明於 p.4
		專書論文		章	
		其他		篇	
參與計畫人力	本國籍	教授	1	人次	
		副教授	2		
		助理教授	3		
		博士後研究員			
		專任助理	1		
	非本國籍	教授	1		
		副教授			
		助理教授			
		博士後研究員			
		專任助理			
		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。)					



## 一、國內

### 期刊論文

施孟賢、段人鳳、鍾曉芳。已接受 (202107)。「中文新聞文本之宣傳手法標記與分析」(The Analysis and Annotation of Propaganda Techniques in Chinese News Texts)。《中文計算語言學期刊》。(人文學核心期刊-第二級 THCI-Tier 2 Journal)。(修改自 The 32nd Conference on Computational Linguistics and Speech Processing (ROCLING 2020) 海報文章)

王炳勻 (Ben Pin-Yun Wang)、許展嘉 (Chan-Chia Hsu)、龍水水 (Shuishui Long)、丁曉穎 (Xiaoying Liles)。(2020)。「語料驅動學習融入華語課堂之教學設計 (Designing data-driven learning activities for the Chinese as a second language classroom)」。《華語文教學研究》(*Journal of Chinese Language Teaching*)。17 (3) : 103-137。(THCI)。

Lin, Kuo-Chin, Duan-Shin Lee, Shao-An Chien, Yueh-Hui Vanessa Chiang, & Nian-Shing Chen. (2020). The effect of a badminton teaching system using Kinect on badminton skill performance, *Physical Education Journal*, 2020, 53(2):161-178.

林昕柔、鍾曉芳。已接受 (202107)。「忙」的語意特徵及認知概念之語料庫研究——以語料庫語言學及語意側重為架構之研究」(The Semantic Features and Cognitive Concepts of Mang2 'Busy': A Corpus Linguistic Study of Semantic Profiling)。《語文與國際研究》。

### 研討會論文

施孟賢、段人鳳、鍾曉芳 (2020)。中文新聞文本之宣傳手法標記與分析。The 32<sup>nd</sup> Conference on Computational Linguistics and Speech Processing (ROCLING 2020), 台北科技大學, 台灣, 台北。9月24-26日。(原海報文章)

李柏彥、劉慧雯、紀明德 (2021)。英雄所見略同? 社群媒體使用者共現行為之研究。「中華傳播學會 2021 年年會」, 政治大學, 台北。6月25-27日。

劉慧雯 (2020)。臉書上的「反送中」: 臉書編輯貼文分析。「回歸後的中港關係」工作坊, 政治大學國際關係研究中心, 台北。6月12-13日。

施孟賢 (2021)。具詞類標記之客華平行語料庫。第七屆青年學者台灣語言學術研討會, 國立臺灣師範大學, 台北。10月15-16日。

Chiang, Yueh-Hui Vanessa. (2020). An investigation of the usage of “thank” in a Chinese online social network environment. *The 11th International Conference of Digital Archives and Digital Humanities (DADH 2020)*, Taipei, Dec 1-4.

Chiang, Yueh-Hui Vanessa. (2020). Development of a social networking inquiry tool driven by digital humanities research collaboration, *The 11<sup>th</sup> International Conference of Digital Archives and Digital Humanities (DADH 2020)*, Taipei, Dec 1-4.

## 二、國外

### 期刊論文

Chung, Siaw-Fong, Meng-Hsien Shih, Hui-Wen Liu, Chi-ling Lee, & Yueh-Hui Vanessa Chiang. Accepted (202101). The Use of *Hēi diào* ('To Turn Black') and its Related [V *diào*] Forms in Social Media: A Corpus-based Study. *International Review of Pragmatics*. 13(2). (ESCI, Scopus)

- Chung, Siaw-Fong.** Accepted (202002). Emotion Terms in Malay: Patterns in [*meN-*], [*meN- -i*], [*meN- -kan*], and [*-kan*]-only Forms.” *Language and Linguistics*. (SSCI, Scopus, THCI)
- Hsu, Chan-Chia.** (2020). Exploring recurrent frames in written Chinese. *Corpora* 15(3): 291-315.
- Huang, Chu-Ren & **Siaw-Fong Chung.** (2021). Delimiting durative events with manner: A comparative study of two Verbs of rushing in Mandarin. *Cognitive Linguistic Studies*. 8(1): 18-64.
- Lin, Yen-Yu & **Siaw-Fong Chung.** (2021). A Corpus-based Study on Two Near-synonymous Verbs in Academic Journals: PROPOSE and SUGGEST. *English Teaching & Learning*. 45: 189-216. Springer. DOI 10.1007/s42321-020-00072-0 (Scopus, ESCI)

