・新聞學研究・　第一六九期
2026 年 10 月　頁 55-110　研究論文

華裔認同協商的爭吵劇場：基於 YouTube 評論語料庫的論述分析*

柯鐘晏**

投稿日期：2025 年 12 月 22 日；接受刊登日期：2026 年 4 月 15 日。

* 本文作者感謝二位匿名審查人的寶貴建議，特此致謝。

** 柯鐘晏任職於華僑大學新聞與傳播學院，ORCID: https://orcid.org/0009-0004-6065-7749，e-mail: duress.ko@gmail.com。

本文引述格式：
柯鐘晏（2026）。〈華裔認同協商的爭吵劇場：基於 YouTube 評論語料庫的論述分析〉，《新聞學研究》，169，55-110。https://doi.org/10.30386/MCR.202610.0016

《摘要》

網路防火長城的隔離，導致中國境內用戶在全球主流網站上缺席；而這個結構因素，阻礙了作為整體的全球華裔社群的線上認同協商活動。然而，在 2021 年發表，諷刺激進中國國族主義網民「小粉紅」的 YouTube 音樂影片〈玻璃心〉，意外引發大量（自稱）中國用戶冒險翻牆參與評論，與全球其他用戶展開激烈的華裔身分認同協商。

本研究將〈玻璃心〉與其 209,744 則評論，視為一個原生線上媒介事件與資料邊界明確的全量語料庫，透過語料庫輔助的論述分析（Corpus-Assisted Discourse Studies, CADS），針對其中華裔身分協商的各種論述進行系統分析，並提供方便理解這些論述之間關係的地圖。

研究結果顯示，〈玻璃心〉評論是一個爭奪彼此華裔相關身分邊界的即興爭吵劇場，參與的草根主角們透過各種語藝實踐，宣告自我華裔相關身分，並且爭論這些集體身分的實際內涵。而這些爭吵的焦點，是一套強調漢人血緣、華語、中文與中共黨國政治意識形態一體化的中國人身分敘事。無論是認同或拒斥，都是透過回應這套敘事的主張，使彼此身分的實質邊界浮現。

關鍵詞：大語言模型輔助內容分析、身分協商、原生數位資料、媒介事件、語料庫輔助論述分析、諸華裔社群

壹、研究背景與目的

全球主流網站為分散各地的各族裔（ethnicities）社群提供了一個能夠協商彼此認同的媒介場域。然而對諸華裔社群（ethnic Chinese communities）而言，這個場域有一個特殊限制：防火長城（Great Firewall, GFW）阻擋了大部分主流網站進入中國，[1] 使得占全球華裔人口絕大多數的中國華裔社群，在這類全球社交平臺上結構性缺席，而這個場域所中介的華裔認同協商因此呈現不對稱狀態。然而，某些在主流網站上引爆的華裔認同協商事件，居然吸引了許多中國用戶冒險「翻牆」參與爭論。[2] 考慮到翻牆行為在中國可能面臨的後果，這種「願意承擔風險也要發聲」的現象，不僅標示出這些事件觸及了參與者認為值得冒險表達的核心議題，更提供了一個難得的機會，讓研究者能夠觀察到原本結構性缺席的中國華裔社群，如何與海外社群進行認同協商。本文在此採用協商一詞的本義，不只是充滿溫情的共識追求，還包括涉及對抗與爭奪的權力博弈。

2021 年 10 月，由華裔馬來西亞歌手黃明志（Namewee）與華裔澳大利亞歌手陳芳語（Kimberley Chen）首發在 YouTube 上，諷刺被稱作「小粉紅」的激進中國國族主義網民的官話（Mandarin）[3] 流行歌曲

[1] 防火長城是中國為規範國內網路而實施的政策與技術組合，其主要功能是阻止防火長城內的網民近用中國政府所不允許的網路服務，包括 YouTube 等全球主流社交媒介（Clayton et al., 2006）。

[2] 「翻牆」是中國網路用語，意思是中國網民繞開俗稱「牆」的防火長城限制，取用牆外的網路服務。

[3] 官話是明清帝國時代對於官員語言的稱呼，經葡萄牙譯為 Mandarim，再傳入英文成為 Mandarin，即「官員的語言」。進入中華民國後，政府決定將北京官話標準化並稱作「國語」，中華人民共和國改稱為「普通話」（Tam, 2020）。早年新馬

MV〈玻璃心〉（Fragile），在一個月內就獲得超過 3 千萬次播放與 20 萬則評論（Namewee，2021 年 10 月 15 日）。這些評論顯示出語言與文字的多樣性：多以漢字書寫，但涵蓋了不同的語言使用情境，包括官話、粵語、臺語、日語，以及使用英文、馬來文等非漢字系統書寫的評論。這種多語言、多文字的參與，結合評論中用戶自我揭露的地理位置與身分認同，如「我在台灣」、「身為大馬華人」、「翻牆來看」等，顯示此事件確實吸引了來自不同地區、不同背景的參與者，不僅包括海外華裔、自述翻牆參與的中國用戶，也包含非華裔評論者。

更重要的是，這些評論所呈現的討論內容，大量圍繞在對於華裔身分的反身思考周圍展開：我是誰？我們是誰？我們與其他華裔社群是什麼關係？由於〈玻璃心〉是身為海外華人的黃明志對「小粉紅」的諷刺，這些被它激發的論述不只觸及文化面向，更觸及國族效忠的政治面向。〈玻璃心〉在 YouTube 上引發的大規模討論，為研究者提供了一個獨特的窗口，得以用來觀察來自不同族裔背景、甚至冒險翻牆的大量參與者，如何在此事件中進行身分協商。而這些評論作為一個邊界明確的全量語料庫（census corpus），使得對這些協商論述的系統分析成為可能。而本文的研究目的，即在呈現〈玻璃心〉事件中諸華裔社群的認同協商論述，並提供方便理解這些論述之間關係的地圖。

貳、文獻回顧

身分／認同（identity）是通常被歸類到某個名稱（name）下，用來

中文教育原以方言私塾為主，後來逐漸轉型為現代華校。戰後隨著國家獨立與教育在地化，為了區隔國家語言，將「國語」改稱為「華語」，並逐漸在海外華人社群之間通行（庄妙菁，2005）。

定義個人或群體的一組性質。根據社會建構論（social constructivism），這些性質並非不變的本質，而是在具體的社會互動，包含直接或透過媒介中介的社會互動中，不斷被塑造、再生與動員（Cerulo, 1997, pp. 387-391）。身分的發生是目標導向的，使人將聚焦在當前溝通脈絡中的某些性質而非另一些性質上（Leary & Tangney, 2012, p. 69），「具有某個身分」是指某人或群體透過協商而被塑造成具有相關性質的類別（Antaki & Widdicombe, 1998, p. 3）。由於身分現象涉及從個體到社會的多個層次，不同學科發展出不同的研究取徑。心理學取徑關注個體認知與情感過程，微觀社會學取徑聚焦互動情境中的身分協商，而巨觀社會學取徑則探討媒介、文化與社會結構如何型塑集體身分（Antaki & Widdicombe, 1998; Cerulo, 1997; Leary & Tangney, 2012）。本研究採取微觀與巨觀之間的取徑，以〈玻璃心〉為案例，聚焦於華裔身分如何在 YouTube 用戶之間被協商。為此，以下回顧媒介如何中介集體身分中「族裔與國族」建構的相關研究與理論基礎，再檢視媒介化華裔身分（mediated ethnic Chinese identities）的既有研究及限制。

在媒介所中介的集體身分研究中，Anderson（1983/2006）關於印刷資本主義所創造出來穩定、匿名且共時的活動，如何將前國族的各種社群，包括宗教與王朝社群，與作為殖民政府統治分類的族裔，轉變為國族認同的分析，奠定了相關研究的基礎。延續 Anderson 關於媒介使「想像的社群／共同體」得以型塑的理論洞見，後續研究將焦點轉向電子媒介，包括廣播與國族（Hayes, 2000）、電視與國族（Hutchings & Tolz, 2015）的關係。

隨著媒介技術從印刷、廣播到電視等大眾媒介，轉向以網路為基礎的分散式媒介，研究所關注的社群，從國族延伸至跨地域、跨文化社群，但以國族為核心的研究傳統仍繼續發展，包括離散（diaspora）、

種族、族裔的相關研究。例如研究早期網站與論壇，如何成為一把兩面刃，在維繫國內少數族裔的國族認同的同時，也創造了全球化效果，從而削弱了他們的國族認同（Saunders, 2010）；研究網路如何以其社群機制與身分協商的不同方式，創造出不同於傳統離散社群的數位離散社群，包括數位強化與數位原生的離散社群（Brinkerhoff, 2009）；研究演算法，包括社群媒介的推送演算法，如何系統性地強化不平等，尤其是種族隔離，並同時具有反抗不平等隔離的潛力（Benjamin, 2019）。[4]

綜上所述，媒介技術的演變持續重塑集體身分的不同建構方式。從印刷到廣播電視創造的是 Anderson 所謂「想像的社群／共同體」，而網路技術則使中介化人際互動（mediated interpersonal interaction）成為可能，讓人們得以在網路空間中直接協商彼此認同。與此同時，演算法也在網路空間中系統地塑造身分之間的隔離。對於在物理空間中隔離的跨國族裔社群而言，這意味著網路空間既可以是跨越邊界的認同協商場域，也可以是被演算法切割的身分隔離空間。本文以下檢視近年來關於華裔線上認同的經驗研究，探討既有研究如何應對這種既統合又隔離（包括既有的物理隔離與演算法隔離）的雙重性。

隨著網際網路成為各跨國族裔社群建構與協商認同的重要場域，關於華裔身分的線上協商研究也持續累積。這些研究使用的方法可分為三類：

第一，檢驗網路媒介使用與認同之間關係的量化調查。如「社交媒介使用」與「香港人的單一或雙重認同（香港人與中國人）」之間關係的調查研究（Kobayashi, 2020）。

第二，深入理解認同協商實踐的深度訪談與數位民族誌等質化取徑。如

[4] Benjamin（2019）全面性探討了演算法如何強化不平等隔離，社交媒介演算法只是其中一種，但也是最重要的一種。

對在日本的中國網紅進行的訪談與數位民族誌（Yan Du, 2025），對參與線上公民運動的紐西蘭華裔的深度訪談（Yu Du, 2025），針對中國與海外諸華裔社群用戶的深度訪談與數位民族誌，探究社交媒介平臺 VRChat 中「華人性」的認同協商（Leo-Liu & Wu-Ouyang, 2025）。

第三，內容分析或資料探勘，用以從鉅量線上文本中識別認同協商的論述模式或網絡結構，如利用微博（Weibo）語料分析用戶如何討論歸化中國運動員的國族血統與文化認同問題（Li et al., 2024）；利用微博語料指出中國產遊戲在中國玩家之間激發的文化與國族認同討論（Gao et al., 2025）；透過主題標籤（hashtag）的網絡分析，指出推特簡體中文圈中，中國國族主義論述的碎片化與極化（Zeng & Fang, 2025）。

在上述取徑中，不同方法在應對網路空間既統合又隔離的雙重性時，各有其優點與限制。首先，調查與訪談法雖然成熟可靠，然而它們一方面依賴受訪者的事後生產資料，這些資料本來就遠不如線上原生資料能夠直接捕捉線上認同協商複雜且動態的過程，另一方面更面臨非機率抽樣的代表性（訪談法）與母體範圍設定（調查法）問題，尤其後者使得調查法通常只被用在特定物理空間範圍的社群之間，如香港（Kobayashi, 2020）。其次，數位民族誌雖然使用數位原生資料，但同樣有非機率抽樣的問題，研究者選擇進入特定網路空間，如 VRChat 追蹤特定參與者（Leo-Liu & Wu-Ouyang, 2025），不得不受限於所設定的研究時段中偶發的認同協商事件。其三，線上內容分析或資料探勘雖然能透過處理鉅量資料避開抽樣問題，還能聚焦於特定認同協商事件，如中國製全球頂級遊戲激發的線上討論（Gao et al., 2025），或是中國主辦奧運期間對於歸化中國運動員的討論（Li et al., 2024），但這類方法多

採用關鍵詞導向的資料聚焦方法，這種演算法會遺漏未使用特定關鍵詞但仍參與認同協商的討論。

本研究採用全量語料庫（census corpus）取徑，蒐集〈玻璃心〉事件的 20 萬則評論，避開了上述方法論限制：（1）使用數位原生資料，直接捕捉協商過程；（2）作為母體而非樣本，無抽樣代表性問題；（3）收集完整事件所有評論，而非基於特定關鍵詞的篩選。

