



「從人文到數位人文」 跨領域對談學術研習營紀要

何立行*

因應數位時代知識微縮革命的推進，面對各式各類的學術資料庫與大量的數位文獻，以數位文本作為研究的文獻類型，在「人文研究思維」結合「數位推論技術」下所形成的「數位思維邏輯」，無疑開啟一種新的人文研究範式。在以人文研究為主體並結合數位運算技術的數位人文研究方法，或許是為人文研究提供了一種新方法與新視野，也可能帶來人文研究範式新的轉向。為此由科技部人文司中文學門召集人廖美玉的推動規劃，邀集中文學會理事長葉國良教授、科技部「觀念·事件·行動：中國近現代觀念形成與演變的數位人文研究」整合型計畫主持人鄭文惠教授、臺灣大學資工系項潔特聘教授與政治大學中文系主任陳逢源教授共同籌辦「從人文到數位人文」跨領域對談學術研習營，於2014年12月5、6日兩天假政治大學行政大樓七樓第一會議室舉行。

藉著這個難得的機會，資深文史學專家、資訊工程與統計學專家，以及正在開發數位人文研究方法的學者和有志從事相關研究的年輕學者和研究生得以齊聚一堂，探討此一學術新疆域的發展現況，分享彼此對運用數位工具和方法從事人文研究的實作心得、審慎思考與樂觀期待。為使討論能聚焦而深入，研習營規劃從「知識微縮革命與數位人文研究」、「文學抒情／敘事結構的數位人文考察」、「思想觀念的數位人文考察」、「研究實務操作」、「從人文到數位人文：人文研究新方法與新範式」等五大主題，探討知識微縮革命與大數據時代下，全球數位人文語境的開展，以及數位時代下對人文研究的挑戰與因應之道。面對數位史料的海量資料，數位人文的優勢何在？數位人文如何有效提高人文學者論述的證據力道？如何幫助人文學者於海量資料中

* 國立清華大學中國文學系博士後研究員

迅速掌握問題核心？又如何進行跨領域研究合作與交流？希望透過本研習營，可以重新測繪臺灣人文學界在國際漢學研究的新位置。

「知識微縮革命與數位人文研究」第一場由中央研究院數位文化中心副召集人兼歷史語言研究所副所長林富士主持並兼任對談人，臺灣大學資訊工程系特聘教授項潔主講，另一位對談人是成功大學中國文學系特聘教授王三慶。項教授以〈唐宋類書與帝國的知識想像：一個數位人文的觀察〉為題，分享多年來以數位方法研究唐宋類書的心得和所遇到的困難。項教授認為，類書具現的是成書當時知識界的知識結構與知識想像，藉由數位工具，我們得以處理類書巨量龐雜的資料。項教授舉《藝文類聚》和《太平御覽》的比較研究為實例，展示如何與文史學者合作，建置類書分析資料庫；並運用分析資料庫比較這兩部類書部類的變化、條目的長短詳略等，論述唐宋兩代世界觀的變化。根據比較結果，提出(1)宇宙觀的改變：符命、祥瑞部的消解；(2)宗教觀的變化：道教地位提高佛教地位下降；(3)「異己」的體系化：四夷觀念演進等三項觀察。項教授並藉著這次研習營的機會，宣布「類書對應查詢系統」(<http://leishucis.digital.ntu.edu.tw/>)與「《藝文類聚》、《太平御覽》資料庫」(http://thdl.ntu.edu.tw/L303_YiWen_YuLang/RetrieveDocs.php)兩套由臺灣大學數位人文中心建置的資料庫正式上線，公開接受全球研究者申請免費帳號密碼使用。

首先回應的是中文學界運用電腦從事研究的先驅王三慶特聘教授。王教授回顧過去以人工方式製作圖表研究敦煌學的辛苦，對數位人文研究的最新發展表示驚喜。但他也指出，就細節而言，資料數位化過程中造成的訛誤、出處來源誤植等，必須更仔細校訂，以避免錯誤的資料影響機器分析結果。從大處著眼，則必須充實背景知識才能掌握研究方向。舉例而言，類書的部次編排方式和蒐羅內容與其編纂目的密切相關。宋代編了四部類書，各有目的，如《文苑英華》收的詩文，《太平御覽》通常就不收了。單單比較《藝文類聚》和《太平御覽》的部類恐怕無法逕論知識結構轉變。以條目長短詳略論世界觀變化的問題則是：類書編纂時期因受限制，手抄本時常有所省略，以數位方法研究類書必須更審慎。