### 研討會論文

- Shih, Meng-Hsien & Siaw-Fong Chung.** (2021). Chinese Semantic Role Sketch based on Sinica Treebank. *Corpus Linguistics International Conference 2021*. University of Limerick, Ireland. July 13-16.
- Shih, Meng-Hsien & Siaw-Fong Chung.** (2021). A Malay Daily News Corpus with Morphological and POS Tagging by the MorphInd Tagger. *Corpus Linguistics International Conference 2021*. University of Limerick, Ireland. July 13-16.
- Yang, Wei-Ting & **Siaw-Fong Chung.** (2021). A Corpus-based Approach to Shell Nouns Analysis in Dcard Social Platform. *The International Conference on Technology-enhanced Language Learning and Teaching & Corpus-based Language Learning and Teaching*, The Education University of Hong Kong, Hong Kong. July 7-9.
- Lee, Chi-Ling & **Siaw-Fong Chung.** (2021). An Analysis of Mandarin Resultative Verb Compound *V-jin* ( 盡 )—A Corpus-based Study. *The International Conference on Technology-enhanced Language Learning and Teaching & Corpus-based Language Learning and Teaching*, The Education University of Hong Kong, Hong Kong. July 7-9.
- Shih, Meng-Hsien & Te-hsin Liu.** (202108). Constructions in Chinese four-character idioms in cross-strait varieties of modern Chinese. *The 11<sup>th</sup> International Conference on Construction Grammar*, University of Antwerp, Antwerp. August 18-20.
- Chung, Siaw-Fong, Meng-Hsien Shih, Yu-Hsiang Shen & Wei-Ting Tseng.** (202010). Metaphoricity Rating of Chinese KIND Metaphor Expressions. *The 34<sup>th</sup> Pacific Asia Conference on Language, Information and Computation*, VNU University of Science, Hanoi, Vietnam. October 24-26.
- Chiang, Yueh-Hui Vanessa, Ying-Zu Lin & Nian-Shing Chen.** (2021). An investigation of group, rater and ratee effects on peer-/self-assessments in a collaborative learning environment in higher education: a cross-classified multilevel analysis. *The 21<sup>th</sup> IEEE International Conference on Advanced Learning Technologies*, Abstract retrieved from <https://tc.computer.org/tclt/icalt2021/>, July 20-21, 144-148.

### 專書論文

- Duann, Ren-feng.** (Accepted 2021). On the Behavior and Event Structure of ‘Eating Steamed Buns Dipped in Human Blood’. In Liu, M., Kit, C., & Su, Q. (eds.) *Chinese Lexical Semantics. Lecture Notes in Artificial Intelligence 12278*, pp. 215-222. Springer.
- Shih, Meng-Hsien, Siaw-Fong Chung, Yu-Hsiang Shen & Heng-Chia Liao.** (Accepted 202010). A Study of KIND Metaphor and Simile Annotation based on Dependency Parsing and ConceptNet. In Liu, M., Kit, C., & Su, Q. (eds.) *Chinese Lexical Semantics. Lecture Notes in Artificial Intelligence 12278*, pp. 525-535. Springer.
- 廖恒佳(Heng-Chia Liao) & 鍾曉芳 ( **Siaw-Fong Chung**). Forthcoming. A Pragmatic Study of ‘JING SHI’ in Social Media (「竟是」在社群媒體中的語用研究).” In Weiqun Wang, Lili Chen, & Shejiao Xu. (eds.) *New Theories, New Approaches and Methodologies, New Issues and New Strategies in the Pedagogy of Teaching Chinese as an International Language (Applied Chinese Language Studies X)*. Selected Papers from the British Chinese Language Teaching Society, the 2018 and 2019 International Conference (pp. 224-230). London: Sinolingua London Ltd. [109-2410-H-004-163-;109-2410-H- 004-032-MY2]

## 摘要

本研究群為一跨領域研究群，希冀透過傳播學門、語言學門以及資訊學門交流合作，深入探討分析「社群媒體」的傳播議題與新興語言現象。隨著資訊科技發展，網際網路越發快速、便利，加上行動裝置普及化，網路使用幾乎無身份限制，大眾能夠隨時隨地上網，自由發表言論、傳遞消息，因此網際網路成為交流訊息最快速便利的平台，網路媒體甚至取代傳統媒體，成為現代大多數人取得新聞、第一手資訊的管道。過去媒體傳播權利僅限於媒體產業中（例如：新聞、電視台），屬於「單向」的資訊流通，大眾對於特定議題的態度與意見較難在傳統媒體中被傳達，然而，網際網路開啟了訊息交流「雙向」管道，讓大眾能夠在網路上傳達一手資訊，或是發表評論。

社群媒體的便利性，使得無時無刻都有使用者上傳文本，透過網頁的紀錄成為語言學門研究的語料，成為語言研究的材料寶庫。網路語言介於口語與書面之間，有其特殊性，而網路語言交流頻繁也導致了語言變化快速，研究網路語言有助於應用語言學與語言理論之分析，而網路傳播之研究也能透過語言表徵之初步分析，來探討傳播議題。透過資訊科技與語言學門合作，除了能大量處理網路社群媒體上的語料外，也能進一步透過計算方式分析語料（如：詞頻分析、中文新聞標記）。