此外，〈玻璃心〉事件提供了一個在當前媒介環境中罕見的觀察機會。正如 Benjamin（2019）對於演算法強化種族隔離的提醒，防火長城將不同華裔社群隔離在不同平臺中，牆外平臺（如 YouTube 以及原名 Twitter 的 X）缺乏一般中國用戶參與，而牆內平臺（如微博、微信）雖然通常允許牆外用戶使用，但這些平臺的審查機制可能影響討論的完整與真實，像是敏感言論會被屏蔽，參與者會自我審查（參見 Yang, 2022）。〈玻璃心〉事件則提供了獨特條件：大量（自稱）中國網民冒險翻牆到 YouTube 參與討論，使原本結構性缺席的中國用戶，與海外用戶罕見地在同一牆外平臺上進行華裔身分的認同協商。在相對不受中國審查機制限制的 YouTube 環境中，不同社群得以更自由地表達立場。結合全量語料庫取徑，本研究得以系統、完整地觀察這個跨社群認同協商的動態過程，填補既有研究在方法論與研究場域上的雙重缺口。

參、研究問題

本文研究旨在以〈玻璃心〉的 YouTube 評論作為全量語料庫進行分析，呈現其中諸華裔社群的認同協商論述。在提出具體研究問題前，本節將首先說明以〈玻璃心〉的 YouTube 音樂影片與其評論作為「原生線上媒介事件」及其資料的理論與方法論意義，以確立研究問題的合理性。

本文所謂「原生線上媒介事件」是兩個理論概念的綜合：「媒介事件」與「原生數位資料」，首先說明什麼是媒介事件。

媒介事件的原始定義來自 Dayan & Katz（1992），他們以菁英設計的電視典禮為典型，將媒介事件定義為具有「劇本」與「事先準備」的媒介化事件。然而他們也承認，某些涉及集體主角（collective protagonists）的草根、自發事件，難以歸入典型媒介事件中。在後續理論發展中，Couldry（2000）把這類草根媒介事件，理論化為媒介化儀式（mediatized ritual）。他進一步擴充原本媒介事件的理論範疇，將以「衝突與爭吵」作為象徵行動的事件也納入其中，這類事件透過戲劇化的方式爭奪社會邊界，從而展演其儀式特徵。

綜合並參考 Sumiala（2012）對前述理論家工作的梳理與詮釋，本文將這類草根自發的衝突媒介事件視為一種「即興劇場」，意指雖無剛性劇本，卻仍具備角色、場景與展演的結構，其中包括參與者基於集體認同立場的情感投入（角色）、基於媒介的技術能供性（affordance）所框架的場所（場景）、事件中具有協商或跨越邊界的象徵意義的行動（展演）。進一步說，線上媒介事件的經驗研究，往往偏好本文稱作「即興爭吵劇場」的研究主題，例如 Robinson & Enli（2022）關於瑞典大選，與 Jerslev（2022）關於美國總統就職典禮的研究，都強調線上公眾對官方活動權威意義的詮釋能動性以及顛覆與重塑論述的語藝實踐。

本文即從此理論與經驗研究脈絡出發，將〈玻璃心〉評論中的華裔認同協商，視作一個即興爭吵劇場的媒介事件。而較為不同的是，先前這類線上媒介事件的經驗研究，多屬於線下官方活動在網路上引發的衍生現象。相較之下本文選取的〈玻璃心〉評論，屬於「原生線上媒介事件」，其生成與爭吵都發生在 YouTube 平臺，形成一個相對完整的事件單位。這種「原生」的技術能供性，不僅界定了事件本身的邊界，也決定了資料的性質。接下來本文將基於對「原生數位資料」的相關理論探

討，說明 YouTube 評論作為經驗研究資料的方法論價值。

原生數位資料（natively digital data）指生來即以數位形式存在的資料，有別於將類比資料轉換為數位形式的數位化（digitized）資料；原生數位資料理論上又可區別為用戶生產的內容，與演算法生產的後設資料（metadata），雖然在實踐中兩者相互糾纏，亦即演算法會根據用戶行為產生後設資料，並根據後設資料配置內容的資料結構與流通（Lupton, 2015; Rogers, 2015; Salganik, 2017）。雖然有研究者警告，演算法會扭曲線上社會現象，使研究者難以評估其中無偏見的因果關係（Salganik, 2017），但也有研究者支持，演算法與用戶行為的糾纏本身就是研究重點，只要研究設計得當即可，例如使用無痕瀏覽模式，減少演算法維度（Rogers, 2015）。還有研究者基於 Latour 的行動者網絡方法論，主張將這種糾纏現象的研究，從單純線上擴充到線上與線下集合體的研究，認為單純的線上資料分析不足以捕捉這種糾纏現象的複雜性，建議透過訪談法或數位民族誌所得到的「厚數據」（thick data）來彌補純粹「大數據」（big data）的不足（Lupton, 2015）。但耐人尋味的是，Latour 本人與其團隊反倒示範了如何能夠單純透過鉅量線上資料（如 Web2.0 資料）的分析，來超越傳統社會學理論中「結構與能動斷裂」的問題（Latour et al., 2012）。[5]

綜上所述，單一 YouTube 帳號下的鉅量評論，反映其用戶行為與演算法的糾纏，包括事件的生成、互動、意義協商，構成一個邊界明確、高度可觀察的網絡，因此成為線上媒介事件的理想觀察場域。〈玻璃

[5] Latour 與其合作者基於 Gabriel Tarde 的「單子」（monads）理論主張，新出現的數位資料集使得研究者得以透過追蹤和部署行動者的「檔案」（profiles）或其「行動者網絡」（actor-network），來檢驗他們提出的「單層」（1-LS）方法論，從而超越傳統社會學區分「行動者／個體」與「結構／社會」，而兩者之間又難以跨越的二層（2-LS）（微觀或巨觀社會學）方法論（Latour et al., 2012）。

心〉的二十萬則評論使得研究者能夠無須借助外部資料，即能觀察到多元的華裔認同論述如何被表述、爭吵與再組織。

為檢驗〈玻璃心〉的二十萬則評論，確實構成一個相對完整且封閉的「華裔認同即興爭吵劇場」，本文首先以詞頻分析進行初步觀察，以揭示其論述焦點。在去除數字、標點與介詞後，前 50 高詞頻詞彙，依序排列如下表 1。

表 1：語料庫中前 50 高頻詞彙

#	詞彙	詞頻	#	詞彙	詞頻	#	詞彙	詞頻	#	詞彙	詞頻
01	你[a]	68,377	14	user	10,801	27	还	6,453	40	知道	5,044
02	我[a]	51,454	15	小粉紅[a]	9,386	28	没有	6,368	41	大陆[a]	5,037
03	人	24,488	16	聽	9,194	29	沒有	6,315	42	年	5,020
04	歌	18,154	17	台灣[a]	9,179	30	我們[a]	6,263	43	吃	4,870
05	首	14,368	18	你们[a]	8,387	31	来	6,139	44	很多	4,837
06	自己	13,961	19	想	7,911	32	听	5,786	45	个	4,808
07	中國[a]	13,754	20	中共[a]	7,305	33	没	5,766	46	洪水	4,726
08	看	13,101	21	台湾[a]	7,186	34	碎	5,683	47	一个	4,693
09	玻璃心	12,761	22	哈哈哈	7,042	35	做	5,657	48	自由	4,670
10	说	12,686	23	沒	6,729	36	我们[a]	5,619	49	爆炸	4,662
11	你們[a]	12,250	24	万	6,699	37	会	5,537	50	国家	4,636
12	中国[a]	11,896	25	the	6,668	38	中國人[a]	5,325	—	—	—
13	說	11,580	26	用	6,634	39	讓	5,144	—	—	—

註：詞彙依順序排列，[a]標示人稱代詞與華裔身分相關詞彙。
資料來源：本研究彙整。

初步分析結果顯示，最高頻的詞彙主要分為三類：人稱代詞（我、你、我們、你們）、諸華裔相關的政治實體與地理身分（中國／中国、中國人、中共、大陸、台灣／台湾）、政治意識形態（自由）與相關的諷刺性集體標籤（小粉紅）。

由於人稱代詞是說話者用來表達身分認同與社會關係的重要語言指標（Davies & Harré, 1990/1999），本研究參照相關研究中利用自我身分宣告所使用的人稱代詞作為身分標記（identity marker），檢索語料庫中集體身分論述的作法（Davis et al., 2019），在語料庫中以「我是」作為檢索詞進行共現分析，前 50 高機率（以 Z-Score 為準）詞彙，依順序排列如表 2。

表 2：在語料庫中與檢索詞「我是」共現的前 50 個詞彙

#	共現詞	Z-Score	#	共現詞	Z-Score	#	共現詞	Z-Score
01	我	20.112	18	你們	8.557	35	支持	6.516
02	人	20.011	19	大陆 [a]	8.201	36	自己	6.479
03	你	17.409	20	歌	7.960	37	越南人 [a]	6.479
04	中國台灣省 [a]	14.211	21	不知道	7.650	38	一个	6.428
05	说	12.416	22	覺得	7.584	39	台灣 [a]	6.377
06	說	12.341	23	玻璃心	7.559	40	华人 [a]	6.304
07	大陆人 [a]	11.780	24	觉得	7.451	41	馬來人 [a]	6.162
08	中國人 [a]	11.323	25	想	7.406	42	首	6.158
09	中国人 [a]	10.976	26	知道	7.229	43	小粉紅 [a]	6.135
10	user	10.219	27	蒙古人 [a]	6.998	44	中國海南省 [a]	5.999
11	台灣人 [a]	9.931	28	沒	6.930	45	听	5.999
12	看	9.853	29	香港人 [a]	6.915	46	来	5.844
13	說	9.835	30	但是	6.884	47	句	5.840
14	中國 [a]	9.327	31	爲	6.836	48	看到	5.794
15	高山族 [a]	9.110	32	你们	6.766	49	沒有	5.780
16	马来西亚人 [a]	9.051	33	个	6.656	50	出生	5.752
17	小粉红 [a]	8.739	34	中国 [a]	6.654	—	—	—

註：[a]標示華裔身分相關詞彙。
資料來源：本研究彙整。

分析結果顯示，與華裔身分相關的政治實體、族裔、意識形態與地域詞彙大量出現。基於「自我身分宣告」（如「我是台灣人」）與「自

我包括語言」（self-inclusive language，如「我們台灣人」）作為指標，能有效從文本中識別出正在展現其集體身分的論述（Davis et al., 2019），[6] 表 2 的分析結果可以初步詮釋為：語料庫中用戶所自我宣告的集體身分與華裔身分高度相關。

綜合以上初步的詞頻與共現分析結果，可以看出〈玻璃心〉爭吵劇場中的互動，往往以用於自我身分宣告的人稱代詞與華裔相關身分標記為核心。為進一步描述與理解這些身分標記如何開展為華裔認同協商的論述，本文提出以下研究問題：

RQ：〈玻璃心〉評論中出現了哪些具統計代表性的華裔相關身分指涉模式，而參與者又如何透過這些模式，在爭吵劇場中爭奪身分認同邊界？

肆、研究方法

為了回答研究問題中對「統計代表性」的要求，本研究需要採用一種能具統計意義地橋接語料庫中統計模式與論述詮釋的研究方法，以下說明之。

語料庫輔助的論述分析（Corpus-Assisted Discourse Studies, CADS）是使用貯存在電腦中，包含大量自然生成的語言資料的語料庫，並使用處理語料庫的程式工具來進行語言研究的方法，其目標在揭示語言模

[6] 此結論是根據該研究（Davis et al., 2019）方法論與發現所推論得出。該研究採用「自我命名」與「自我包括語言」作為集體身分（素食者）標記，對資料進行保守的編碼標準篩選，以確保只納入身分明確的集體身分。透過這些標記篩選出的樣本，成功支持了所有身分理論假設，從而間接證實該指標在識別並操作化理論預期的集體社會身分方面具有有效性。

式，從而理解語言被用於建構論述的方式。相較於傳統論述分析，CADS 在方法論上的優勢在於，透過語料庫所提供具統計意義的鉅量證據來支持所發現的論述模式，有效限制研究者的認知偏差；並且允許研究者採用混合研究方法進行三角驗證，即使研究者只進行單一文本的質化研究也可以利用大型參照語料庫作為基準，快速且有效地檢驗其對論述模式的猜測（Baker, 2006）。