主持兼對談人的林富士教授接著回應，認為項教授的研究量化分析做得不錯，但對文獻的解讀則需要更細緻。最關鍵的問題是文史研究中最強調的「脈絡式閱讀」。符應、祥瑞部的消解反映的可能不是不同時代宇宙觀的差



異，而是兩部類書的編輯及預設讀者的差異。放回歷史脈絡，宋太宗因得位不正，忌談天命、壓抑符瑞，《太平御覽》的編輯當是受此影響。太宗之後，宋代君臣講符瑞者其實有增無減。四夷觀念的改變和民族主義的萌芽，在杜佑《通典》中已有呈現，未可視為新現象。宋太宗一登基便剃度十七萬僧人，大興佛教，佛、道勢力的消長的觀察，顯然也與歷史事實有出入。

第二場對談由中正大學中國文學系主任賴錫三教授主持，南京大學藝術研究院副研究員、美國萊斯大學(Rice University)趙氏亞洲研究中心訪問學者陳靜代表不克親自出席的萊斯大學歷史系教授湯尼·白露(Tani E. Barlow)，以〈“Ephemera: Current and Future”及其相關〉為題，介紹萊斯大學「中國商業廣告檔案庫(1880-1940)」(Chinese Commercial Advertisement Archive, 簡稱 CCAA)的建置和相關研究問題。擔任與談的兩位學者為中央研究院近代史研究所副研究員張哲嘉及中正大學中國文學系教授黃錦珠。CCAA 是魯斯基金會資助的「蜉蝣計畫」(The Ephemera Project)下的子計畫，將建置一個收錄廣告圖像和相關資料的數位資料庫，預計 2015 年夏天完工。目前已經完成約 45,000 張報紙版面圖片的數位化，以及 40,000 多張廣告圖片元數據(Metadata, 或稱「後設資料」)的輸入。陳靜副研究員指出，報刊廣告的蜉蝣性為建置資料庫帶來一些問題，也引發許多思考。數位化本身無可諱言是一個捨「原真性」(authenticity)取「文本性」(textuality)，捨「物質性」媒介取「文本性」信息的選擇。正因如此，CCAA 力圖在展示廣告圖像的同時，以元數據的方式紀錄信息。然而，元數據的建立相當具挑戰性。CCAA 元數據包括：(1)關於廣告內容的「描述性數據」(descriptive data)；(2)與廣告商品相關的「語境信息數據」(contextual information data)；(3)與廣告刊報紙相關的文獻數據(bibliographical data)；(4)數位化過程中產生的技術數據(technical data)。所有內容識別都採人工標註，主題詞和內容描述更是倚賴人力。目前已有的主題詞並不能涵蓋廣告圖像中的所有訊息，這部分對於廣告圖像研究而言是很大的問題。

張哲嘉副研究員從理論的角度回應報告，認為「蜉蝣」的提出正反映西方學者在建構理論時，往往能先忽略細節，採取獨特切入點，抽取概念。西方的抽象理論因而得以應用於全世界。黃錦珠教授則關注資料詮釋問題。廣告常以女性為主角，主角就代表具有主體性嗎？經過一百多年，我們看這些廣告的角度是不是跟當年接近呢？關於廣告文案的文藝手法是否還有許多有

待開發的研究，如真實與虛構的析辨。為什麼香菸廣告總是使用西方女性圖像？面對大量資料，不同的詮釋，常常會產生南轅北轍的結果。

第三場主講者為臺灣大學臺灣文學研究所所長黃美娥教授，由中央大學中國文學系李瑞騰教授主持並兼任對談人，另一位對談人為政治大學臺灣文學研究所講座教授陳芳明。黃教授以〈數位資料庫與臺灣文學研究新趨向〉為題，從新興學科本土化、國際化、趨新性等特色談起，指出臺灣文學作為一門新近體制化的學科，在方法論上深受西方文學／文化理論和跨領域研究的影響。隨著全球數位化時代的來臨，研究者獲得了許多新工具、新史料，發掘出許多新問題，新的研究領域也正在形成中。數位科技累積大量資料，便於研究者快速查閱，善用資料能突破原本人力無法到達之處。資料庫的運用正在使研究習性發生變化，從原本的「理論依賴」逐漸轉向「資料庫依賴」。報告中列舉臺灣文學相關的國內外資料庫。在研究方法上，展示除了資料檢索外，可以使用的各種計量研究方法，或激發如 GIS、人際網絡、社群分布、詞綴、作者風格量化分析等新嘗試。黃教授並舉出個人經驗，分享資料庫如何增益自己過去的研究，而資料庫的運用也使跨時段、跨地域、跨國界的研究日益方便，促進學術社群跨界連結、交流與合作，擴大臺灣文學研究的邊界。最後，黃教授呼籲建立大型「臺灣文學作家作品資料庫」，讓「資料」取代「理論」，作為國際交流的平臺。

