

什麼是好的數字人文*

姜文濤

在 2000 年左右,“數字人文”這個術語開始進入人們的學術視野和話語之中。作為一種將計算、統計和人文闡釋結合在一起的方法,其包涵的內容非常寬泛,甚至可容納媒介文化研究、軟件研究、圖書館博物館的基礎設施建設、數字閱讀與寫作平台、數字教學等。這使得數字人文必然具有跨學科的特徵,有人認為它會帶來某種人文社會學科內部及與理工學科之間的融合。但數字人文的發生發展,也一直伴隨著各種爭議和質疑。有人指出,人文學科尤其是文學和藝術的研究,依賴於研究者的理解力、同情力、想象力和現實感,而量化研究排除了研究者的主觀性,這會導致“人文”的喪失。還有人認為,數字人文採用量化研究方法迎合量化行政管理,爭取到更多學術資源,從而影響了傳統人文學科的發展。其他的批評還有,數字人文研究未取得劃時代的成果,缺乏學術研究的深刻性和歷史感,只是傳統文本研究的延伸和點綴,等等。與此同時,數字人文的概念也在不斷擴大和模糊化。

在此背景下,我們更應當以積極而又審慎的態度來推動數字人文的發展。筆者並不想急於為數字人文下定義或把某些研究排除在數字人文之外,而是認為,與其追問“什麼是數字人文”或“什麼不是數字人文”,或許更好的提問方式是“什麼是好的數字人文”,以及“如何做好的數字人文”。

我們並不確定數字人文會佔據什麼樣的知識空間,以及會產生什麼樣的問題,但相信好的數字人文研究最終能夠廣續和增補人文學術的傳統。無論是從長時段的人文學的脈絡中定位數字人文,還是從具體的方法和研究出發得出具有人文學普遍意義的分析,這兩個層面上的討論都有助於思考“什麼是好的數字人文”,進而鞏固數字人文研究的基礎。一方面,從學科史的視角考察人文社會學科發展的內在邏輯,可以探究數字人文方法促進人文社會科學研究的可能性;另一方面,數字人文作為研究方法和工具的特殊性,使之具有在人文社會科學各個學科之間搭建跨學科橋樑的可能性。前者是數字人文的縱向拓展側面,後者是其橫向擴展側面,從而構成一種“十”字面向發展的數字人文研究。通過縱橫兩個方向的檢視,我們可以判斷出真正具備人文社會科學學術意識、順應學科發展趨勢的研究,這些才是健康可持續的數字人文研究,也就是好的數字人文研究。筆者不揣淺陋,以自己比較熟悉的文學和文化研究領域的數字人文研究為例,嘗試提出“好的數字人文研究”應具有的幾種意識和素養,以期引起相關的思考和討論。

(一) 人文學術意識

我們常聽到這樣的說法:數字人文時代來臨了,傳統的研究和闡釋方法會被淘汰,人文學術研

* 本文係國家社會科學基金重大項目“基於大數據技術的古代文學經典文本分析與研究”(項目號:18ZDA238)的階段性成果。

究將邁進新的紀元。筆者認為這是一種學術方法上的達爾文主義，它會扼殺人文學術研究的精神和靈魂，而數字人文在此會充當“特洛伊木馬”的危險角色。這是急需要避免的學術研究發展趨向。人文主義的歷史非常悠久，現代人文學科則主要是隨着近代印刷文化的興起而產生的一種知識機制。隨着數字信息技術的發展，印刷文化的基礎設施日漸與數字媒體的基礎設施交叉在一起。在這樣的背景下，我們需要繼承人文闡釋的悠久傳統才能繼往開來，才有可能進而創造出新的知識生產形式和社會文化形式。在英文中，“數字人文”中的形容詞“數字(digital)”之“digit”也是“手指”的意思，這就強調了其中“人”的因素。當然，這並不是說數字人文研究不可以提出新的有價值的問題。事實上，我們逐漸由 19 世紀以來印刷文化佔據主導地位的時代，過渡到數字文化佔據主導地位的歷史時期，日常生活和學術研究中的各項基礎設施都發生了相應的變化，這是數字人文應運而生的社會物質條件。一時代有一時代之人文學術。我們有充分的理由相信數字人文會提出有價值的人文問題、會以自己的方式推動人文學的發展。但需要強調的是，這些問題和發展並不會憑空產生，其基礎仍是人文學悠久的闡釋傳統。這是數字人文研究整體上的意義所在，是根本中的根本。斯坦福大學文學實驗室主任馬克·阿爾吉-休伊特(Mark Algee-Hewitt)教授闡述得很清楚：“‘文化分析學’(按：數字人文研究許多名稱中的一種)不是要用數學嚴謹性取代學者們數百年來發展出的闡釋技巧的虛擬人文學科，而是增強的人文學科，在最好的情況下，能展現最仔細的細讀讀者也往往忽視的新類型的證據和仔細考慮過的理論觀點，二者聯手產生新的批評研究。”(馬克·阿爾吉-休伊特，《文化分析學是增強版的人文學科，不是沒有闡釋技巧的虛擬人文學科》，濟南：《山東社會科學》，2019 年第 8 期)這種增強的人文研究，需要具備傳統人文學科的學術訓練以及數字人文基本方法訓練這兩個方面的準備工作。這是開展好的數字人文研究需要具備的前提意識。