然而，就本研究要求的「統計代表性」而言，既有 CADS 有一個方法論弱點。語料庫分析本質上只能得到非語義的詞頻統計模式。而在實務操作上，研究者會透過關鍵詞脈絡（Keywords in Context, KWIC）挑選樣本，來檢驗統計模式背後的實際語義，然後基於這些樣本對所發現的模式進行論述詮釋。但這種檢驗往往不包含被挑出來的 KWIC 樣本的統計代表性。也就是說，語料庫分析的結果確實有統計代表性，但是被挑出來代表這種代表性的樣本，未必有統計代表性，使得後續的論述詮釋有可能落入「統計編故事」的陷阱中（柯籙晏，2026）。[7]

為解決前述「統計編故事」的方法論弱點，並滿足研究問題對「統計代表性」的要求，本研究採用柯籙晏（2026）提出的方案，在 CADS 的研究設計中，加入大語言模型輔助內容分析（Large Language Model-Assisted Content Analysis, LACA）作為語料庫分析與內容分析的橋接機

[7] 以收錄 CADS 經驗研究的 *The Routledge Handbook of Corpus Approaches to Discourse Analysis* 一書為例，其導論明確指出，研究者必須詮釋語料庫分析的結果，但對如何確保詮釋的有效性與一致性卻缺乏具體說明（Friginal & Hardy, 2020）。在這本手冊中，也收錄了一些缺乏檢驗語料庫分析與論述分析之間的語義關係的研究，例如 Williams（2020）在其對美國國會英語統一法案辯論語料庫的論述研究中，直接從語料庫的關鍵詞與共現分析跳到論述分析的意識形態批判，而沒有進一步檢驗關鍵詞脈絡（KWIC）語義分析結果的統計代表性，這使得該研究可能只是在所有符合統計模式的 KWIC 樣本中，挑選了符合其意識形態批判的研究目的的樣本來進行論述分析，卻用前期的統計結果來為這些立意挑選的樣本背書。此即本文所謂「統計編故事」的 CADS 方法論陷阱。

制。內容分析能夠系統地檢查具統計模式的 KWIC 樣本的實際語義，以確保在語料庫中所發現的華裔認同協商論述具有統計代表性，而透過大語言模型（Large Language Model, LLM）的輔助，則更能大幅降低內容分析的成本。以本研究針對 3,567 條 KWIC 樣本的二元編碼為例，僅需研究者投入兩個星期進行提示詞優化；同樣的工作若使用傳統內容分析進行，包括建立標準編碼本、訓練編碼員、執行編碼與品管編碼結果等流程，則約需 650-700 工作小時。

一、研究方法操作步驟

本研究方法操作步驟如下：

1. 語料預處理：選擇〈玻璃心〉的評論作為語料庫，進行資料清理，包括建立斷詞辭典，增加如「小粉紅」、「玻璃心」等語料庫中重要的新詞彙後，進行中文斷詞。
2. 詞頻與共現詞分析：對語料庫進行詞頻與共現詞分析，並根據分析結果決定下一階段 KWIC 的檢索詞。如前所述，初步的分析結果顯示，〈玻璃心〉語料庫中「自我宣稱身分」詞「我是」，與華裔相關身分詞如「台灣人、香港人、中國人」高度共現。
3. 關鍵詞脈絡檢索（KWIC）：定位理論（positioning theory）指出，人稱代詞是身分認同論述的有效指標，第一人稱用於宣告自我身分並指出自己的生活史，第二人稱用於劃定身分邊界並分配他人角色（Davies & Harré, 1990/1999）；基於此，本研究選擇「我是、我們／我们、你是、你們／你们」等人稱代詞作為檢索詞，檢索這些詞的左三右七個詞的所有 KWIC 條目，[8] 共得 38,915 條，去除重複資料後為

[8] 左三右七個詞，指的是以檢索詞為中心，取左邊三個詞與右邊七個詞的 KWIC 範圍。

34,880 條，用於檢視這些詞在脈絡中的實際意義。本研究使用 CORPRO（Chueh & Chen, 2016）作為語料庫分析工具，實現上述語料預處理、詞頻分析、共現分析與 KWIC 等工作。

4. 抽樣：為解決 KWIC 條目數量龐大的問題並兼顧統計代表性，在去除重複後，各檢索詞（我是、你是、我們／我们、你們／你们）的 KWIC 條目，依 95% 信心水準與 3% 抽樣誤差分別隨機抽樣，合計得 3,567 條。
5. 大語言模型輔助的內容分析（LACA）：[9] 為確保 KWIC 樣本具有華裔認同協商的論述意義，本研究使用 Claude Sonnet 4 模型對 3,567 條 KWIC 樣本進行編碼，識別並篩選出其中包含「宣稱自我或指涉他人的華裔相關身分」的樣本，如「我是台灣人」、「你們小粉紅」編為 A 類（A 樣本），無關樣本則編為 B 類（B 樣本），。注意，無關樣本除了「我是覺得無言」這樣無身分詞的陰性樣本之外，還排除了偽陽性樣本，如「你們喜歡台灣人」包含身分詞但其人稱代詞並非指涉身分詞。經過信度檢驗，本研究的 LLM 編碼與人工標準編碼的 Cohen's κ 介於 .869 至 .979，顯示高度一致性。最終選出 937 條 A 樣本。LACA 編碼工作的完整提示詞見附錄二。
6. 論述分析：以具有統計代表性的「宣稱自我或指涉他人的華裔相關身分」937 條 A 樣本為基礎，進行論述分析，以呈現語料庫中的不同華裔認同論述。其中包含三個步驟：（1）基於紮根理論（Charmaz, 2014），對 A 樣本進行聚焦編碼（focused coding），透過持續比較法（constant comparative method）編碼直到飽和，共得到七個編碼軸

[9] 大語言模型輔助內容分析（LACA）可應用於各種需要批次語義驗證的內容分析場景。本研究中採用的 LACA 檢驗機制，是基於作者將 LACA 嵌入 CADS 流程中，作為解決統計代表性問題的輔助檢驗機制。關於該 LACA 檢驗機制的理論基礎、人機協作的提示詞設計原則與信度驗證方法，請參閱柯籙晏（2026）。

（coding axes）；（2）關連分析（relational analysis）與論述建構：對「其他」軸以外的六個編碼軸進行關連分析，旨在闡明各軸之間關係的敘事結構，即「誰，在做什麼，怎麼做」，以建構出一個用以描述語料庫中華裔認同協商論述的結構；（3）以前述結構作為檢索指南，在整個語料庫中補充案例，所選用的案例包括但不限於 A 樣本。

二、語料庫描述

本研究採取 2021 年 10 月 15 日至 2022 年 3 月 4 日之間張貼在〈玻璃心〉YouTube 官方音樂影片下的 209,744 則評論作為語料庫，其中包含 73,760 則初階留言以及 135,984 則二階留言，總共約 230,000 個詞彙。如前所述，在將〈玻璃心〉視作一個線上媒介事件的理論前提下，這二十萬則評論構成了一個相對完整且封閉的「全量語料庫」。

有鑑於繁／簡字體是諸華裔社群之間身分協商的重要指標，臺港慣用繁體，中國與東南亞慣用簡體（Sun, 2005）。為評估語料庫的繁簡組成，本研究選取 153 對具代表性的日常高頻的繁簡異形詞進行統計分析。結果顯示，繁簡比例為 50.6% 與 49.4%，呈現高度均勻的混合狀態。[10] 此一比例顯示這個論述空間並未被單一群體主導。

在後設資料方面，採取了初階留言的用戶帳號名（總共 51,819 個不重複帳號），評論者可以使用指令「@帳號名」表示自己的評論是要回

[10] 從完整詞頻表（57,336 個詞彙）中選取 153 對繁簡異形詞，包括聽／听、中國／中国、億／亿、好聽／好听、你們／你们、說／说、讓／让、萬／万等。統計結果：繁體詞彙 68,381 詞次（50.6%），簡體詞彙 66,867 詞次（49.4%），總計 135,248 詞次。計算方法：繁體比例 = Σ 繁體詞次/（Σ 繁體詞次 + Σ 簡體詞次）。透過 Bootstrap 重抽樣檢驗（n = 10,000）確認穩定性，95%信心水準的信賴區間為 46.2%-54.2%。

應特定帳號的評論，這使得研究者可以在某種程度上重建評論者之間的互動。

三、〈玻璃心〉音樂影片描述

本研究的語料庫由〈玻璃心〉的 YouTube 評論組成，亦即，〈玻璃心〉是觸發評論的初始訊息，因此需要先說明〈玻璃心〉說了什麼。

〈玻璃心〉的敘事，是由敘事者與缺席對話者之間的想像對話所構成。這種結構常用於情歌中，用來表達敘事者與缺席對話者之間的關係溝通（relational communication）。在〈玻璃心〉的對話中，敘事者為自己說話太過直白、傷害了缺席對話者的感情而道歉。但在描述缺席對話者如何不講理之後，敘事者堅決拒絕缺席對話者要敘事者向他道歉的要求。歌詞中沒有明白指出缺席對話者是誰，但對於他的特徵描述，如「你擁有粉紅色純潔的心、跟你說話像對熊貓彈琴」，以及在音樂影片中的各種視覺隱喻，例如粉紅色的畫面調性以及一位穿戴大熊貓（中國官方象徵之一）服裝的演員，都指向被稱作「小粉紅」的激進中國國族主義網民。這首歌是用情人間失敗的關係溝通作為隱喻，諷刺海外華人——以兩位海外華人演唱者為代表——與小粉紅之間失敗的關係溝通。如果把〈玻璃心〉歌詞的以下段落當作整首歌的訊息框架（message framing）（方形括號〔 〕中的內容為本文作者補充）：

> 你說我屬於你／別逃避快回家裡／〔你想要的〕一點都不能少／〔我〕都〔得〕讓你贏／〔你〕不講道理……〔你〕要我跪下去〔都聽你的〕／Sorry 我不可以〔接受你的要求〕

那麼，這部音樂影片可以被詮釋為，海外華人與激進中國國族主義

者（小粉紅）之間有劇本的衝突劇場。雖然黃明志本人諷刺地不承認對於〈玻璃心〉的諷刺詮釋，說這首歌「只是一首甜蜜的情歌」，但大多數聽眾都覺知這種諷刺，並在評論區中演出諸華裔之間爭奪身分邊界的即興爭吵劇場。

伍、研究發現

一、〈玻璃心〉語料庫中諸華裔身分認同論述軌跡：[11]基於 A 樣本的描述

本研究透過 LACA 得到 937 條 A 樣本，亦即使用第一、二人稱代詞宣稱自我或指涉他人華裔相關身分，且具統計代表性的 KWIC 樣本之後，繼續透過人工的持續比較法，對 A 樣本進行聚焦編碼。結果歸納出七個編碼軸，這些軸不互斥，容許單一樣本被標記在多軸上。其中包括：

1. 自我身分宣告

「我是台灣人」（A 樣本 13）、「我们中国人」（A 樣本 710）等。在自我身分宣告樣本中出現的身分，可以排列出從中華核心到中華邊緣的光譜。本研究將中國人／大陸人／華人編為「中華核心」，核心以外者編為「中華邊緣」，若同時出現核心與邊緣的複合身分，如「馬來西亞華人」，則同時編為「中華核心與邊緣」。值得注意的是，這個

[11] 本研究在描述基於語料庫資料所建構的論述分布時，將具先驗與靜態隱喻的「結構」一詞，替換為強調經驗與動態的「軌跡」（trajectories）。此舉旨在說明這些論述並非預設的框架，而是由〈玻璃心〉評論者們共同踩踏出的論述實踐路徑。

劃界原則本身就是爭吵劇場的焦點。因此這個原則是基於語料庫中，這些身分被賦予的協商功能，即以中／華身分為爭吵基點。至於找出這些劃界的進階內涵，例如「我是台灣人」的身分標記是否與特定的政治與文化立場掛勾，則屬於後續論述分析的工作。

2. 諷刺他者身分或自嘲身分

為了避免過度詮釋，本研究只編碼明確使用污名化標籤的諷刺或自嘲，如「你们可是一群小粉红呢」（A 樣本 829）、「我是墙国人，哈哈」（A 樣本 397）。由於缺乏背景知識的讀者會看不懂諷刺，基於這個保守的編碼原則，語料庫中諷刺與自嘲論述實際上會更多。