李瑞騰教授對「資料庫依賴」的說法頗有疑慮。他以作家資料庫為例，指出作家的認定與選擇，資料收集如年表製作、作品頁數認定等，都尚未標準化，導致資料庫在細節上常有很多問題。他認為，黃教授不僅嫻熟資料庫的駕馭，自身的學養也足以彌補資料庫的不足；而一般研究生若無足夠學養，使用資料庫可能會做出一些奇怪的結論。陳芳明教授也同時呼應，認為數位對人文研究衝擊之巨大，令人難以想像；但再怎麼資料化，都比不上閱讀。不能太信任數位方法，就像不能太信任理論。理論只能幫助我們發現問題，解決問題要靠研究者用自己的興趣大量閱讀。研究一位作家必須做全集式的閱讀，數位資料只能作為輔助。針對兩位對談人的疑慮，黃美娥教授澄清「依賴」主要是放在國際化的脈絡下來談，這些年「理論」一直是臺灣文學研究者與國外對話交流的平臺，她嘗試提出的是，「資料」是否也可以作為國際交流的重要平臺。

第二個主題「文學抒情／敘事結構的數位人文考察」，第一場由銘傳大學



資訊工程系謝育平副教授主講〈同位詞夾子：主題式分類詞庫萃取演算法〉，由中央大學中國文學系康來新教授擔任主持人，臺灣大學中國文學系教授歐麗娟與臺灣清華大學中國文學系副教授祝平次擔任對談人。「同位詞夾子」是謝教授發明的一種數位工具。同類詞彙通常會出現在句子的特定位置，具有很高的同位性。「同位詞夾子」由「前文、前綴、中綴、後綴、後文」五個部件組成，用以在文件中萃取同位詞。執行時先利用「種子範例」掃描文件以產生某類詞彙的詞夾子，再利用產生的詞夾子掃描文件、夾出該類詞彙的「候選詞」，然後以人工確認「候選詞」是否與「種子範例」同類。確認的詞彙納入擴充的「種子範例」，再度啟動演算法。如此互動循環，直至滿意為止。謝教授展示了幾個詞夾子成功萃取的例子。平均而言，從一個百萬字文庫中萃取一個分類詞庫可以在兩小時之內完成。

對談人歐麗娟教授認為資料庫可以讓論點的前提堅固，減少漏洞；但多半作為一種索引，真正的功夫在於之後的細讀。有了精讀對文本和問題意識的掌握，加上科技的幫助，才能滴水不漏。她覺得詞夾字的設計最動人的是利用人工。歐教授對此提出了兩個問題：一是語法規範會因時地而異，如果地名不在「到……去」裡，而用了別的語法結構，是不是就無法被夾出來？針對這一點，謝教授回應：詞夾子是根據文本本身的用法造出來的，不需要預設句構。第二個問題是：事件有沒有辦法夾出來？例如，從動作來看林黛玉的形象，林黛玉有一些特定的舉止，是別的女孩沒出現過的動作，如摔、啐、蹬等，集中在小說前四十回，程式有沒有辦法找到這些動作特徵？謝教授表示：事件無法夾，除非文本全面 tag。另一位對談人祝平次教授指出，謝教授的報告觸碰到數位人文的核心問題。過去我們從不懷疑我們看世界的方式，電腦卻提供另外一個看世界的角度。我們怎麼教會電腦做人會做的事，又如何能從機器看待世界的方式學到新東西？這是對人文學科知識基礎反省的契機。祝平次教授認為詞夾子如果可以做成一個大家都使用的應用程式（APP）應該很有意思。至於詞夾子的精準度，謝教授回答：詞夾子的召回準確率很難有一個具體數字，因為一開始一定很差，在運作的歷程中，準確度卻會越來越高。主持人最後總結表示讓「賽先生」來扮演協助人文學者的角色是一個很好的發展。