本研究群這一年來共舉辦了十一場聚會，前半年五場、後半年六場，透過與資訊科技專業、傳播媒體專業與語言學專業等學者，共同調閱社群媒體上的大量文本資料，試圖了解傳播領域網路社群的形成過程、動態、目的以及結果。同時，也能幫助語言學者專家取得社群媒體之大量語料，並進一步處理、分析資料，爬梳社群媒體上文本之語意網絡，描繪出更全面、完整的語言現象，為應用語言學與語言理論帶來研究新意。

關鍵詞：社群媒體、語料庫、語意分析、計算語言學、資訊傳播

## 目錄

報告內容.....	1
個別場次之討論題目 .....	1
7/15 媒體文本的語料庫輔助分析：從經濟新聞到旱災報導的實證研究.....	1
9/8 中文抽象名詞研究的現況與挑戰.....	2
9/8 面向AI，尋找人類智能的定位.....	3
10/26 言談篇章中的抽象名詞：立場、評價、與人際互動.....	4
11/30 LEARNING TO GENERATE NEWS HEADLINES WITH MEDIA'S STANCE .....	5
1/19 訊息理解之應用與挑戰.....	6
2/22 構式之無所不在，認知與計算觀點.....	7
3/12 詞彙搭配關係與信息量：以客語類別詞為例 .....	8
4/16 同"鮭"於盡:從壽司郎鮭魚之亂看社群迷因傳播 .....	9
5/14 資料科學方法輔助民意趨勢分析：戰略及戰爭風險感知的網路民意與心理效果 .....	10
6/28 語言中的語意網絡 .....	11
計畫執行心得、收穫.....	12

## 報告內容

### 各別場次之討論題目

本研究群透過跨領域合作，從 109 年 7 月至 110 年 6 月，邀請了 10 位不同領域的專家學者共同參與聚會，分析多元的社群媒體資料，探討社群媒體的傳播模式以及語言使用情形。並以視覺化呈現研究數據，透過語言學與傳播學門理論，分析新興語言以及傳播模式。以下共為十一次聚會的題目與心得。

#### 第一場：109年7月15日

#### 媒體文本的語料庫輔助分析： 從經濟新聞到旱災報導的實證研究

講員：中國文化大學新聞學系 郭文平教授

會議地點：大勇樓 404 室

參與人員：劉慧雯老師、鍾曉芳老師、江玥慧老師、段人鳳老師、許展嘉老師、施孟賢老師、紀明德老師、李季陵助理、蔡依霖博士生、林意玲碩士生、黃曼華碩士生、廖恒佳碩士生、謝挺昊碩士生



郭文平老師本次聚會分享以語料分析為本的新聞研究，以 2014 年至 2015 年的「十年旱災」之報紙報導以及經濟新聞為分析對象，探討旱災、經濟新聞在的媒體再現與論述。郭老師提出語料庫對新聞文本分析的優勢在於，語料庫的統計分析能夠得到一些語言現象，並且這些語言現象是只透過觀察所得不到的，而相較於傳統的文本論述分析，語料庫的量化與質性分析，有更堅實的資料基礎。常見的語料庫分析方法有「詞頻統計分析」、「并置詞分析」以及「共詞句分析」，詞頻統計分析是透過觀察詞彙出現的頻率分析作者的語言使用偏好與新聞文本特徵；并置詞則是透過共現詞分析來觀察詞彙搭配情況；共詞句則是分析脈絡中的關鍵詞（key word in context）。

郭老師的研究工具主要為 Wordsmith 軟體，以旱災的新聞文本來說，旱災有趣的高頻詞有「抗旱」，這個詞彙顯示了一種將旱災是為戰爭的對抗意識。另外，「水庫」與「進帳」在旱災新聞中共現頻率高也顯示了，媒體對水庫的「量化關切」。郭老師的分享相當仔細，鍾曉芳老師表示，語料庫的研究較常見於語言學領域，在新聞領域的應用令人耳目一新，也相信對新聞研究有所助益。劉慧雯老師則提出，語料庫的應用能與大學報的課程結合，加深學生對新聞文本閱讀與撰寫的文字敏銳度，也進一步詢問 wordsmith 工具使用細節。

## 第二場：109年9月8日 中文抽象名詞研究的現況與挑戰

講員：國立臺灣科技大學語言中心 謝承諭助理教授

會議地點：研究大樓 421室

參與人員：鍾蔚文老師、鍾曉芳老師、劉慧雯老師、紀明德老師、江玥慧老師、段人鳳老師、施孟賢老師、李季陵助理、許弘諺博士生、黃曼華碩士生、廖恒佳碩士生、沈鈺翔碩士生、曾瑋庭碩士生、林熙堯碩士生、梁元齡碩士生、李翊綺碩士生