3. 反串或扮演身分

反串或扮演其實是諷刺的一種，但比諷刺或自嘲的解讀更需要背景知識。幸好語料庫中的反串或扮演有一個明顯的成語化形式，如被少數帳號大量使用的「我是中國台灣省人」占樣本 8.8%（A 樣本 17），以及對前者的形式反諷如「我是福建省的榕紅斑」（A 樣本 4）。這種高度重複的固定套路，即使缺乏背景知識也能明確辨認。除此之外，明顯不對應現實的扮演，如「我是巴蜀利亚大公国的公爵」（A 樣本 35）也被本研究編入此軸。同樣地，由於這個保守的編碼原則，語料庫中這類論述實際上更多。

本研究之所以把反串與扮演從諷刺與自嘲中獨立出來，是因為此軸與下一軸「討論帳號身分」高度關連，構成爭吵劇場的重要情節。

4. 討論帳號身分

在匿名評論區中，互相質疑帳號背後的身分本來就是常態，而這種情況在〈玻璃心〉語料庫中更加明顯。此軸包含了對身分的質疑與反質疑，如「我們都知道你是牆內人裝的」（A 樣本 429）、「为什么你认

为我是韭菜」（A 樣本 5）。

此軸與「反串與扮演」構成論述序列：反串引發質疑，質疑引發反質疑的爭吵情節。

5. 討論集體身分

使用後設陳述的形式討論集體身分，包括質疑、反質疑與假設性描述的樣本，如「中國人說我不配，說我是胡人後代」（A 樣本 42）、「我是马来西亚人如果你要种族的应该是华族吧！华族不是中国人哦」（A 樣本 205）、「如果套用西方世界〔的說法〕，那我是温和中国人，出征的是极端中国人」（A 樣本 178），被編入此軸中。

6. 直述他者身分

直述而不討論他人身分的樣本被編入此軸，如「你们马来华人貌似这些年过的不如意哦，多关心关心自己吧」（A 樣本 820）。

7. 其他

基於各種其他原因需要被特殊標記的樣本。如「我是藍營家庭」（A 樣本 264）、「我是一個學習普通話的日裔美國人」（A 樣本 297）等自我宣稱身分不直接與中華核心與邊緣的光譜有關；「垃圾就是垃圾说句话 这歌热度一退他在你们台湾照样 18 线不如」（A 樣本 778）是在討論〈玻璃心〉本身。

編碼結果綜合統計如下表 3 所示：

表 3：基於 A 樣本的聚焦編碼結果與分布

編號	編碼軸名稱	在 A 樣本中的占比
1	自我身分宣告：中華核心認同	27.0%
	自我身分宣告：中華邊緣認同	28.1%
	自我身分宣告：核心與邊緣之間認同	6.3%
2	諷刺他者身分或自嘲身分	20.8%
3	反串或扮演身分	8.8%
4	討論帳號身分	16.9%
5	討論集體身分	35.5%
6	直述他者身分	10.8%
7	其他	2.4%

註：因採用非互斥編碼，單一樣本可同時標記於多個編碼軸，故占比總和超過 100%。

資料來源：本研究彙整。

這些具統計代表性的編碼軸占比，不僅勾勒出評論區中身分指涉的模式，也呈現出參與者在爭吵劇場中爭奪身分邊界的手段。這符合了本研究基於定位理論（Davies & Harré, 1990/1999）所篩選出的 A 樣本既有結構——由第一與第二人稱代詞引導的華裔相關身分定位，其本身就具備「對話性」敘事結構，宣告了說話者的自我身分並指出其生活史；「你是」與「你們」樣本則被說話者用於劃定與對話者之間的身分邊界。因此，根據 A 樣本歸納出來的六個編碼軸（排除「其他」軸）並非彼此獨立，而是一套相互關聯的模式，共同構成語料庫中的華裔身分協商論述。其中角色的部分為軸 1 與 6，論述展演的部分為軸 2、3、4、5。比較特別的是軸 4（討論帳號身分）同時指向場景本身：在匿名的網路平臺上，質疑對方背後身分的真實性，是一種只有在這個場景中才有意義的語藝實踐，場景與展演在此不可分割。

據此，A 樣本呈現角色、場景（YouTube 評論區）與展演等媒介儀式的關鍵特徵（Couldry, 2000; Sumiala, 2012）；這個分析結果讓〈玻璃心〉評論區可以被理解為一個評論者之間諷刺又不信任地討論何謂華裔身分，彼此爭奪身分邊界的即興爭吵劇場。

二、〈玻璃心〉的即興爭吵劇場：誰與誰在吵？吵什麼？怎麼吵？

有鑑於 A 樣本內建的敘事軌跡與功能特徵，並且基於區分了故事（story）與敘事論述（narrative discourse）的敘事分析法（Genette, 1972/1980），本研究將 A 樣本六軸的論述軌跡進行以下轉換：

故事：聚焦於情節內容，即角色（評論者）的行動序列。這包含參與協商者的身分（軸 1、6）與所協商內容（軸 5）。這兩者共同定義了爭吵劇場的行動者與實際行動。

敘事論述：聚焦於表達形式。這對應協商策略（軸 2、3、4），即透過語言進行邊界爭奪的語藝展演工具。

本研究接下來結合語料庫中的具體案例，進一步聚焦〈玻璃心〉的即興爭吵劇場進行分析：誰與誰在吵？吵什麼？怎麼吵？

（一）中華認同敘事：從核心到邊緣

以下本文分析所展示的具體案例，全部來自 A 樣本（標示為「A 樣本編號」）以及語料庫中的其他語料（標示為「B 樣本」），各案例的裁切，如省文、加註或跨評論展示，依分析需要而定。又，評論者帳號名，一律經匿名化處理。

1. 中國人身分的標準敘事：在黨國領導下對抗美帝的中國人

檢視 A 樣本，其中出現一個特殊的身分詞「深红」，如「我是深红，出国之后更红」（A 樣本 65）。透過 KWIC 可以確認，「深红」與其變體「大红」、「赤红」等，是對「小粉紅」這個詞的反質疑：紅是共產黨的代表色，「小粉紅」被用來諷刺中共支持者，中共支持者於是自稱「深红」表示對黨的支持毫不動搖。這顯示「中共就是中國，遵從黨國領導的就是中國人」的認同敘事在〈玻璃心〉爭吵劇場中，是用以爭奪身分邊界的重要語藝資源，如以下例 1：

> 例 1：〔〈玻璃心〉〕挣的也只是舆论风口的钱，反正现在 YouTube 嘴〔註：網路用語，罵的意思〕中国嘴共产党就有流量了……我们中国人知道共产党哪里做得好，哪里做得不行，但是轮不到〔黄明志〕你这个……外人，否定一切说三道四（B 樣本）

「我们中国人知道共产党」顯示評論者自我宣告為「遵從黨國領導的中國人」，透過「嘴中国嘴共产党」的併陳，表達其「黨即是國」的認知，並將透過〈玻璃心〉諷刺激進中國國族主義網民「小粉紅」的黃明志他者化為「你这个外人」。

例 2 進一步說明，在「遵從黨國領導的中國人」敘事中，還有哪些他者：

> 例 2：前言：如果你是中国人，爱国，爱党，请停止观看此视频〔〈玻璃心〉〕……作为一位中国人，这种狗屁不通的言论与这首令人作呕……的音乐实在是令人羞耻……共产党在不知中国奋斗了多久，来成就中国现在的富强……美国人……眼红

了，于是他不厌其烦地给中国制造麻烦……宣传反共文化，支持港独台独，等等等等〔令人〕不齿的下三滥行为（B 樣本）

例 2 尤其有意思地將「美國人」拉進諸華裔爭奪身分邊界的爭吵劇場中。推論鏈為：（1）中國富強的歸因「共产党成就中国」→（2）外部威脅的動機「美国人眼红」→（3）實際威脅行動「宣传反共文化，支持港独台独」，這條推論鏈把爭吵劇場視作「西方帝國主義陰謀」而非諸華裔社群之間的草根自發行動。在語料庫中，「美國／美国」的詞頻是 2,054／4,564，是「中國／中国」的詞頻 13,754／11,896 的 25.80%，可見其重要性。

綜上所述，若以中國官方國族主義作為參照點（參見 Huang & Liu, 2018; Moody, 2007），這類敘事可以稱作中國人身分的「標準敘事」，定義了遵從黨國領導，對抗西方帝國主義的中國人核心認同，並將違反標準敘事者他者化。

2. 標準敘事的偏差：認知黨國相異的中國人

若「遵從黨國領導」是中國人身分的標準敘事，那麼在語料庫中有從標準到偏差的一系列變異。例如例 3、4、5：

例 3：你得清楚〔小〕粉红具体……指的是“中国无条件最棒第一名”的那种，况且政府真的不代表中国，我永远以我是中国人而自豪，但是政府究竟做的怎么样就一言难尽了……按照我的角度看的话我就觉得中共做的很不错了（B 樣本）

例 4：我是中国人，从 2011 年开始反共“未來不是內戰, 就是第三次世界大戰”一起面對這個人類恥辱的魔鬼政權吧（B 樣

本）

例 5：我是大陆人我是华人我精神上不认为我是中国人，那天中国发生战争在不伤及平民的情况下，我绝对无条件支持别国（B 樣本）

認同「我是中国人」的例 3 評論者雖然支持黨國「我就觉得中共做的很不错了」，但已經偏離標準敘事「政府真的不代表中国」。例 4 評論者強烈反黨，但仍認同自己是中國人。而例 5 評論者則甚至否認自己是中國人「我精神上不认为我是中国人」。把例 2、3、4、5 排列起來，可以看到從標準到偏差的敘事光譜，其中爭吵劇場的焦點從標準敘事對「黨—國—中國人身分」的綁定，轉向透過解構標準敘事來再定義中國人身分，例 5 評論者甚至表示「那天中国发生战争在不伤及平民的情况下，我绝对无条件支持别国」來諷刺標準敘事中的帝國主義陰謀論。

（二）標準敘事的激進與溫和詮釋

〈玻璃心〉是身為海外華人的黃明志向小粉紅爭奪身分邊界所演出的爭吵劇場，而〈玻璃心〉的評論則往往以〈玻璃心〉是否「辱華」，或更精確地說，評論者是否感到〈玻璃心〉羞辱了中華核心認同作為爭吵的起點。由於個人會對其所認同團體的恥辱與自豪感同身受（Salice & Montes Sánchez, 2016），對於〈玻璃心〉是否辱華的答案，可以作為檢驗中華核心認同的試金石。在語料庫中，「辱華」與其同義詞的總詞頻是 3,220，甚至比演唱者「陳芳語」的 2,534 更高。〈玻璃心〉是否辱華的論述，可以用來區分評論者對中國人身分標準敘事的激進或溫和詮釋，以下分別說明。

1. 中國人身分標準敘事的激進詮釋

例 6：看到下面评论一大帮〔支持〈玻璃心〉的〕华人津津乐道谈辱华，却不想想自己的民族身份也是华人……使用中文、讲华语、流着中国血的华人丝毫没有民族荣誉感……作为中国人的我……情绪是有的……1 是对……共产党政府的一种敬意，是他们的努力使得大陆华人……可以和西方人……平起平坐……2 是遗憾……很多……外迁的华人……对故乡对中国的概念渐渐淡忘，以至于现在根本不清楚自己的身份，所谓辱华其实是在侮辱包括自己在内民族文化认同……让西方人有机可乘……这种状态下受影响最大的还是那些海外的华人啊（B 樣本）

注意自我宣告為中國人的例 6 評論者如何分析自己的情緒。首先重述標準敘事「共产党政府使得大陆华人可以和西方人平起平坐」，其次責備「用中文、讲华语、流着中国血」的海外華人居然不願意和中國人共享榮譽，反而「津津乐道谈辱华，却不想想自己的民族身份也是华人，让西方人有机可乘」，亦即主張海外華人應該要和中國人一樣認同標準敘事。本文將這種敘事稱作「標準敘事的激進詮釋」，這種詮釋的激進體現為一種大漢文化沙文主義，以漢人血緣、使用官話與漢字作為劃分「中華核心身分」的排他性邊界，徹底無視非漢中國人的存在以及海外華人的母語多樣性（參見 Simpson, 2007）。令人訝異的是，確實也有海外華人認同激進詮釋，例如例 7：

例 7：我是泰國华裔.听了这首諷刺辱华歌曲.不禁让我心生怒火.为在马來西亞华人社会出現一個逆仔〔黃明志〕而喊冤.更为生在马来西亞几代华人为對抗英帝国主义的欺詐压迫而献身