(二) 社會科學素養

有一些學者將數字人文文學研究歸結為 20 世紀六七十年代以來文學研究社會學轉向的一個部分。(參見 James English, *Everywhere and Nowhere: The Sociology of Literature After "the Sociology of Literature"*, *New Literary History*, 2010, 41)這種學術史上的定位不無道理。數字人文研究的許多個方面的確是與社會科學密不可分的，比如統計學、心理學、社會學等。在這些學科中，量化研究與質性研究之關係的討論、方法和實踐已經相當多了。正如澳大利亞國立大學的凱瑟琳·伯德(Katherine Bode)所說：“文學的意義不是單一的，而是社會構成性地鍛造而成，在這個社會裡，特定時刻中特定範式(歷史的、哲學的、心理學的、社會學的，現在是統計學)的突出地位塑造了我們所知的一切和理解這一切的方式。任何情況下，學科的純潔性都無法保護貧乏的方法論，跨學科性能增加方法論意識。”(凱瑟琳·伯德：《需要對用機械方法研究文學數據、統計學和機器學習進行更多的批判》，濟南：《山東社會科學》，2019 年第 8 期)好的數字人文研究者應該具備一定的社會科學學術研究的素養，有意識地從這些學科汲取方法論上的能量，以防止停留於數字炫技或者是過度人文闡釋的層面上。當然，這並不意味著文學研究或者總體上的人文研究領域要成為社會科學方法的附庸或者是殖民地。相反，好的數字人文研究，應該對所研究的問題以及所使用的研究方法在本學科學科史上的意義有充分之瞭解。

(三) 學術史和學科史的視野

有價值的學術問題的組織和提出一定是具有學術史和學科史視野的，建立在對學術史上對該問題或相關問題已有討論的基礎之上。自 19 世紀以來，作為知識生產制度的科學技術、人文與社會科學成為大學裡不同的學科，它們所涵蓋的知識討論對象、範圍和方法各自不同，形成了彼此不

一樣的知識體系和傳統。1990 年代以來,跨學科研究和討論(尤其在人文學科方面)成為一種學術潮流,也產生了一批有價值的學術成果。數字人文具有天然的跨學科和學科融合的潛質。然而,我們需要注意到,所有成功的跨學科研究無不建立在本學科嚴謹的學科訓練的基礎之上。從學者學術生涯生存策略上來講,對學術史、學科史的重視也是很必要的。目前在數字人文研究領域活躍的多為年輕學者,如果他們的研究不獲得本學科內部的承認,很難想象他們能在本就壁壘森嚴、競爭激烈、考核嚴格的學術界生存下去。具體到人文學科中之文學研究方面,它與數字人文方法的結合尤其容易引起爭議:一邊是統計技術和量化方法的抽象,一邊是非常複雜的傳遞情感與美感的語言修辭。這種人文與技術的機械性對立是較為常見的。在最近的一場有關數字人文的學術論爭中,美國聖母大學笈章難教授就重提了語言修辭和闡釋的複雜性,並以基本統計原則從實證層面討論了計算文學研究中較為典型的學術作品,通過指出其中許多的技術問題、邏輯謬誤和概念缺陷,論證計算文學批評的方法論和理論前提並不適用於分析文學、文學史和語言學的複雜性。(笈章難:《以計算的方法反對計算文學研究》,濟南:《山東社會科學》,2019 年第 8 期) 如果没有人文學科研究方法和學科史方面系統的訓練,而只通過統計或計算技術進行炫技般的學術表演,那也許並不符合基本的學術研究標準,遑論數字人文研究。

(四) 數字方法和技能

遠在“數字人文”這個術語在 2001 年出現之前,在其前身“人文計算”方法之中,實驗的先鋒性就是一個很凸顯的特徵(關於“人文計算”的歷史,參見 Susan Hockey, *The History of Humanities Computing, A Companion to Digital Humanities*, Ed. by Susan Schreibman, Ray Siemens, and John Unsworth, Oxford: Blackwell, 2004),它發展出了一系列的研究方法。比如社會網絡分析、GIS、機器學習、可視化、情感分析等。這些方法是量化和模型的方式,是數字技能的體現。這是作為“增強版”的數字人文對於從事人文研究的學者提出的新的技能要求,即學習 Python 或 R 語言等編程方面的知識,或者對統計學的高階掌握。在理想的情況下,數字人文研究者應該自己學會編程語言,這是一種腳本編寫語言素養。在腳本編寫的時候,某個概念也會逐漸成形。這個過程是思索數字人文研究問題很重要的一個側面。如果研究者並不具備編程語言方面的技能,那他或者她至少要與具備這方面技能的人員有較為深度的學術合作。這是因為,不同的方法和技能常常意味着不同的問題意識,或者是處理問題的不同的路徑,往往會導致不同的研究結果。