本次的聚會中，謝承諭老師著墨於抽象名詞在學術研究中的現況。抽象名詞在語言學領域中，近年來的研究可以分為三類，分別是「抽象名詞構成的功能詞串」、「個別抽象名詞」與「整體抽象名詞」。抽象名詞的功能詞串例如「事實上」、「問題是」、「這樣子/的樣子」、「結果」與「之類的」；個別抽象名詞例如「意思」、「原因」、「理由」、「緣故」、「人」、「東西」、「事(情)」等；整體抽象名詞例如「X的是」。而研究主要針對抽象名詞的「知識立場」、「評價」、「回應」、「互動性」、「構式」與「字串」等面向進行分析，也著墨於其語序及冠詞等句法研究。

抽象名詞在中文語言學領域中較為新穎，在理論上尚未有標誌性的框架與論述，因此也面臨一些困難與挑戰。在中文裡哪一些詞語屬於抽象名詞？如何透過語料庫或自動化方式找到抽象名詞？如何分析標記與分類？而抽象名詞的功能具體為何亦是語言學者的挑戰之一。

最後，鍾曉芳老師表示，希冀與謝承諭老師及紀明德老師團隊合作，結合近年來最受年輕族群歡迎的社群媒體Dcard中的語料，深入分析社群媒體使用抽象名詞之現象。

此外，郭文平老師和謝承諭老師也對研究群所探討的議題十分關注，後續也經常參加研究群所舉辦的聚會，分享其相關研究成果與心得，為研究群帶來知識與技術的提升，也溢注研究群研究能量。



### 第三場：109年9月8日 面向AI，尋找人類智能的定位

講員：國立政治大學傳播學院 鍾蔚文名譽教授

會議地點：研究大樓 421室

參與人員：鍾蔚文老師、鍾曉芳老師、劉慧雯老師、紀明德老師、江玥慧老師、段人鳳老師、施孟賢老師、李季陵助理、許弘諺博士生、黃曼華碩士生、廖恒佳碩士生、沈鈺翔碩士生、曾瑋庭碩士生、林熙堯碩士生、梁元齡碩士生、李翊綺碩士生



本次聚會中，鍾蔚文老師以歷史角度出發，探討AI與人類智能的關係。人類智能的發展主要是為了使人類適應環境並解決生活問題，以及回應其他動物感官與神經系統行為的發展。鍾老師認為，當今AI人工智能的發展，某種程度上只是5億年前前寒武紀故事的重演，而現代的智能發展進入了一個有趣的轉折點，我們透過電腦、智慧型手機下載智能，但同時也創造了自己的對手。鍾老師將AI定義為認知工具的最新成員，若運用得當也是最純正的認知工具。而AI觸及三個根本議題，一是「智能/知識是什麼？」、「『我』是誰？」以及「智能物種如何競爭或共同演化？」。

人類智能主要透過「思考」來影響活動，而機器只能則是以「符號」、「程式」來運作，而「身體」是智能進行活動的必備條件，AI智能因此經歷了「身體轉向」的過程，「全人原則」表現了「身體」所需具備的元素，「觀看」、「聆聽」、「溫度」等。智能與身體的發展，主要有三個階段，「對焦」、「顯現」與「創造」。以人類智能的角度來看，「學習」是人類「對焦」的方式，透過學習的資訊傳遞來認識環境，並進一步聚焦環境。接著，「顯現」階段則是利用環境工具與資源來解決問題，而對人類智能來說，口語、文字與電腦都是用來解決問題的工具。最後的「創造」階段，人類的生態工程與知識工程皆是創造階段的發揮。

最後，鍾老師以新聞工作為例，提出兩個人機共同演化的策略，一是「分工與互補」，機器負責「監控」、「連結」與「檢索」等功能，而記者則是致力於「人機互動」與「系統設計」。第二是「演化新智能」，創造出更好的認知工具，以及新思維、新的傳播想像。



## 第四場：109年10月26日

### 言談篇章中的抽象名詞：立場、評價、與人際互動

講員：國立臺灣科技大學語言中心 謝承諭助理教授

會議地點：研究大樓 421室

參與人員：謝承諭老師、鍾曉芳老師、劉慧雯老師、紀明德老師、江玥慧老師、許展嘉老師、施孟賢老師、李季陵助理、黃曼華碩士生、廖恒佳碩士生、沈鈺翔碩士生、曾瑋庭碩士生



謝承諭老師在本次聚會中為我們深度認識「抽象名詞」。名詞可以分為抽象與具體名詞，而抽象名詞又可分為特定與非特定，特定名詞例如democracy、linguistics、minimalism等，而非特定名詞例如fact、idea、problem等，而特定與非特定名詞之差異不僅在語意上，語法功能上也略有不同。而抽象名詞也常形成功能詞串，例如英文in fact、the thing is、the fact is，中文的例如「問題是」、「我的意思是」、「事實上」……，而抽象名詞時常反應了語句的立場與評價。