的义士而挽惜（B 樣本）

與例 6 相較，例 7 評論者直接表達自己的受辱感「这首諷刺辱华歌曲让我心生怒火」，並在這種情緒下，評論者把華裔馬來西亞人黃明志稱作父祖的背叛者「逆仔」，而就邏輯而言，他的「泰國华裔」身分首先必須與「马來西亞华人」等同，才有資格以宗族長輩的身分說這種話。其次，他又把「马来西亞华人」與「對抗英帝国主义」的標準敘事聯繫在一起。

綜合來看，例 6、7 評論者表達了「所有海外華人都應該要和中國人一樣認同標準敘事」的激進詮釋，不僅鞏固了標準敘事的核心，更透過訴諸血緣、文化與歷史反帝敘事，要求海外華人接受標準敘事的集體榮辱共享。

2. 中國人身分標準敘事的溫和詮釋

當然不是所有評論者都認同激進詮釋，如例 8：

例 8：我又爱国又是共产党员，这不影响我喜欢黄明志的歌，也不影响我长久以来也对各种动不动就“被辱”和闹剧一样的“出征”〔註：小粉紅對辱華者的集體線上言語攻擊〕感到十分无语……我不认为这些“出征”的小伙伴〔註：中國網路用語，意為親愛的朋友〕能代表中国的爱国主义者，更不可能代表中国人……希望海外的小伙伴可以意识到，如果你在互联网上遇见了很多和你吵架的玻璃心中国人，这些人的背后还有十倍百倍的可以做朋友的不是玻璃心的中国人（B 樣本）

合理推論應為中國籍的例 8 評論者認同標準敘事「我又爱国又是共产党员」，但明顯區分海外華人「海外的小伙伴」與中國人「玻璃心中

国人、可以做朋友的不是玻璃心的中国人」的不同。本文將這個符合中華人民共和國在法律事實上區分中國公民與海外華人的敘事，稱作標準敘事的溫和詮釋。認同溫和詮釋的海外華人，如例 9 明確指出：

例 9：我身为海外华人不明白为什么小粉红每次不爽就说辱华……又不是全世界的华人（华裔）都来自他们的国家啊，人家〔黄明志〕说的是你的国家，又不是那个种族，可以不要擅自代表全世界的华人（华裔）吗？（B 樣本）

從標準敘事是否適用於海外華人的詮釋差異，就可以理解馬來西亞華裔社群內部的中華核心與邊緣身分的爭吵，如例 10：

例 10：马来西亚〔華裔〕有分两类人，比较亲共的我们都称为中华胶。基本上 80 前〔註：1980 年前出生〕的还有这种思想，80 后〔註：1980 年後出生〕的基本不是对政治〔冷感？〕就是不会有那种我是中国人的想法（B 樣本）

然而即使否定激進詮釋，許多評論者也只是想從「我們都是中國人」退回到「我們都是華人」的狀態，體現「中華核心與邊緣之間認同」的游移不定，一方面拒絕激進詮釋要求的黨國認同與反帝敘事，一方面又不願意割斷與大中華區（Greater China）的連結，如例 11：

例 11：与其说〔〈玻璃心〉〕反华 反中 不如说 反共 反极权……我们都是华人 我们从来不会反中 更不会反华（B 樣本）

（三）標準敘事悖論：中國籍的少數族裔敘事

A 樣本中出現了少數卻有關鍵意義的「非漢人中國人」。雖然中華人民共和國號稱五十六族，但一般所謂中國人與華人指的就只是漢人（Gladney, 1994）。這種定義上的混淆在自我認同為漢人之間溝通時問題不大，但與中國籍非漢人溝通時就需要額外強調，如例 12、13、14：

例 12：我也不是汉人，我是蒙古族😊在少数民族预科学校里认识了一大堆新疆朋友，维族、哈族、俄罗斯族……都有，到现在联系也都很紧密（B 樣本）

例 13：我是朝鮮族人，不是漢族，你們〔漢人之間〕的事隨意，我不過是看熱鬧~我早不在中國了，留著〔中〕國籍不過是工作生意方便，哈哈哈（B 樣本）

注意 12、13 評論者對「不是漢人」的身分邊界強調，以及非言語訊息「😊」與「哈哈哈」，顯示他們在爭吵劇場中事不關己的姿態。中國籍的少數族裔敘事揭示了：標準敘事與其說是黨—國—中國人認同，不如說是黨—國—漢人認同。當然，即使在血緣與文化上與華人／漢人毫無瓜葛，非漢中國人可能接受標準敘事的愛黨、愛國、反帝，但也可能反對標準敘事，如例 14：

例 14：我是维吾尔族出生在新疆喀什.感谢台湾人民的勇敢和坚强.共产党溺〔滅？〕亡（B 樣本）

（四）替代敘事：與中華核心劃清界線的身分

前述分析，是以中國國族主義為基礎參照點編排的。然而既然不只中國有國族主義，爭吵劇場中當然也可以看到，與中華核心身分徹底劃

清界線的替代敘事。事實上，前述例 14 已經接近這個範疇，只差沒有直接表達。以下案例則是直接表達：

例 15：越南華裔我們寫字就用繁體〔漢字〕, 說話大部分是用粵語, 還有我們常叫我們自己是 “唐人”, “唐話” . we dont use “中国人” 來說自己（B 樣本）

例 15 評論者一一駁斥了激進詮釋的血緣、文字與語言理論，用「唐人」、「繁體」、「唐話」來與中華核心身分劃清界線。「唐人」是大清國禁止治下臣民出海的「天朝棄民」時代，海外華人由下而上的自發（autonomic）認同，不像「華人」「中國人」是中國官方國族主義由上而下的發明。「越南華裔我們」與其說是評論者的自稱，不如說是評論者為了與其他參與爭吵者溝通而採用的他稱，「唐人」才是評論者真正（authoritative）認同的身分。此外評論者提醒，漢字使用者未必說官話「我們說話大部分是用粵語」，繁體字則被用來與簡體字使用者劃清界線。尤其值得注意的細節，原本寫繁體字的例 15 評論者在與中國人身分劃清界線時刻意切換成英文「we dont use」與簡體字「中国人」，然後馬上切換回繁體字「來說自己」。

在語料庫中，最顯著地與中華核心身分劃清界線的是馬來西亞人：共現分析的詞彙配對「我是，马来西亚人」的 Z-Score 是 8.215（$p < .001$），遠高於「我是，华人」的 4.335（$p < .001$）；而在 A 樣本中「我不是中国人我是马来西亚人」這樣明確劃清界線的樣本就有 6 個。合理的推測是，華裔馬來西亞人黃明志的〈玻璃心〉吸引了更多華裔馬來西亞人參與爭吵。如例 16：

例 16：我是马来西亚华人，出国自我介绍是 “Malaysian” ……

在种族是一名华人之前，我是马来西亚人。虽然我们不是什么强国，但国民的自尊与底线我还是有的。不能理解为何我国领土被入侵〔註：馬中兩國在南中國海的主權爭議〕，中华胶还能帮中国说好话……我不挺共就被人标签为崇洋媚外亲美国（B 樣本）

在馬中兩國主權爭議的問題上，認同馬來西亞人身分的例 16 評論者，與認同標準敘事激進詮釋的「中華膠」徹底劃清界線。

例 15、16 只是與激進詮釋劃清界線，例 17 更與溫和詮釋也劃清界線：

例 17：〔用戶 Hoi ○ Wong〕：……我〔是〕馬來人,中國入侵〔南中國海〕九段線,〔把新冠〕肺炎傳播〔到〕馬來西亞,我都十分森七七〔註：臺灣網路用語，生氣的意思〕……你們中國人喜歡……大一統思想,〔我們〕馬來人叫〔這種人〕大中華膠

〔用戶 A〕你用华文名〔Hoi ○ Wong〕，你说你马来人？你反串？

〔Hoi ○ Wong〕：在馬來人〔之中〕我是娘惹,我有馬來血統,當然是馬來人（B 樣本）

例 17 評論者首先在血緣上認同馬來人身分「我是娘惹，我有馬來血統，當然是馬來人」。在地理大發現時代，馬來人稱所有外國婦女為 Nyonya，包括華人移民和馬來土著的混血後裔，之後這個詞彙越來越特指這個群體的婦女，而「娘惹」則是用漢字注音檳城福建話的 nō-niâ（Mahmood, 2004）。其次，在馬來西亞的中文教育是使用簡體字的情

況下，她使用繁體字留言，而她在評論中的省字「我馬來人、在馬來人我是娘惹」可看出她不很會用中文，其三，她與標準敘事認同者劃清界線「你們中國人喜歡大一統思想，馬來人叫大中華膠」。另一方面，大概和案例 9 評論者主張「用中文」、「讲华语」、「流着中国血」就該認同標準敘事一樣，用戶 Hoi ○ Wong 的對話者無視馬來西亞人身分複雜的族裔、文化與歷史內涵，單憑根本不是漢字，只是看似羅馬字母拼音漢字的名字「Hoi ○ Wong 」，就質疑對方「用华文名」，是反串的馬來人。

爭吵劇場中的替代敘事再例如：

例 18：〔用戶 B〕：翻牆出來的中國同胞們，我是台灣人，我不贊成我們國家塔綠斑〔註：網路用語，諧音諷刺以綠色作為象徵的民進黨政府是塔利班〕的一些政策，我也不爽美國賣高價垃圾武器給我們，但是中國也別想用武力控制我思想的自由，希望有更多的粉紅同胞出來呼吸自由的空氣。
〔用戶 C〕：台灣人才不會說什麼中國同胞
〔用戶 D〕：〔你反串台灣人〕裝的很差勁哦！（B 樣本）

例 18 評論者一方面接受標準敘事「中國同胞們，我也不爽美國賣高價垃圾武器給我們」，一方面認同與中國劃清界線的臺灣國族身分「我是台灣人，我不贊成我們國家的一些政策，但是中國也別想用武力控制我」。這兩個敘事未必矛盾，但例 18 評論者的對話者仍然質疑他是反串的臺灣人「裝的很差勁哦」，因為他沒有與標準敘事徹底劃清界線「台灣人才不會說什麼中國同胞」。

〈玻璃心〉爭吵劇場也吸引了英文使用者參與，在以 American 為檢索詞的全部 97 條 KWIC 樣本中，有 17 條明確自我宣告 American，如：

例 19：〔用戶 E〕：I’m American but I’m not gonna lie this is fire 🔥, support the free spirit! ✪〔譯文：我是美國人，但不得不說這首歌真讚🔥，支持自由精神！✪〕

〔用戶 F〕：Hi, I’m a American too but I grow up in the China. This song tell you true about the china. The CCP own the China so this country is super evil.〔譯文：嗨，我也是美國人，但我在中國長大。這首歌告訴你中國的真相。中共擁有中國，所以這個國家超邪惡〕（B 樣本）

在例 19 中，一位與爭吵劇場的身分邊界爭奪行動無涉的美國人評論者，很美國地以自由政治意識形態來評論〈玻璃心〉「support the free spirit! ✪」；另一位連「I’m a American」都文法錯誤，顯然英語非母語的美國人評論者，只願意說「I grow up in the China」而非「I’m Chinese American」，然後從邏輯上徹底解構標準敘事：「The CCP own the China so this country is super evil」在承認綁定了黨—國—中國人的標準敘事是正確的前提下，這個結論就會是正確的。

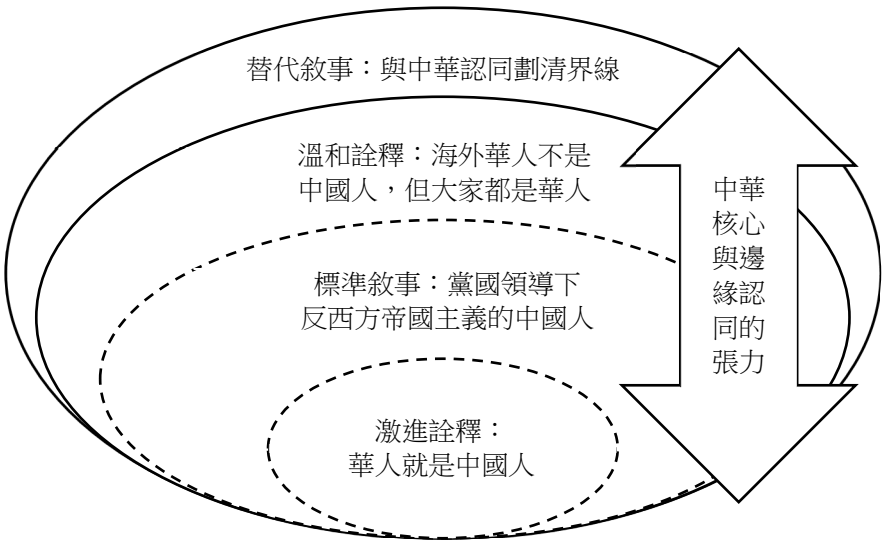
綜上所述，替代敘事以爭吵劇場的草根主角們各自的多元論述，從歷史與文化（例 15、17）、國家主權（例 16、17）、血緣（例 17）、政治意識形態（例 18）到邏輯（例 19），全面否定了標準敘事，無論其激進或溫和詮釋。