第二場由政治大學資訊科學系特聘教授劉昭麟主講〈語文分析方法在數位人文的一些應用——《紅樓夢》、《西遊記》及其他〉，由政治大學中國文學

系特聘教授高桂惠主持並兼任對談人，另一位對談人則是政治大學中國文學系博士後研究員邱偉雲。劉教授首先分享自己近年與人文學者協作過的一些語文分析研究，包括利用「近現代思想史專業數據庫」做的各種主權、官制、立憲、選舉、平等關鍵詞叢的分析，以及音譯詞彙的研究。他歸納語文分析的基本步驟為：定義問題、選定初步關鍵詞彙、讓電腦擷取相關文獻。比較進階的則包括詞類標記、斷詞處理等等技術。劉教授舉「《紅樓夢》裡誰最愛笑？」這個問題為例，說明比例分析有時比頻率分析重要。又示範如何用數位語文分析方法知道「寶玉與金陵十二金釵關係遠近」——統計在 40 個字的範圍內，同時提到寶玉跟十二釵之一的次數——得出關係最近的依次是寶玉與黛玉、寶玉與寶釵、寶釵與黛玉、寶玉與湘雲。此外，又展示了「《西遊記》裡哪一些妖怪要吃唐三藏」、「《西遊記》妖怪戰力排行榜」的量化研究。

高桂惠教授回應時指出，中國文學系的強項是咀嚼文本，跟「腦筋迴路被數位開過的」想法很不一樣。十多年前她開始從事跨領域研究時，跨的領域是文、史、哲、社會學一起合作研究情慾與禮教，從中看到各學科視野的局限。現在跨的是數位與人文，不同的思維方式間需要更多的溝通。邱偉雲博士則從數位人文研究者的角度，分享他與劉昭麟教授長期合作獲得的啟示，希望能扮演資工與人文學者之間橋梁的角色。開放觀眾討論時，參與十分熱烈。祝平次教授提醒大家數位人文並不僅是量化研究。若能把斷句細讀跟標記文本結合，不僅達成精讀訓練，標記的文本還可以嘉惠後人。王三慶教授分享自己做過的緊鄰字相關性研究呼應劉昭麟教授的報告，並期待未來能有更多進展。此外，他建議建立資料時能多留意版本問題，並指出已有學者利用資料庫做版本比對工作。

第三場〈唐詩的數位人文研究〉，由逢甲大學中國文學系教授兼科技部中文學門召集人廖美玉主持，政治大學中國文學系教授鄭文惠主講，臺灣大學中國文學系教授蔡瑜、中央大學中國文學系助理教授李宜學擔任對談人。鄭教授從唐詩資料庫建置的歷程與方向談起，示範十二個唐詩數位人文研究個案分析，接著討論唐詩數位人文研究的可行性與可能性，並提出未來的展望。伴隨著電腦技術的發達，全唐詩的檢索系統已從「字符串」檢索提升成為具有多功能、便於交叉檢索與分析的系統。最新的發展則是對本體庫的重視。本體庫考慮概念體系，從語義標記著手，進行概念分類，架構知識體系。以語意概念分類所建立的檢索功能，可以檢索到字形不同但語義相同的資料。



鄭教授談到數位人文研究的 QMTA 合作路徑：從提問 (Question)，提出傳統研究方法或議題；到方法 (Methodology)，將傳統研究方法和議題轉化為量化研究方法和議題；再引入科技 (Technology)，編寫方法程式、計算研究結果；最後是應用 (Application)，應用新方法、提出新問題或新詮釋。在這個合作模式下，目前適合唐詩利用的數位人文研究方法包括關鍵字詞、詞的網絡、意象與意象叢、詩人社會網路、文學地理、事件與作品關聯、典故與隱喻、詩派風格及傳承、詩歌與思想觀念、詩歌與時代脈絡化轉向等等，都可以借重數位人文研究方法繼續開展。