謝老師也曾以抽象名詞的相關研究為基礎，探究社群媒體網路中的語言現象。謝老師觀察在PTT電子佈告欄中男女版的100篇文章，整理出常見的Cyberspace Utterance Final Expression (CUFE)有「～～」、「=\_=」、「科科」等，可以發現社群媒體上的CUFE類別為「非傳統標點符號」、「表情符號」、「狀聲詞」與「括號語」等，這些語言表達方式在功能上皆表達了說話者的立場。另外，在PTT健身版中，謝老師觀察鄉民們的不禮貌留言，研究整理出鄉民常以其專業性與男性氣概之相關言論來表達負面意見。

最後，謝老師分享其「種」字的抽象名詞與立場表達研究，研究顯示「種」常與其他抽象名詞合用，並出現於描述句、評論句與比較句中，而當「種」與問題共現時，表達評價的現象最為明顯，而在社群上也常出現「各種X」的創新表達方式。謝老師希冀未來抽象名詞研究能進一步朝「構式」與語料庫標記與因素對應之方向前進。

109年11月30日

## Learning to Generate News Headlines with Media's Stance

講員：國立政治大學資訊科學系 黃瀚萱助理教授

會議地點：大勇樓 403室

參與人員：黃瀚萱老師、鍾曉芳老師、劉慧雯老師、紀明德老師、江玥慧老師、許展嘉老師、李季陵助理、盧慧娟老師、蔡介立老師、古孟玄老師、林怡瑩博士生、陳岳紘碩士生、黃曼華碩士生、廖恒佳碩士生、楊惟婷碩士生



自然語言處理（NLP）是透過資訊科技方式處理語言的研究，黃老師在本次聚會中，介紹如何以文章內容與數據計算方式，自動化產出新聞標題。由於神經網路（Neural Network）的開發，自動生成新聞標題的技術已經有所斬獲，然而，仍有一些問題待解決，其中包含了如何在生成新聞標題的過程中，帶入新聞媒體的立場，以及如何在新聞文章中擷取適當的數值，成為具有代表意義的標題元素。

關於新聞媒體立場的標題生成，黃老師以三家報社為研究對象，分別是自由時報、聯合報與中國時報，以大量的新聞文章與新聞標題為機器訓練資料，透過統計、機率等方式，計算出適當的訓練模組，產生具有立場的新聞標題。而研究結果可以發現，中國時報的標題與其它報社的標題差異較大。新聞系劉慧雯老師表示，這是可以預見的情況，由於中國時報政治立場鮮明，因此加上新聞立場後，自動化產生的中時新聞標題也凸顯了其文本特色。

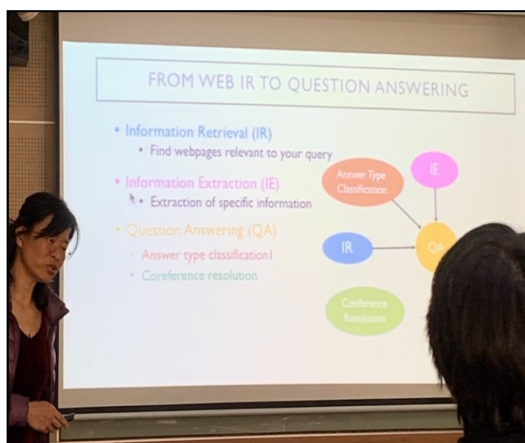
在新聞標題數值的擷取技術上，則需考量原文數值在標題中的呈現方式，在巨量資料中可以發現，標題中的數值常是原文內數值經過「四捨五入」或是「簡化」的結果，因此黃老師的研究透過機器學習與統計分析方式，嘗試建立擷取原文內數值，並進一步轉化為對標題具有價值之數值元素。

## 第六場：110年1月19日 訊息理解之應用與挑戰

講員：國立中央大學資訊工程學系 張嘉惠教授

會議地點：研究大樓 421 室

參與人員：鍾曉芳老師、劉慧雯老師、江玥慧老師、段人鳳老師、許展嘉老師、施孟賢老師、謝承諭老師、鄭家瑜老師、楊惟婷碩士生、廖恆佳碩士生、李季陵碩士生



本次聚會邀請了中央大學資訊工程學系的張嘉惠老師，為我們講解機器學習的操作原理，並介紹 WebETL / DS4NER 兩個搜尋網絡上的資訊、擷取出特定資訊及命名實體的工具。張老師以活動事件搜尋做為應用範例，介紹如何使用這兩個工具建構 EventGo 系統，並且將搜尋到的活動加以分類，提供活動搜尋的服務。另外一個應用則是社群媒體的聲量統計，若想找出含有特定實體(如歌手、家電品牌)的文章，只需要輸入關鍵字，系統就會幫忙找出相對應的文章，並且萃取出比對相符合的貼文。