三、小結

本研究的發現指出，〈玻璃心〉的即興爭吵劇場呈現如下敘事軌跡：參與爭吵的草根主角們，將爭奪彼此之間身分邊界的演出，聚焦於

中華核心認同的「標準敘事」，亦即綁定「黨—國—中國人」，並將這種綁定與作為他者的「美帝」對立。〈玻璃心〉所觸發的辱「華」與否辯論，揭示了不同主角對標準敘事的不同詮釋。激進詮釋者主張將「黨—國—中國人認同」擴張至血緣與文化邊界，要求海外華人共享中國人的集體榮辱。溫和詮釋者則反對激進詮釋，主張區隔政治上的中國人與文化上的華人。而在中華邊緣，反標準敘事者透過多元論述的替代敘事，從歷史、文化、國家主權、血緣、政治意識形態到邏輯等各種面向否定標準敘事，無論激進或溫和詮釋。據此，〈玻璃心〉中的華裔相關身分敘事，可以圖示化為中華核心與邊緣認同之間的張力模型，如圖 1 所示。

圖 1：〈玻璃心〉評論中的華裔相關身分敘事地圖



資料來源：本研究繪製。

陸、討論

整合認同協商理論（Integrative Identity Negotiation Theory）指出，認同協商中的爭吵並非只是溝通失敗，它實際上是身分邊界的「警報器」：當彼此不相容的價值觀在同一場域碰撞時，所產生的摩擦與爭執，正是界定「我／我們」與「你／你們」之間區隔的實質邊界所在（Ting-Toomey & Dorjee, 2018）。〈玻璃心〉觸發的爭吵劇場就像一場大規模的破壞常規實驗（breaching experiments, Garfinkel, 1967/1991），[12] 透過諷刺極端中國國族主義網民，強行中斷標準敘事的溫和詮釋所建構的認同——在後冷戰全球化時代，這種認同被稱作「大中華區」（Greater China）——來維持表面上的和諧，迫使華裔相關身分之間的實質邊界浮現。本研究的發現，即描繪了這場實驗所揭示的邊界地圖。透過參照這張地圖，許多在〈玻璃心〉爭吵劇場中看似缺乏意義的評論，便不再只是噪音，而是可理解的語藝實踐。以下討論透過三個案例，重新定位這些實踐。

首先，本研究雖然是以中文評論作為繪製地圖的起點，然而一旦地圖建立，便能定位出那些功能等同的非中文評論，如例 20：

> 例 20：Sudah bukan kali pertama china menceroboh kawasan laut/zon udara kita. Tapi setelah kita malaysia patriot nak mempertahankan kedaulatan negara malaysia sendiri, penjilat

[12] 破壞常規實驗，是俗民方法論創始人 Garfinkel（1967/1991）研發的實驗程序，透過刻意「製造麻煩」，破壞被成員視為理所當然且「看而未察」的背景期待，使原本隱形的社會運作邏輯，在成員試圖修補混亂、恢復秩序的互動過程中變得可供分析。

komunis terus menuduh kita sebagai pengkhianat cina.〔譯文：這已經不是中國第一次入侵我們的海空領域了。但當我們這些愛國的馬來西亞人要捍衛自己國家的主權時，那些舔共的卻罵我們是華人的叛徒〕（B 樣本）

表面上看起來，例 20 評論者只是用馬來文重複了例 16、17 的評論內容，但馬來文作為後設訊息的意義是，它讓這則評論不會落入「用漢字、講官話」的標準敘事激進詮釋中，避開例 17 評論者被質疑「你用华文名，你说你马来人？你反串？」的情況。這不是試圖與他者爭吵，而是在尋求我群（馬來文使用者）的認同。

其次，地圖能使一些隱諱的評論背後的敘事軌跡顯現出來。如例 21：

例 21：我是上海人，我非常喜欢这首歌，我还把它分享给了我美国和新加坡的同事。美国同事是白人，我把有人写的每个时间点的梗也发给她了，她非常喜欢，谢谢二位马来西亚歌手的贡献（B 樣本）

乍看之下，例 21 只是在陳述跨文化交流的日常瑣事；然而一旦放入地圖，例 21 便可被理解為一個隱諱的替代敘事：透過宣告「我是上海人」而非中國人、把滿滿都是辱華「梗」的〈玻璃心〉分享給「美国和新加坡的同事」，還強調「美國同事是白人」而非華人、稱呼黃明志與陳芳語是「馬來西亞人歌手」而非華人歌手。例 21 沒有正面宣告任何反對標準敘事的立場，但透過讓標準敘事元素缺席（absent）的語藝實踐，產生出與例 5 具體表達的「我精神上不认为我是中国人，我绝对无条件支持别国」功能等同的論述效果。

最後，地圖能使乍看之下難以理解的推論鏈清晰浮現，如例 22：

例 22：我曾经是小粉红，老共的支持者，目前是深红，我觉得这首歌非常好听。查了黄明志的经历，我仿佛看到了一个哗众取宠版的保尔柯察金，在演绎《哥们儿的心是铁打的》虽然我听出了这首歌所表达的爱恨难分，但看看共产党的历史，再看看台湾社会发出的政经议题的声音。塔绿班们是打算对共产党人走过的死胡同都打一遍卡吗？而且你们确定要最后打卡“改革开放”这条胡同吗？怎么说呢，我们这类政见者“被”玻璃心碎了，你们不光能听见了声音，不信你低头，你们还能看见玻璃渣子别踩哦，踩了会有呱呱呱的声音〔註：中國網路用語，諷刺台灣人是「井底之蛙，台蛙」〕～～（B 樣本）

若缺乏適當的參照，例 22 的評論很容易讓人覺得只是在進行淺薄的反諷與輕率的口頭威脅。然而一旦放進本研究繪製的認同邊界地圖，其推論邏輯便顯現出高度自洽：例 22 評論者指出〈玻璃心〉「爱恨难分」，表面上試圖與小粉紅劃清界線，但使用了大量來自華語流行音樂產業（Mandopop industry）的資源，而這個產業本身是後冷戰全球化搭配中國改革開放的產物。[13] 在此背景下，例 22 評論者利用共產主義革命英雄的符號，將黃明志比作「哗众取宠版的保尔柯察金」，[14] 諷刺其

[13] 1990 年代後期成形的「華語流行音樂產業」（Mandopop industry）是標準的大中華區產物，整合了港台的創作人才與包裝行銷能力，以及中國大陸的廣大市場與資本（Moskowitz, 2010）。身為華裔馬來西亞人的黃明志與華裔澳洲人的陳芳語的演藝生涯軌跡，以及〈玻璃心〉所使用的各種象徵符號，都是這條產業鏈的產物。

[14] 保爾・柯察金（Pavel Korchagin）是蘇聯作家 Nikolai Ostrovsky 的小說《鋼鐵是怎樣煉成的》中的主角，在 20 世紀中葉曾作為蘇聯、中國等共產國家青年的集體精

反抗姿態只不過是商業邏輯下的膚淺表演。他並未像其他支持標準敘事的評論者那樣陷入自我辯護，而是透過「塔绿班」一詞將以臺灣人為代表的海外華人他者化，宣告「大家都是華人」這條改革開放之路，已經被〈玻璃心〉的自以為幽默堵死。他化為深紅，向昔日同胞、今日他者發出警告：曾對海外華人滿懷期待、如今卻傷透了心的中國人，已經用自己破碎的玻璃心渣子鋪就了一條新的認同邊界；誰要是輕率踐踏這條邊界，可是會痛得「呱呱呱」叫。

柒、結論

本研究以〈玻璃心〉音樂影片的 YouTube 評論作為一個邊界明確的原生線上媒介事件，透過全量語料庫取徑，說明諸華裔社群如何在這個（自稱）中國用戶冒險翻牆大量參與的特殊媒介事件中，協商彼此身分認同。研究結果顯示，語料庫中的互動並非零散意見的集合，而是一個具備草根主角、舞台場景與即興展演的爭吵劇場，參與者們透過自我身分宣告與他者身分指涉的肯定、質疑與後設討論，以及各種語藝實踐如諷刺與自嘲、演出與反串等，爭奪彼此身分邊界。

在此意義下，本研究關注這些認同論述如何以可理解的互動方式被組織起來。透過經驗資料的分析，本研究指出〈玻璃心〉爭吵劇場中的即興演出，都以在黨國領導下反對西方帝國主義的中國人認同敘事，即「標準敘事」為基礎參照點。在這個論述空間中，不同身分敘事並非單純並列，而是圍繞在標準敘事周遭被辨識、回應與評價；所謂「替代敘事」，就是對標準敘事所主張的血緣、語言文字、國家主權與政治意識

神偶像與革命英雄典型。

形態等元素的內涵進行解構與重組，迫使彼此之間身分認同的實質邊界顯現出來。

在研究方法上，本研究引入 LACA，針對基於語料庫統計模式篩選出的樣本進行語義檢驗。此舉補強了傳統 CADS 容易落入的「統計編故事」風險，確保本研究對〈玻璃心〉即興爭吵劇場的描述，是建立在具統計代表性的樣本上。這項方法實踐不僅穩固了本研究的詮釋信度，也為研究者在未來處理鉅量線上語料研究中遭遇同樣問題時，提供一條門檻合理的技術取徑。

本研究是對單一原生線上媒介事件的詮釋分析。全量語料庫所保證的統計代表性，其範圍止於〈玻璃心〉媒介事件本身。而本文的認同邊界地圖，描繪的是 2021 年 10 月至 2022 年 3 月間，諸華裔社群在〈玻璃心〉評論區的特殊條件下所展演的論述軌跡。本文的詮釋能否延伸至其他事件、平臺或時期，甚至在某種程度上代表華裔認同論述的趨勢發展，有待後續研究檢證。本研究提供的只是在此限制條件下，一個可供未來研究作為參照的基礎。

參考書目

Namewee（2021年10月15日）。〈黃明志 Ft. 陳芳語【玻璃心】（Official Music Video）〉【影音】。上網日期：2026年3月1日，取自https://www.youtube.com/watch?v=-Rp7UPbhErE

庄妙菁（2005）。《“华语”一词的历史演变与发展》。南方学院出版社。

柯籙晏（2026）。〈橋接統計與詮釋：大語言模型輔助內容分析用於語料庫輔助論述分析的概念驗證〉，《新聞學研究》，167，123-180。https://doi.org/10.30386/MCR.202604.0008

Anderson, B. (1983/2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso.

Antaki, C., & Widdicombe, S. (1998). Identity as an achievement and as a tool. In C. Antaki & S. Widdicombe (Eds.), *Identities in talk* (pp. 1-14). SAGE.

Baker, P. (2006). *Using corpora in discourse analysis*. Bloomsbury Academic.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Wiley.

Brinkerhoff, J. M. (2009). *Digital diasporas: Identity and transnational engagement*. Cambridge University Press.

Cerulo, K. A. (1997). Identity construction: New issues, new directions. *Annual Review of Sociology*, *23*, 385-409.

Charmaz, K. (2014). *Constructing grounded theory*. SAGE.

Chueh, H.-C., & Chen, K.-H. (2016). CORPRO: A Chinese language corpus tool and a case study of media representation of organic agriculture. In J. Xiang (Ed.), *Digital humanities: Between past, present and future* (pp. 285-313). National Taiwan University Press.

Clayton, R., Murdoch, S. J., & Watson, R. N. M. (2006). Ignoring the great firewall of China. In G. Danezis & P. Golle (Eds.), *Privacy enhancing technologies* (pp. 20-35). Springer Berlin Heidelberg.

Couldry, N. (2000). *The place of media power: Pilgrims and witnesses of the media age.* Routledge.