蔡瑜教授回應表示，數位工具彙整資料的功能超過人工，但如何運用資料則視研究者的學術關懷，研究者本身有其主體性。詩學有其特殊性，就詞彙、語義方面，無法依賴資料庫。以同位詞夾子的使用為例，「到……去」的「到」、「去」二字在詩歌中往往省略，必須從前後文語義去揣摩。詩人又會不斷創造新詞彙，很多技術在這種狀況下會遇到一些困難。不過，一些過去不敢問的問題，現在的確可以藉由搜尋資料庫初步看看有無展現。蔡教授舉陶淵明為例，為何陶淵明總是說「園田」而非「田園」？一經查找，發現這兩者的確不同，而著名的〈歸去來兮辭〉裡的「田園將蕪」，最早在《宋史》中出現時也是「園田」。數位工具給我們很大的幫助，但要從語意、意義去切入，還有很大的困難。再者，或許我們該更重視韻律的研究。韻律因為需要大量統計，過去很少人研究。北京大學開發的全唐詩分析系統雖然可以把平仄分析出來，但會受限於它的格式設定。楊秀芳教授開發的漢字古今音資料庫可以做很多分析。不過，這是為語言學研究設計的，詩歌研究者必須用人工一首一首貼上去。如果能把詩歌文本資料庫跟漢字古今音資料庫結合起來，提供判讀系統，就可以用來統計，研究詩歌格律發展史。如果可以這樣研究，應該可以進一步建構韻律美學作為支撐，而非只是格律規範。第二位對談人李宜學教授則舉出三個數位研究案例回應：首先是數位人文研究先行者羅宗濤教授的唐詩研究。羅教授當年利用大陸兩套數據庫，做了許多主題、意象的量化研究。李教授認為資料收集可以利用數據資料，但觀察聯想、情趣與蘊含意義仍有賴人文學者的功力。第二個案例是武漢大學做的「唐詩宋詞排行榜」，用統計的方式，做了加權、標準化處理，探討讀者的關注程度。研究者很自覺地表示，其結果僅能反應關注程度，不能代表藝術成就或思想高度。第三個案例則是北京大學的中國古代聲律分析系統。這個系統採用人工提前

干預的方式，又增加詩歌文本校勘的功能，首次解決了四聲跟平仄標註及一字多音的問題。

第二天進入第三個主題「思想觀念的數位人文考察」。首先由政治大學中國文學系博士後研究員邱偉雲偕同政大中國文學系碩士生莊勝涵主講〈先秦儒道思想論著文本結構與觀念系統的數位人文考察——以《論語》、《孟子》、《荀子》、《老子》、《莊子·內篇》為中心〉。邱博士指出，思想領域的數位人文研究可以分為三階段，最早是紙本逐字索引研究；第二階段則是系統的逐字索引與數量統計研究；現在進入第三階段，發展中的是概念共現和概念分布並用。以往的檢索邏輯得出的字頻如散布四處的點狀資料，未來的研究邏輯則是希望將字頻的分布結構、各字詞關係結構加以連結，使研究者得以分析運用。邱博士舉《老子》的字頻為例，加上標準差運算可以看出特定字詞在八十一章中是遍布或集中出現。如果再比較上下經關鍵字的辛普森指數分布圖，則可看出上下經用字差異。此外還可利用CUSUM圖，找出在全書中穩定成長的基本觀念詞。電腦亦可快速比對資料，找出上下經各自的獨有用字。接續邱博士的報告，莊勝涵展示如何用統計方法進行《論語》、《孟子》、《荀子》三書文本結構的承轉分析；又比較了《老子》與《莊子》內篇的論述結構與觀念系統；最後再展示先秦儒道思想論述結構的異同。邱博士最後做總結，討論以數位人文方法進行思想研究的偏、執、得。

對談人陳逢源教授認為經學研究的確適合做數位分析，細節在比對之下顯現出來，也開展出跨越先秦不同思想流派的大視野。不過，他提醒報告中還有不少值得推敲的部分。如少說不一定不重要，多少與重要性的對應必須釐清。再者，必須留意資料可能會導致論述歧出，無所依歸。在構思研究時必須重視傳統研究成果。「亂」字可以訓為「治」，統計時也需要區隔。其次，並列先秦各家學說加以討論時，也需要將時間軸納入思考。另一位對談人蔡振豐教授偏向於將數位方法視為工具。他認為電腦無法掌握文本意義，中文詞彙的歧義會影響統計結果。對統計資料進行人為判讀牽涉到研究者的閱讀量多寡和詮釋能力。不過，就工具而言，研究工具當然是越多越好，在研究遇到瓶頸時可以使用。希望科技部能建立一個平臺，提供分析工具，公開研究成果，讓大家可以避免重複，並互相學習研究方法。開放現場討論時，祝平次教授提醒習慣傳統研究方式的學者：電腦也是可以學習的，而以資料驅動的研究方式跟過去的想法也與傳統思維方式有別。