但是若是文章採用代名詞，這就不是關鍵字比對所能處理的情形。這也就需要交互參照解析(co-reference resolution)，將指示代名詞的部份轉換為其所替代的真正實體，這也就是指代消解任務。目前英文在「指代消解」問題上已有現存工具可以使用，中文的指代消解工具則有中研院 CKIP 可以參考。

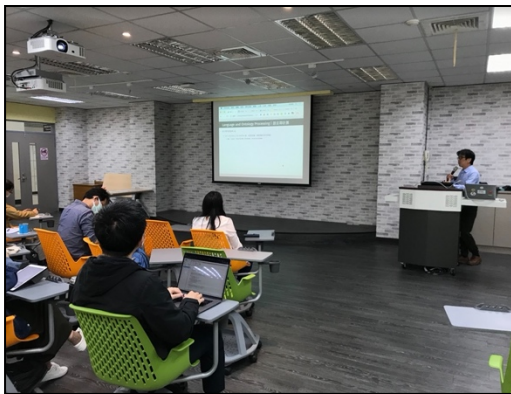


## 第七場：110年2月22日 構式之無所不在，認知與計算觀點

講員：國立臺灣大學 語言學研究所 謝舒凱副教授

會議地點：研究大樓 304 室

參與人員：鍾曉芳老師、劉慧雯老師、紀明德老師、江玥慧老師、許展嘉老師、張瑜芸老師、賴瑤瑛老師、郭文平老師、楊惟婷碩士生、廖恆佳碩士生、李季陵碩士生、黃曼華碩士生、沈鈺翔碩士生、林怡瑩博士生、徐翌珍碩士生、曾昱翔碩士生、張鈺琳碩士生



本次聚會邀請了臺灣大學語言學研究所的謝舒凱老師，謝老師提出語言學是一門廣泛的學問，包含構詞、詞意區分、概念知識本體、變異與變遷、情意處理與社會網路，每一項概念底下都包含著豐富的議題。同時更指出由於時代的變遷，語言研究也應具備時代精神，尤其新時代的網絡發展帶來了社群媒體，催生不同語言版本的民主化、語言與資訊間的關係以及人際關係已跳脫出傳統人際互動，而進入到「人機互動」的時代，在不停變化的情況下，謝老師講解語法中的「構式」能補足傳統語言學理論的不足，作為分析各式語料的工具。

謝老師指出「構式語法」（Construction Grammar）近年來收到語言學界的關注，為語言研究的新趨勢，探討語法及語意之間的互動。援用了認知語言學的理论之外，也更強調語用和對話的因素，希望在解釋人類學習和理解語言的過程中，更貼近真實的樣貌。

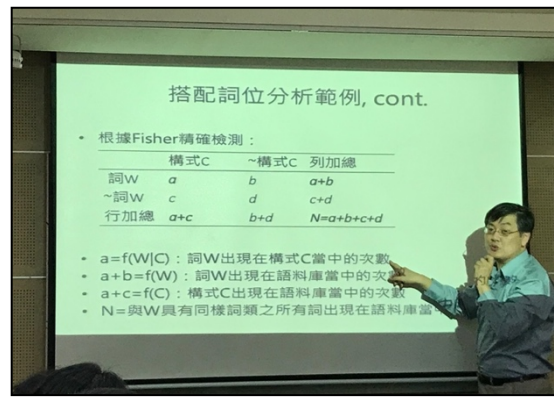
## 第八場：110年3月12日

### 詞彙搭配關係與信息量：以客語類別詞為例

講員：國立清華大學 英語教學系 黃漢君副教授

會議地點：研究大樓 420 室

參與人員：鍾曉芳老師、劉慧雯老師、江玥慧老師、紀明德老師、段人鳳老師、許展嘉老師、施孟賢老師、謝承諭老師、賴惠玲老師、鄭家瑜老師、楊惟婷碩士生、廖恆佳碩士生、黃曼華碩士生、沈鈺翔碩士生、林怡瑩博士生、陳彥宏碩士生、林彥伊碩士生、周俊杰學士生、黃亦晨碩士生、蔡依霖博士生



本次聚會邀請到清華大學英語教學系的黃漢君老師，為研究群講解語言中的搭配關係，由於語言同時亦是一種符號表示，若能分析語言構式中詞位間配搭關係，有助於辨析詞彙語意。

探討詞彙間搭配關係之前，首先必須先釐清什麼是詞？什麼是字？詞之間的相關性為何？這時需要介紹檢測相關性假說的「虛無假設」與「Fisher 精確檢測」，在樣本數偏少的情況下，使用「Fisher 精確檢測」能有效計算出事件的相關性，降低誤判的機率。以英文中的搭配詞為例，根據 Fisher 精確檢測檢驗語料庫中構式出現的次數，能從次數多寡找出強度最強的搭配詞，一一做出排序與比對。