(五) 數據意識

認為數字人文就是做數據庫,這是一種對二者都不了解的錯誤看法。實際上,文本與數據都是需要探討的理論問題,並非是不證自明的、可以簡單拿來研究和分析的“原材料”。正如並不存在原始的文本一樣,也不存在所謂的原始數據。文本和數據都會經歷文獻整理和校勘的過程。而且,信息數據化而成為數據庫的過程也許比文本化的過程還要更為複雜,它涉及所建立數據庫的存儲、訪問、擴寫、查詢、維護等各個側面。有些研究者認為文學作品就是文本,文本和數據是單獨的、穩定的、自足的實體;不重視文本和數據記錄資料的多重多樣性,包括忽略文獻學以及學術文獻校勘能對文本物質形態歷史所做出的貢獻。這種對文本和數據體系的有限的、抽象的、甚至是非歷史化的處理,忽視了它們自身的社會、經濟、機構和技術結構。提倡數據文本文獻校勘學的學者凱瑟琳·博德就反對這種對於學術研究材料的抽象和簡約,認為“數字化的學術研究材料,就如文獻目錄一樣,是具有闡釋性性質的建構,它們依然在變化之中,不僅僅是內容上面,也包括形式上面,在這個過程中為文學史研究提出重要的實踐和概念上的挑戰”。(Katherine Bode, *The Equivalence of "*

Close" and "Distant" Reading; or, Toward a New Object for Data- Rich Literary History, *Modern Language Quarterly*, Vol 78, 1 March 2017)“基礎設施研究”(Infrastructure Studies)正在成為一個人文社會科學的新研究領域和方向。“‘基礎設施’,即既能讓人類經驗得以實現,又能對這種經驗加以限制的社會及技術環境,如今和‘文化’概念本身一樣有着大體相同的規模、複雜性和普遍文化影響。”(參見艾倫·劉:《通往思辨的基礎設施研究》,濟南:《山東社會科學》,2019年第6期)好的數字人文研究,包括數字人文基礎理論研究,都需要具備對作為學術研究基礎設施的數據和數據庫的思辨性意識。

(六)反思性

任何學術研究都需要一定程度的反思性,這是體現該研究意義的時刻,即:它突出了、解決了、反映了什麼問題;所採取的方法具有什麼層面的有效性、有限性;該項研究在學術史上的位置和價值;等等。在“增強版”的數字人文研究中,尤其需要提倡學術研究的反思性。數字人文研究的優勢顯而易見,即它能大規模地、長時段地、以數字語料庫為基礎、採用量化或模型的方式來規劃和探討人文研究中的問題。與此同時,它在各個層面上也都容易受到來自傳統人文學研究者的攻擊:太實證主義、不够批判性細讀、不能體現人文主義的價值、體現了新自由主義式的項目化學術生產方式;等等。事實上,任何一項研究都具有局限性,並不存在任何解決了所有問題的、具備了上帝視角的研究,並沒有橫空出世、無中生有的重大的科學研究理論發現。知識的生產更多地是一種漸進的、不斷修正的進程。數字人文研究應該坦言本身進行研究的數據基礎設施前提、採用的方法和步驟,以及這些所帶來的局限性或者說是未來研究可以修正的空間。這才是一種誠實的學術研究品格,是學術研究者的基本素養,浮夸的、斷言式的學風是需要避免的。

(七)開放性與協作性

人文研究者常常具有一種身份上的克里斯瑪(Christma)。這常常體現在研究者似乎是靈光一現的、個人天才式的觀點上,或者研究者常以國家民族精神、文化傳統的代言人自居。按照澳大利亞學者伊安·亨特(Ian Hunter)的考察,無論是20世紀的新批評還是取代新批評的理論時代,在人文主義研究者那裡都存在一個非常明顯的教導的層面,即:(學生/讀者)觀察,(學生/讀者)辨認、認同,(人文主義批評者)糾正、管教,(人文主義批評者)例證、例示。(參見 Ian Hunter, *The History of Theory, Critical Inquiry*, 33, Autumn 2006)這使得啓蒙時代以來世俗現代性語境下的人文主義教育成為一個有關道德教育和自我成長的、享有特權的社會場域。好的數字人文研究應該是以其反思性、技能性和數據的公開性,敞開整個研究過程和步驟,展示可重複的人文研究成果,是一種遠為民主化的知識勞動。同時,因其規模龐大,一項數字人文研究常常需要多個研究者合作完成,這就需要各個成員放棄自己的文化偏見和過重的主體性,而具備開放與協作的特性。這是新的學術人格與知識倫理產生的過程。事實上,19世紀晚期德國歷史學家特奧爾多·蒙森及其領導的“大人文學科”,以及其同時代人稱為“工業化”模式的學術研究,就已經體現了這種知識生產的模式。(參見 Chad Wellmon:《忠實的工人和杰出的學者:大人文學科與知識倫理》,澳門:《澳門理工學報》,2018年第3期)

以上所羅列的七點,只是拋磚引玉,未必全都合宜。期待更多學者積極參與相關探討與實踐,共同審慎而樂觀地推動健康的數字人文研究,同時也促進人文學術整體的良性發展。

(作者係浙江大學國際聯合學院講師,博士)

[責任編輯 桑海]