第二場由政治大學統計學系余清祥教授以〈人文基因圖譜：隱喻或方法〉為題，談統計在數位人文研究中的應用。對談人為政治大學英文系特聘教授賴惠玲及美國萊斯大學歷史系講師趙沈允。主講人余教授從「資料驅動」(Data Driven) 形成新的「數位思維邏輯」談起。表示運用數位技術分析大量文獻，探索新的知識議題，或可視為測繪「人文基因圖譜」。如何測繪？從統計學習理論出發，有監督學習與非監督學習兩種。前者以給定的測量值(變數)預測目標變數，常用迴歸分析等統計方法；後者則是不給定目標變數，但尋求變數特性及之間的關連，常用多變量分析。余教授接著舉中國近代思想史研究的案例，談統計如何協助人文學者從文獻數量龐大的字詞中，偵測「潛在關鍵詞」，而後由人文學者去除不相關的關鍵字，再引入「生物多樣性」及「物種演化的概念」，進一步篩選出核心關鍵詞。對談人賴惠玲教授表示，統計是進行數位研究不可或缺的工具。從語言學的角度來看，意義必須是鑲嵌在歷史社會和文化脈絡中才能產生。另一位對談人趙沈允教授則指出，伴隨數位時代到來的是一個重新建構知識的新契機，但當每個方程式背後都有許多理論和假設，卻沒有統一標準時，人文學者難免會對資料如何被對待有所疑慮。如何能讓統計工具與方法透明化，無疑是數位人文研究的重要課題。

第三場由北京師範大學文學院特聘教授方維規主講〈概念史的理論、方法和實踐〉，臺灣師範大學國文系陳麗桂教授擔任主持人，對談者為中研院近史所副研究員潘光哲和日本關西大學外國語學部教授沈國威。方教授從「概念史」的來龍去脈談起，主要介紹德國概念史研究的起源和發展，透過「鞍型期」的理論來呈現概念史的特點和追求。方教授強調，概念史十分關注語境，概念是一個代號，是思想的出口。他評述了德國概念史模式從思想史研究向思想的社會史研究的轉向，並進一步闡釋概念史的理論依據和主要特徵，討論概念史研究與劍橋學派對歷史語義學的不同理解及其分析語言與歷史之關係的不同切入點。接著談到概念史研究必須具備的學術功底和學術準備和概念史的新近發展與國際影響，最後簡要談論概念史中依然存在的疑難問題，並對近現代西學東漸與歷史語義學研究提出看法。

對談者潘光哲副研究員表示，概念史的論述宏大，但在研究時需要「趴在地上研究」的功夫。他舉 president 一詞為例，說明很多相關論述因為音譯選字不同，根本無法在資料庫中查到。另一位對談者沈國威教授認為語詞是思維的外殼。近代概念史是西方新概念的接受史，東亞的新容受是通過漢字

來實現的。新詞的理據反應了造詞者的理解和主張，近代詞彙研究與概念史研究相得益彰。就數位人文研究而言，沈教授提到曾用數位方法做過《遐邇貫珍》等資料庫，以及查詢資料庫可看到現代漢語常用抽象詞高頻出現在1904年，因為科舉取消，知識菁英都到日本速成，學了很多雙音節的詞彙。1904到1920年可能是漢語詞彙的鞍型期。方教授回應並補充，在這次的會裡一直看到做數位人文的人並不排除細讀，但是常常下的結論都像是沒細讀過的人寫出來的。有些涉及到理論的問題，例如題目的價值：我們想透過數據庫發現什麼東西？有些文章看起來像做到一半，可能是思想高度的問題。