最後，黃老師為了計算類別詞與構式之間的關係，舉了客語中的類別詞說明，首先假設越典型的類別詞M，越常出現在「DET-M-N」的構式中，客語中四個常見的類別詞為「個」、「隻」、「儕」、「位」，使用「Fisher 精確檢測」後發現，「個」檢測成果數值相當高，證實「個」是客語中常用的類別詞。

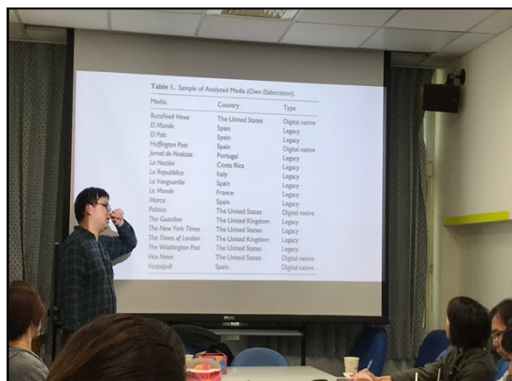
## 第九場：110年4月16日

### 同"鮭"於盡:從壽司郎鮭魚之亂看社群迷因傳播

講員：國立中正大學 傳播學系 劉榮樺助理教授

會議地點：大勇樓 404 室

參與人員：鍾曉芳老師、劉慧雯老師、江玥慧老師、紀明德老師、段人鳳老師、許展嘉老師、楊惟婷碩士生、廖恆佳碩士生、黃曼華碩士生、沈鈺翔碩士生、林怡瑩博士生、陳彥宏碩士生、周俊杰學士生、黃亦晨碩士生、蔡依霖博士生、林瀚軒碩士生、許弘諺博士生



本次聚會邀請到中正大學傳播學系的劉榮樺老師，劉老師所帶來的題目非常有趣而且貼近時事，討論時下社群媒體上正興起的「迷因」，劉老師首先說明迷因的理論基礎，迷因的功能為一種新的複製器（replicators），通常出現在社群媒體上，表達對某些事物的看法，同時也展現了當時的社會現象。劉老師舉了鮭魚的迷因圖為例，網友大量使用成語、諺語創造鮭魚迷因圖（例：同「鮭魚」盡）短時間內在社群平台上製造跟風的效應，網友們紛紛作出更多、更豐富的鮭魚迷因圖響應，並且如同限時動態般短暫分享圖片，帶來轉瞬即逝的效果，這種短時間內大量分享、製圖的行為如同儀式行為一樣，展現出文化動力學。

另外，劉老師也指出研究迷因議題的多元面向，首先是迷因對社群帶來的力量不容小覷，甚至促使人們更改自己的姓名，影響著我們如何做決定；而僅在短時間內出現的迷因圖片竟然能夠帶來巨大的影響力，也是值得關注的研究議題，現在的媒體上出現不少「轉瞬即逝」的資訊分享現象，例如限時動態等等，這些短時間內出現卻又馬上消失的資訊如何引起人們的關注，也相當值得探討。



## 第十場：110年5月14日

### 資料科學方法輔助民意趨勢分析：戰略及戰爭風險感知的網路 民意與心理效果

講員：國立國防大學新聞學系傅文成副教授

會議地點：大勇樓404室

參與人員：鍾曉芳老師、劉慧雯老師、江玥慧老師、紀明德老師、段人鳳老師、邱淑怡老師、楊惟婷碩士生、廖恆佳碩士生、林怡瑩博士生、周俊杰學士生、黃亦晨碩士生、許弘諺博士生



本次聚會邀請到國防大學新聞學系的傅文成老師，傅老師為我們講解了大數據、資料科學與人工智慧的應用方式，經常應用於分析網路上的資訊，例如：照片、文字等等，傅老師以一張充滿汽車的照片為例，使用大數據的分析方式有助於挖掘出藏在表層以下的資訊，因此不僅能辨識出照片中的物體為汽車，同時也能分辨出汽車的車型、車款、自用車還是警車，進一步判斷這張照片的拍攝背景。傅老師也提出了分析社交媒體使用者分佈的理論「冪律分布」，這個理論將每一個社群使用者視為一個節點，而節點之間彼此的聯繫越多、越複雜，可判斷出這個節點的使用者是一名公眾人物、名人或是生產者；相反地，如果節點間的交集不多，則可以推測出這個節點的使用者是一名普通人或是消費者。透過這樣的分析方式，有助於掌握社群媒體上的「離散人」，分析使用者在社群上活躍的情況，也能找出社群中具有高度連結的使用者。

從使用者的行為、連結等等「足跡」，能對他們的人生歷程、政黨傾向、政策態度等做更進一步的判斷，傅老師舉了一個例子，收集使用者的足跡，能判斷他們對於軍人年改等政策的態度，同時也能從中找出網路謠言的散播者，這些資料也進而成為國防部研擬相關政策的重要考量與依據。