Davies, B., & Harré, R. (1990/1999). Positioning and personhood. In R. Harré & L. v. Langenhove (Eds.), *Positioning theory: Moral contexts of intentional action* (pp. 32-52). Blackwell.

Davis, J. L., Love, T. P., & Fares, P. (2019). Collective social identity: Synthesizing identity theory and social identity theory using digital data. *Social Psychology Quarterly*, *82*(3), 254-273. https://doi.org/10.1177/0190272519851025

Dayan, D., & Katz, E. (1992). *Media events: The live broadcasting of history*. Harvard University Press.

Du, Yan. (2025). Homeland social media as a field of negotiation: Creative strategies of China's Wanghong in Japan. *Global Media and China*, *10*(3), 353-371. https://doi.org/10.1177/20594364241293216

Du, Yu. (2025). Practising citizenship through online media: An interpretive case study of Chinese New Zealanders' civic engagement online. *Media International Australia*, *195*(1), 137-151. https://doi.org/10.1177/1329878x231185563

Friginal, E., & Hardy, J. A. (2020). Corpus approaches to discourse analysis: Introduction and section overviews. In E. Friginal & J. A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 1-4). Routledge.

Gao, Y., Li, J., & He, J. (2025). Hybridizing tradition and technology: Global reception of "Black Myth: Wukong" as a vehicle for Chinese cultural identity and international media flows. *Global Media and China*, *0*(0). https://doi.org/10.1177/20594364251385203

Garfinkel, H. (1967/1991). *Studies in ethnomethodology*. Wiley.

Genette, G. (1972/1980). *Narrative discourse: An essay in method*. Cornell University Press.

Gladney, D. C. (1994). Representing nationality in China: Refiguring majority/minority identities. *The Journal of Asian Studies*, *53*(1), 92-123. https://doi.org/10.2307/2059528

Hayes, J. E. (2000). *Radio Nation Communication, popular culture, and nationalism in Mexico, 1920-1950*. University of Arizona Press. https://doi.org/10.2307/j.ctv125jssk

Huang, H., & Liu, X. (2018). Historical knowledge and national identity: Evidence from China. *Research & Politics*, *5*(3), 2053168018794352. https://doi.org/10.1177/2053168018794352

Hutchings, S., & Tolz, V. (2015). *Nation, ethnicity and race on Russian television: Mediating post-Soviet difference*. Taylor & Francis.

Jerslev, A. (2022). Contemporary ceremonial media events – time and temporalities of liveness. *Nordic Journal of Media Studies*, *4*(1), 19-36. https://doi.org/10.2478/njms-2022-0002

Kobayashi, T. (2020). Depolarization through social media use: Evidence from dual identifiers in Hong Kong. *New Media & Society*, *22*(8), 1339-1358. https://doi.org/10.1177/1461444820910124

Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). The whole is always smaller than its parts': A digital test of Gabriel Tardes' monads. *The British Journal of Sociology*, *63*(4), 590-615.

Leary, M. R., & Tangney, J. P. (Eds.). (2012). *Handbook of self and identity*(2nd ed.). The Guilford Press.

Leo-Liu, J., & Wu-Ouyang, B. (2025). The good, bad, and ugly in social virtual reality: The alternative global village for identity negotiations in VRChat. *Social Media + Society*, *11*(2), 20563051251340854. https://doi.org/10.1177/20563051251340854

Li, B., Scott, O. K. M., Zhao, L., & Jin, S. (2024). Are they really Chinese? Examining Chinese audiences' emotions and perceptions toward naturalized athletes at the 2022 Winter Olympics. *Communication & Sport*, *12*(1), 40-62. https://doi.org/10.1177/21674795231185061

Lupton, D. (2015). *Digital sociology*. Routledge, Taylor & Francis Group.

Mahmood, E. (2004). *The Nyonya Kebaya: A century of Straits Chinese costume*. Periplus Editions (HK) Limited.

Moody, P. R. (2007). The evolution of China's national interest: Implications for Taiwan. In S. Horowitz, U. Heo, & A. C. Tan (Eds.), *Identity and change in East Asian conflicts: The cases of China, Taiwan, and the Koreas* (pp. 27-48). Palgrave Macmillan US. https://doi.org/10.1057/9780230603134_2

Moskowitz, M. L. (2010). *Cries of joy, songs of sorrow : Chinese pop music and its cultural connotations*. University of Hawai'i Press.

Robinson, J. Y., & Enli, G. (2022). #MakeSwedenGreatAgain: Media events as politics in the deterritorialised nationalism debate. *Nordic Journal of Media Studies*, *4*(1), 56-

80. https://doi.org/10.2478/njms-2022-0004

Rogers, R. (2015). *Digital methods*. MIT Press.

Salganik, M. J. (2017). *Bit by bit: Social research in the digital age*. Princeton University Press.

Salice, A., & Montes Sánchez, A. (2016). Pride, shame, and group identification [Hypothesis and Theory]. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00557

Saunders, R. A. (2010). *Ethnopolitics in cyberspace: The Internet, minority nationalism, and the web of identity*. Bloomsbury Publishing.

Simpson, A. (2007). *Language and national identity in Asia*. Oxford University Press. https://doi.org/10.1093/oso/9780199267484.001.0001

Sumiala, J. (2012). *Media and ritual: Death, community and everyday life*. Taylor & Francis.

Sun, W. (2005). Media and the Chinese diaspora: Community, consumption, and transnational imagination. *Journal of Chinese Overseas*, *1*(1), 65-86. https://doi.org/10.1353/jco.2007.0007

Tam, G. A. (2020). *Dialect and nationalism in China, 1860-1960*. Cambridge University Press.

Ting-Toomey, S., & Dorjee, T. (2018). *Communicating across cultures*(2nd ed.). Guilford Publications.

Williams, E. A. E. (2020). Critical discourse analysis for language policy and planning. In E. Friginal & J. A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 481-498). Routledge.

Yang, F. (2022). From ethnic media to ethno-transnational media: News-focused WeChat subscription accounts in Australia. In W. Sun & H. Yu (Eds.), *WeChat and the Chinese diaspora* (pp. 53-68). Routledge.

Zeng, J., & Fang, H. (2025). Negotiating issue visibility in the diaspora: A study of political discussion within Jianzhong Quan on X, 2014-2023. *Convergence*, *31*(3), 1009-1030. https://doi.org/10.1177/13548565251335412

附錄一：LACA 編碼信度與不一致率表

本表呈現 Claude Sonnet 4 模型對本研究四組 KWIC 樣本進行辨識的信度檢驗結果。各指標說明如下：

（1）樣本數：從全量語料中隨機抽取（信心水準 95%，抽樣誤差 3%）並進行人工標準編碼校對的樣本總數（總計 3,567 條）。

（2）與標準編碼 κ：衡量 LLM 編碼結果與人類專家「標準編碼」的一致性。本研究四組樣本的 κ 值均在 .869 至 .979 之間，根據學術標準（$\kappa \geq .80$），顯示 LLM 的判定邏輯與人類高度一致。

（3）內部κ：衡量同一模型在不同時間對相同樣本進行五次編碼的一致性，藉此檢驗 LLM 的編碼穩定性。數據顯示 Sonnet4 編碼結果穩定。

（4）不一致率：模型判定結果與標準編碼不符的比例。

KWIC 檢索詞	樣本數	與標準編碼 κ	內部 κ	不一致率
我是	781	.979	.987	1.7%
你是	793	.883	.876	9.2%
我們／我们	980	.915	.908	6.2%
你們／你们	1,013	.869	.860	8.4%

附錄二：LACA 提示詞

（1）用於「我是」KWIC 樣本編碼的提示詞（Sonnet4, κ = 0.979）

編碼原則：

研究背景：分析《玻璃心》（2021）YouTube 評論中的身分敘事，聚焦自我身分表達。以「我是」為關鍵詞的左三右七 kwic 集合而成的語料庫

A 類（肯定或探討「我」的身分）：

語句涉及「我是」與顯性身分標記（政治實體、地理身分、族群身分、政治身分、集體標籤）的確認或探討，聚焦「我」的身分定義，涵蓋：

「我是」+ 顯性身分標記（如「我是台灣人」）。

「我是」+ 否定詞（「不是」「不喜歡」「不接受」「不認同」等）+ 顯性身分標記，描述「我是」的身分（如「我是不是小粉紅」「我是不喜歡被說是小粉紅」）。

他人表示「我」是顯性身分標記（如「他們說我是境外勢力」）。

他人表示「我」不是顯性身分標記（如「他們說我不是台灣人」）。

他人質疑「我」的顯性身分標記（如「他們說我是不是台灣人？」）

關鍵：必須包含顯性身分標記（如「台灣人」「小粉紅」），且聚焦「我」的身分確認或探討（不要求認同）。

B 類（不涉及肯定或探討「我」的身分）：

語句完全無顯性身分標記，或身分標記僅作為「我」的行動的對象，如「我是客觀討論，中國會滅亡」「我是討厭共產黨」。

明確規範：若無顯性身分標記，一開始即歸為 B 類，避免推斷隱含身分（如「語言」暗示「華人」，「立場」暗示「統派」）。

身分標記定義

政治實體：

範例：美國、台灣、中國、中共等（含縮寫：US、台、中、ccp）。反例：國家、政府、祖國

地理身分：

範例：歐美、大陸、香港、台北人、中國台灣省、馬來籍、越南籍、牆內、牆國等。反例：國外、海外、地區、地方、山上、島內。

族群身分：

範例：華人、漢人、客家人、少民等。反例：民族、同胞。

政治身分：共產黨、KMT、民進黨、統派、獨派、韓粉、馬列子孫等（正式立場或政黨認同）

集體標籤：

範例：五毛、小粉紅、韭菜、塔綠班、台蛙、1450、牆內人、中華膠、反賊、玻璃心、境外勢力等。反例：老百姓、普通人、底層、群眾、人民。

- 變體：縮寫（台、中）、複合形式（台澎金馬、北上廣深）、修飾性複合詞（真正的中國人、緬甸勞工、正港台灣人）、行動動詞（辱華、反共，僅限如「你們反共人士」）。

根據以下格式編碼以下語料：流水號順序 tab 理由 tab 編碼結果（小寫），範例：001 無顯性身分標記 b、003 含顯性身分標記「馬列同路人」，他人表示「我是」該身分 a、987 「我」肯定自我身分「小粉紅」a

（2）用於「你是」KWIC 樣本編碼的提示詞（Sonnet4, $\kappa = 0.883$）

編碼原則：

研究背景：分析《玻璃心》（2021）YouTube 評論中的身分敘事，聚焦自我身分表達。以「你是」為關鍵詞的左三右七 kwic 集合而成的語料庫

A 類（肯定或探討「你」的身分）：
語句涉及「你是」與顯性身分標記（政治實體、地理身分、族群身分、政治身分、集體標籤）的確認或探討，聚焦「你」的身分定義，涵蓋：
「你是」+ 顯性身分標記（如「你是台灣人」）。
「你是」+ 否定詞（「不是」「不喜歡」「不認同」等）+ 顯性身分標記，描述「你是」的身分（如「你是不是小粉紅」「你是不喜歡被說小粉紅」「你是不認同中國人」）。
他人表示「你」是顯性身分標記（如「他們說你是境外勢力」）。
他人表示「你」不是顯性身分標記（如「他們說你不是台灣人」）。
他人質疑「你」的顯性身分標記（如「他們說你是不是台灣人？」）
關鍵：必須包含顯性身分標記（如「台灣人」「小粉紅」），且聚焦「你」的身分確認或探討（不要求認同）。

B 類（不涉及肯定或探討「你」的身分）：
語句完全無顯性身分標記，或身分標記僅作為「你是」的行動的對象，如「你是認為中國會滅亡？」「你是不是支持台灣？」「你是被中共洗腦」。

明確規範：若無顯性身分標記，一開始即歸為 B 類，避免推斷隱含身分（如「語言」暗示「華人」，「立場」暗示「統派」）。
身分標記定義
政治實體：
範例：美國、台灣、中國、中共等（含縮寫：US、台、中、ccp）。反例：國家、政府、祖國。
地理身分：
範例：歐美、大陸、香港、台北人、中國台灣省、馬來籍、越南籍、牆內、牆國等。反例：國外、海外、地區、地方、山上、島內。
族群身分：
範例：華人、漢人、客家人、少民等。反例：民族、同胞。