第四個主題「研究實務操作」，由中興大學助理教授陳碩文擔任主持，三位數位人文研究構想書發表者為南京大學藝術研究院副研究員暨萊斯大學趙氏亞洲研究中心訪問學者陳靜、逢甲大學中國文學系博士候選人李堯涓、臺灣大學中國文學系碩士生熊蘭祺，對談人為政治大學中國文學系博士後研究員邱偉雲和臺灣清華大學中國文學系博士後研究員何立行。陳靜副研究員的構想書以〈概念網路：從中國商業報紙廣告到中國近現代社會學思想〉為題，從「廣告作為消費文化中心並具體呈現本土化的現代社會思想」這個觀察出發，嘗試結合「中國近現代思想及文學史專業數據庫（1830-1930）」、「晚清民國期刊全文數據庫」及「中國商業廣告檔案庫」三個資料庫，一方面對「中國商業廣告檔案庫」中的元數據如概念、術語、主題詞等進行分析；一方面比較廣告全文中的概念與「中國近現代思想及文學史專業數據庫」、「晚清民國期刊全文數據庫」所搜尋到有關社會學文本中的概念之間的關係，考察廣告與社會學文本中的概念內涵與外延之間的同異之處。李堯涓有鑑於學界過去多半都以個案材料論證寒門之於整體社會的實態，對寒門之士的界定也仍有歧義，希望借重數位人文研究方法爬梳數量龐大的文本，對「寒門」詞群做歷時性研究，觀察相關詞彙出現的時間點與用法，探討由漢至唐的發展脈絡與轉變現象。熊蘭祺聚焦於徐志摩與民國早期畫家的交遊狀況、參加藝術活動的情形和藝術評論書寫，希望透過資料庫搜尋重建徐志摩與民國早期美術界的互動，重新思考徐志摩如何理解和實踐藝術的社會功能，以及如何定位藝術家在社會改造中扮演的角色。邱博士針對三人的構想書分別提出關鍵字切詞、關鍵詞叢、共現詞彙、文字雲等數位人文研究技術方面的實質建議；何立行博士則就研究素材的選取和數位方法的採用是否能緊貼問題意識達成有效的研究提出疑問。



壓軸的是第五個主題「從人文到數位人文：人文研究新方法與新範式」圓桌討論。主持人廖美玉教授首先談到這兩天大家一直在搶時間的現象，說明人文與科技跨領域對談的迫切需求性。她以動畫中的老鷹為例，每一根羽毛都是由技術導演以「一串串極其複雜的公式完成」來說明數據之美，以及其對人文研究的衝擊與挑戰。方維規教授表示計算機發展之速，大家都有共識。目前計算機幫他做了很好的資料收集工作。回應鄭文惠教授說概念史研究有主觀性在，方教授說那就是在考驗史家的功底。沈國威教授分享自己研究嚴復，就為嚴復建立一個語料庫的研究方法。趙沈允教授同意數位方法的重要性已經是共識，也再次強調統計工具與方法需要透明化，才能免除人文學者對研究方法的焦慮。劉昭麟教授說電腦速度快，能夠做的事也很多，資工學者面對人文研究者常講詞頻，只是因為好理解。人工智慧進步神速，我們所用的電腦，速度是三十年前的一百萬倍。當然，電腦不是一步到位，如果告訴電腦我們進行圓桌座談？會場卻找不到圓形桌子，電腦可能就會困惑，這就是我們需要想辦法的。此外，他特別提出，很多想進行數位人文研究的學者缺乏數位文本，科技部應該以國家的高度來解決這個問題。祝平次教授接續電腦發展迅速的主題，說人與機器可以互相學習，我們怎麼知道我們的思維有沒有受到束縛？我們怎麼知道自己不是知識的色盲？祝教授認為，理想的數位人文研究是人文學者應該一直丟問題給資工學者，打破彼此的知識色盲狀態。鄭文惠教授認為知識突破跟研究突破需要巨大的能量去進行創造性的轉換。

回顧研習營提到的幾個議題，學者如何各自有主體性而展開跨領域的研究協作？建置資料庫的知識本體庫須思考全球可以互相嫁接轉換。數位人文合作的學者各有關注的主題與研究專長，如何達到更好的協力合作與學術發展，是可以繼續努力的。尤其如何釋出相關的數位文本以推進數位人文研究的進展，奠定臺灣學界在國際數位人文研究上的地位，有賴科技部以國家的高度予以協助與解決。經過兩天激烈精彩的觀念激盪和意見交流，「從人文到數位人文」跨領域對談學術研習營圓滿落幕。期間提出的研究問題，擦出的火花，都將轉化為這個新興領域繼續發展的動能。