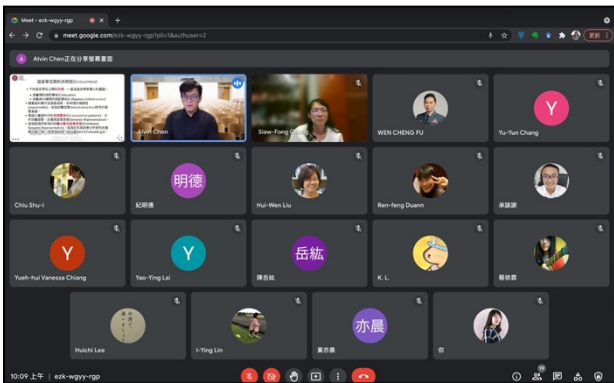
## 第十一場：110年6月28日

### 語言中的語意網絡

講員：國立台灣師範大學 英語學系 陳正賢助理教授

會議地點：線上

參與人員：鍾曉芳老師、傅文成老師、劉慧雯老師、紀明德老師、江玥慧老師、段人鳳老師、許展嘉老師、謝承諭老師、賴瑤鎡老師、張瑜芸老師、邱淑怡老師、李惠琦老師、楊惟婷碩士生、陳岳紘碩士生、蘇冠綸碩士生、蔡依霖博士生、林怡瑩博士生



本次聚會邀請到國立台灣師範大學英語學系的陳正賢老師，陳老師在這次的聚會中講解語言間擁有共現的關係（co-occurrence），不論是詞彙之間的搭配關係（collocation）、還是詞彙與句構之間的搭配關係（collostruction），一直是語言學家所關注的重要議題，而由於語料庫的發展日益成熟，研究字串之間的關聯性也因而成為了解詞彙語意的重要基礎，透過觀察語料庫中詞彙的共現關係、再分析語意，有助於得出語意表徵（semantic representation），因此詞彙與句構之間彼此就像一張複雜交織的網，相互連結、影響。

陳老師舉了中文的「空間句構」說明，用中文表達空間時，常常使用類似的句構「在+物體+空間方位詞」，像是：在期限內、在條件下...等，這樣的空間句構其實是非常抽象的，很難具體表達出他的意義，但是陳老師認為只要透過詞彙與句構的共現，即可一窺語意的端倪。首先需要找出詞彙與句構間的連結，哪些詞彙經常與哪些句構共同出現？下一步即是尋找詞彙與詞彙間的關係，處理意義相近的詞彙，第三步則是尋找句構與句構之間的連結，是否有相似的句構常用來表達相似的語意。透過這三種連結，及有助於我們找出語法知識網絡間的連結，建構出句構的語意網絡。本次為研究群首次辦理線上聚會，少了空間的限制，吸引其他學校對此議題感興趣的老師前來參與討論，包含成大李惠琦老師、國防大學傅文成老師，共同為研究群增加學術研究能量。



## 計畫執行心得、收穫

社群媒體的出現，影響了人類使用語言的方式，社群媒體成為人們溝通交流的主要平台，除了新興詞彙大量派生、變化快速，網路語言風格也較現實生活中談諧甚至諷刺。另外，網路讓一般民眾能夠透過「留言」或「迷因」的方式參與事件，雙向的傳播使得網路篇章結構更為複雜，如欲瞭解特定網友留言之意義，依賴讀者對文章中事件的深入理解，並具有足夠背景脈絡知識，這些現象使得以社群媒體為對象的語言研究更有新意與可看性。除了進一步深究網路文本，分析各式文本了「來源」使用者也成為當前的主要議題，由於社群媒體上資訊流動率又快又廣，許多使用者利用這項特色傳佈謠言或不實消息，透過分析、歸納這些不實消息的語言特色，有助於判斷資訊真假。使用「語料庫語言學」研究方法，便是希冀透過大量語言資料的分析，描繪出更全面、完整的語言現象。

然而網路媒體即時、大量且複雜，若能透過簡單的檢索功能取得所需語料，便能凸顯資訊科技輔助的重要性。透過自動化的資訊探勘，不但能更快速全面地蒐集網路語料，甚至能透過歷時性的角度，能夠觀察分析語言流變之情況。因此本計畫定期邀請傳播領域、資訊科技領域與語言學領域學者與研究生共同參與，討論相關研究進程，相互交流研究心得，跨領域的研究計畫也是本研究群主要的執行活動之一，透過社群媒體傳播研究的基礎，使用不同的角度切入，探討網路語言現象的產生及特徵，幫助語言學者更完整、更深入的分析語言傳播背景脈絡。另一方面，語言學角度的切入社群媒體之傳播研究，能拆解文本中的語言結構，分析特定構式、詞彙結構產生的語意，能使傳播的研究更為細緻生動。資訊科技學門除了提供自動化資料處理工具，有效抓取網路上的資料之外，也使用視覺化的數據呈現，也幫助研究能更多元的展示研究結果。