政治身分：共產黨、KMT、民進黨、統派、獨派、韓粉、馬列子孫等（正式立場或政黨認同）。

集體標籤：

範例：五毛、小粉紅、韭菜、塔綠班、台蛙、1450、牆內人、中華膠、反賊、玻璃心、境外勢力等。反例：老百姓、普通人、底層、群眾、人民。

- 變體：縮寫（台、中）、複合形式（台澎金馬、北上廣深）、修飾性複合詞（真正的中國人、緬甸勞工、正港台灣人）、行動動詞（辱華、反共，僅限如「你們反共人士」）。

根據以下格式編碼以下語料：流水號順序 tab 理由 tab 編碼結果（小寫），範例：001 無顯性身分標記 b、003 含顯性身分標記「馬列同路人」，他人表示「你是」該身分 a、987 肯定「你是」顯性身分「小粉紅」 a、561 「你是」的行動的對象是地理身分「海南人」 b

（3）用於「我們/们」KWIC 樣本編碼的提示詞（Sonnet4, $\kappa = 0.915$）

研究背景：分析《玻璃心》（2021）YouTube 評論中的身分敘事，聚焦身分表達。以「我們」為關鍵詞的左三右七 kwic 集合而成的語料庫。

A 類（「我們」肯定或探討自我身分）：

語句涉及「我們」與顯性身分標記（政治實體、地理身分、族群身分、政治身分、集體標籤）的確認或探討，聚焦「我們」的身分定義，涵蓋：

「我們」+ 顯性身分標記（案例「我們台灣人」）。

「我們」+ 否定詞（「不是」「不喜歡」「不接受」「不認同」等）+ 顯性身分標記，描述「我們」的身分（範例「我們不是中國人」「我們不喜歡被說是中國人」）。

他人表示「我們」是顯性身分標記（如「他們說我們是境外勢力」）。

他人表示「我們」不是顯性身分標記（如「你們不要說我們是中國

人」）。
非標準語序（如「作為台灣人的我們」）等同於「我們」+顯性身分標記，歸為 A 類。
關鍵：必須包含顯性身分標記（如「台灣人」「小粉紅」），且聚焦「我們」的身分肯定或探討（不要求認同）。

B 類（不涉及肯定或探討自我身分）：
語句完全無顯性身分標記，或身分標記僅作為「我們」的行動的對象，範例「我們統一台灣」。

明確規範：若無顯性身分標記，一開始即歸為 B 類，避免推斷隱含身分（如「語言」暗示「華人」，「立場」暗示「統派」）。

身分標記定義
政治實體：
範例：美國、台灣、中國、中共等（含縮寫：US、台、中、ccp）。反例：國家、政府、祖國。
地理身分：
範例：歐美、大陸、香港、台北人、中國台灣省、馬來籍、越南籍、牆內、牆國等。反例：國外、海外、地區、地方、山上、島內。
族群身分：
範例：華人、漢人、客家人、少民等。反例：民族、同胞。
政治身分：共產黨、KMT、民進黨、統派、獨派、韓粉、馬列子孫等（正式立場或政黨認同）。
集體標籤：
範例：五毛、小粉紅、韭菜、塔綠班、台蛙、1450、牆內人、中華膠、反賊、玻璃心、境外勢力等。反例：老百姓、普通人、底層、群眾、人民。
- 變體：縮寫（台、中）、複合形式（台澎金馬、北上廣深）、修飾性

複合詞（真正的中國人、緬甸勞工、正港台灣人）、行動動詞（辱華、反共，僅限如「你們反共人士」）。

根據以下格式編碼以下語料：流水號順序 tab 理由 tab 編碼結果（小寫），如：980 包含地理身分「大陸人」，肯定「我們」的身分 a、111 無顯性身分標記 b、151 「我們」的行動的對象是地理身分「海南人」 b

（4）用於「你們/们」KWIC 樣本編碼的提示詞（Sonnet4, $\kappa = 0.869$）

研究背景：分析《玻璃心》（2021）YouTube 評論中的身分敘事，聚焦身分表達。以「你們」為關鍵詞的左三右七 kwic 集合而成的語料庫。

A 類（肯定或探討「你們」的身分）：

語句涉及「你們」與顯性身分標記（政治實體、地理身分、族群身分、政治身分、集體標籤）的確認或探討，聚焦「你們」的集體身分，涵蓋：

「你們」+ 顯性身分標記（案例：「你們中國人」「中國人,你們」「中國人是你們」、反例：去你們新加坡）。

「你們」+ 否定詞（「不是」「不認為」等）+ 顯性身分標記，否定「你們」的身分（如「你們不是中國人」「你們不認為自己是中國人」）

他人表示「你們」是顯性身分標記（如「誰叫你們共產黨」）。

他人表示「你們」不是顯性身分標記（如「我看你們不是中國人」）。

他人質疑「你們」的顯性身分標記（如「你們說我們不是中國人？」）

關鍵：必須包含顯性身分標記（如「台灣人」「小粉紅」），且聚焦「你們」的身分確認或探討。

B 類（不涉及肯定或探討「你們」的身分）：

語句完全無顯性身分標記（「你們深入不到那種地方」），或「你們」

的行動的對象是身分標記（案例「你們要武統台灣」反例「你們中國要武統」）。

明確規範：若無顯性身分標記，一開始即歸為 B 類，避免推斷隱含身分（如「語言」暗示「華人」，「立場」暗示「統派」）。

身分標記定義

政治實體：

範例：美國、台灣、中國、中共等（含縮寫：US、台、中、ccp）。反例：國家、政府、祖國。

地理身分：

範例：歐美、大陸、香港、台北人、中國台灣省、馬來籍、越南籍、牆內、牆國等。反例：國外、海外、地區、地方、山上、島內。

族群身分：

範例：華人、漢人、客家人、少民等。反例：民族、同胞。

政治身分：共產黨、KMT、民進黨、統派、獨派、韓粉、馬列子孫等（正式立場或政黨認同）。

集體標籤：

範例：五毛、小粉紅、韭菜、塔綠班、台蛙、1450、牆內人、中華膠、反賊、玻璃心、境外勢力等。反例：老百姓、普通人、底層、群眾、人民。

- 變體：縮寫（台、中）、複合形式（台澎金馬、北上廣深）、修飾性複合詞（真正的中國人、緬甸勞工、正港台灣人）、行動動詞（辱華、反共，僅限如「你們反共人士」）。

根據以下格式編碼以下語料：流水號順序 tab 理由 tab 編碼結果（小寫），如：980 否定「你們」的身分「大陸人」 a、111 無顯性身分標記 b、151 「你們」的行動的對象是地理身分「海南人」 b

Negotiating Ethnic Chinese Identities in an Agonistic Theater: A Corpus-Assisted Discourse Analysis of YouTube Comments

Lu-Yen Ko*

ABSTRACT

Background and Objectives

Due to the Great Firewall — China's nationwide internet censorship system—mainland Chinese internet users rarely participate at scale in identity negotiations among ethnic Chinese communities mediated by major global platforms. In October 2021, Malaysian Chinese singer Namewee and Australian Chinese singer Kimberley Chen released their Mandopop music video Fragile on YouTube, satirizing China's radical nationalist netizens, termed the "Little Pinks." The video drew large numbers of mainland Chinese users who circumvented their government's firewall to comment on it, sparking disputes with overseas users over questions of ethnic Chinese identity.

This research treats the identity negotiation mediated by the Fragile comment section as a natively online media event with clear data boundaries. Its emergence, interaction, and meaning-making all unfolded within YouTube, constituting a relatively closed observable unit - namely, a corpus of over 200,000 comments posted under the Fragile music video. Accordingly, the

* Dr. Lu-Yen Ko is teaching at School of Journalism and Communication, Huaqiao University, Quanzhou, China. ORCID: https://orcid.org/0009-0004-6065-7749. e-mail: duress.ko@gmail.com.

research employs CADS to present these identity negotiation discourses and to chart a boundary map of their relations.

Literature Review

This paper reviews three bodies of literature. First, identity theory: drawing on social constructivism, identity is treated as a category shaped, reproduced, and mobilized in social interaction. Second, mediated collective identity: this body of literature traces research on national and ethnic identities mediated by print, radio, and television, highlighting the distinctive duality of online space relative to these mass media—namely, that this space can be both a site for negotiating identities across physical boundaries and a space of algorithmically segregated identities. Third, empirical literature on online ethnic Chinese identity negotiation: pointing out that the methods used in these studies, including surveys, in-depth interviews, and digital ethnography, as well as keyword-driven mining of large-scale online data, face limitations in addressing the simultaneously integrated and segregated duality of online space, and proposing a census corpus approach to circumvent the following limitations: (1) using natively digital data to capture negotiation processes directly, avoiding the limits of retrospective accounts; (2) treating the corpus as a population rather than a sample, sidestepping issues of sampling representativeness; and (3) collecting all comments from a complete event rather than filtering by keywords, avoiding omissions that keyword selection entails. The Fragile video thus offers a rare opportunity to address the dual methodological and field-related gaps in the literature on online ethnic Chinese identity.

Research Question

Theoretically, integrating media event theory and mediated ritual theory, this study treats the video as a grassroots conflictual media event viewed as an

improvised agonistic theater that lacks a rigid script yet retains the theatrical structure of role, scene, and performance. Methodologically, as natively digital data, the comments under this single YouTube video constitute a bounded, observable census corpus, allowing researchers to observe the articulation, contestation, and reorganization of discourses without recourse to external data.

Preliminary analysis further shows that the Fragile corpus contains a high frequency of personal pronouns and ethnic Chinese identity-related terms. Drawing on the positioning theory, which treats personal pronouns as key linguistic indicators of identity claims and boundary work, this distribution suggests that interactions in the corpus are organized around self-identity claims and the ascription of identities to others. Accordingly, this study asks the following questions.

RQ: What statistically representative patterns of ethnic Chinese-related identity reference appear in the Fragile video comments? How do participants use these patterns to contest identity boundaries within the agonistic theater?

Method

This research employs CADS, introducing Large Language Model-Assisted Content Analysis (LACA) as a bridge between corpus statistics and discourse interpretation, addressing the statistical storytelling risk endemic to conventional CADS, wherein the statistical patterns identified in the corpus are representative, but the samples researchers extract to interpret these patterns are not. The procedural steps are: (1) corpus preprocessing; (2) word frequency and collocation analysis; (3) extracting keywords-in-context (KWIC) entries using first- and second-person pronouns as search terms; (4) drawing random samples independently for each search term based on a 95% confidence level and a 3% margin of error, yielding a total of 3,567 entries; (5) binary coding of the sampled entries by a large language model (LLM) to

identify A-samples that claim self-identity or ascribe identity to others in ethnic Chinese-related terms, yielding 937 A-samples with inter-coder reliability Cohen's κ ranging from .869 to .979; and (6) focused coding and relational analysis of A-samples based on grounded theory to identify narrative trajectories within the agonistic theater, then returning to the full corpus for cases reflecting who is arguing with whom, about what, and how.

Findings

The findings show that the performances within the improvised agonistic theater all take a standard identity narrative - a Chinese identity that opposes Western imperialism under the leadership of the Chinese Communist Party (CCP) party-state - as their reference point. In this discursive space, other identity narratives are not merely juxtaposed with the standard narrative, but are recognized, responded to, and evaluated in relation to it. Alternative identity narratives are deconstructions and recombinations of the elements the standard narrative asserts - ancestry (Han Chinese), language and script (Mandarin and Sinographs), state sovereignty (the People's Republic of China), and political ideology (party-state ideology) - forcing the enacted boundaries of identity to surface in the tension between the standard narrative and alternative ones.

Discussion and Conclusion

From the perspective of the integrative identity negotiation theory, when incompatible values collide in a shared field, the resulting friction marks the enacted boundaries that separate participants. The Fragile agonistic theater can thus be read as a breaching experiment: through satire of radical Chinese nationalists, it forces the boundaries among ethnic Chinese communities into view. The boundary map produced here allows comments that appeared scattered or meaningless, including non-Chinese-language comments,

comments seemingly unrelated to the disputes, and superficially flippant verbal attacks, to be located within the map.

By introducing LACA as a bridge between corpus analysis and discourse analysis within CADS, this paper methodologically supplies the alignment procedure that conventional CADS empirical work often lacks, reducing the risk of statistical storytelling and offering future researchers facing similar problems with massive online corpora a path. This study does not predict trends in ethnic Chinese identity. Rather, it maps the discursive trajectories of one specific natively online media event, providing a comparative basis for subsequent research.

Keywords: Large Language Model-Assisted Content Analysis, identity negotiation, natively digital data, media events, Corpus-Assisted Discourse Studies, ethnic Chinese communities