

# 從史料量化到數位人文： 明清社會經濟史研究中的數位方法演進

## From Quantification to Digital Humanities: The Evolution of Research Approaches to the Social-Economic History of the Qing Dynasty

趙思淵 (Zhao Siyuan)、潘芸淇 (Pan Yunqi) \*

### 引言

傅斯年在《歷史語言研究所工作之旨趣》中說：「近代的歷史學只是史料學，利用自然科學供給給我們的一切工具，整理一切可逢著的史料。」<sup>1</sup> 這段論述對史學與史料之關係的理解頗受爭議，在現代學術史中也聚訟紛紜。但是，這段話還有另一面向，並且也是現代史學發展不可忽略的一個層面：引入多種交叉學科的研究方法與工具，的確是現代中國史學長足發展的重要推手。

明清社會經濟史研究自其學科初創期就與社會科學方法的關係緊密，特別是注重運用數理統計方法與相關分析工具。所以在數位技術不斷更新迭代的情況下，這一領域如何進一步演變，就特別值得關注。數位人文是近年來方興未艾的方法論，數位人文對明清社會經濟史帶來的助益，主要在於史料整理與分析的工作方法將更為高效且豐富。學者可以較低成本處理大規模資料群，亦可能由分析工具的發展，提出新的研究議題與視角。

### 一、明清社會經濟史研究中的史料與分析工具

明清社會經濟史的發展存在兩個重要的學術傳統——賦役制度史與鄉村經濟研究，這兩者都顯著地受到相關社會科學方法論的影響。梁方仲的學術訓練來自經濟學，傅衣凌則受到社會學很深的影響。<sup>2</sup> 此外，廣義的中國社會經濟史研究，其早期發展也特別看重社會科學方法。以清華學人為核心的史學研究會奠定了中國社會經濟史研究的早期樣貌，其最初成員是湯象龍、吳晗、羅爾綱、朱慶永、谷霽光、梁方仲、劉雋、夏鼐等人。<sup>3</sup> 他們在當時的一系列研究中，都廣泛援引經濟學的概念、術語，以及量化統計研究方法。<sup>4</sup>

同樣是在交叉學科的研究議題以及方法的驅動下，社會經濟史研究自最初就特別看重收集、整理、研究多樣化及可進行量化分析的資料。陶孟和、湯象龍主持的《中國近代經濟史研究集刊》是當時史學研究會的發表

\* 作者趙思淵現為上海交通大學人文學院歷史系教授；潘芸淇現為上海交通大學人文學院歷史系碩士生。

1 傅斯年，〈歷史語言研究所工作之旨趣〉，《中央研究院歷史語言研究所集刊》1:1 (1928.10)：3。

2 陳春聲、劉志偉，〈天留迂腐遺方大，路失因循復倘艱——梁方仲先生的中國社會經濟史研究（代序）〉，《梁方仲文集》（北京：中華書局，2008）；陳支平，〈《傅衣凌著作集》與中國社會經濟史學派（代序）〉，《傅衣凌著作集》（北京：中華書局，2008）。

3 王世民、夏素琴整理，〈夏鼐日記〉第1冊（上海：華東師範大學出版社，2011），頁240。

4 陳峰，〈兩極之間的新史學：關於史學研究會的學術史考察〉，《近代史研究》1 (2006.1)：93-114。

陣地之一，其〈發刊詞〉顯示出與傳統史學研究所不同的取向。他們特別批評經濟類資料在傳統中國史料類型中的缺乏。「或者是以先的人對於經濟事實或經濟現象不加注意，沒有記載；或者有記載而人們不認識他的價值，未能保存。」所以建設中國經濟史研究，首先就是史料的建設，也即注意搜集家庭流水帳、店鋪生意帳、工料清單、戶口錢糧清冊等材料，「以先為人所拋棄的，至少不理會的，現在都變成最有趣的，最可寶貴的經濟史料了。」<sup>5</sup> 值得一提的是，梁方仲所開創的賦役制度史研究中，就已經有若干圍繞前人所不注意的文獻進行的專題研究，如戶帖、易知由單、魚鱗圖冊等等。這些文獻不少是他個人收集。從賦役冊籍的形成考察賦役制度的運行，也成為此後賦役制度史的重要路徑。<sup>6</sup>

傅衣凌所開創的契約文書研究則有著更為人熟知也頗具傳奇色彩的故事。1939年傅衣凌在福建永安黃曆鄉所發現的一箱契約文書，開啟了以契約文書研究鄉村經濟的學術脈絡。<sup>7</sup> 陳支平評論「這是我國學者第一次引用民間契約文書研究中國社會經濟史的著作，也是傅衣凌開創中國社會經濟史學派的奠基之作。」<sup>8</sup> 科大衛則更強調這一發現的實質是研究視角的轉變，是一個「牛

頓和蘋果的故事案例」。<sup>9</sup> 也就是說，1939年之前中國學者當然早已瞭解鄉村中廣泛存在著契約文書。在法學領域，也早有人將契約文書作為認定民事習慣的依據，進行系統整理與研究。<sup>10</sup> 傅衣凌的開創性在於，以契約文書的內涵激活了明清史的一種新的解釋路徑，也令契約文書成為煥發活力的「新史料」。這與陶孟和、湯象龍等人所論內在精神一致，也即過去學者不看重的材料，在新的經濟史研究中，都變得有趣、有價值。

多樣化、大規模、量化的史料在社會經濟史研究的發展中始終是重要的動力。近年來社會經濟史的長足發展，也伴隨著史料的迅猛擴增。傅衣凌的開創性工作無疑對此後的社會經濟史研究有著深遠的感召力：「一代史學大師，在鄉間農舍的黑暗中，探索一個已經逝去長久的時代，由此開創出中國社會經濟史的新領域。」<sup>11</sup>

事實上，最近二十年民間文書的收集、整理、刊布都迅猛增長。即使按照比較保守標準估計，目前披露數量超過萬件的民間文書群也已經有：臺灣古地契、徽州文書、清水江文書、浙南文書、太行山文書、閩東文書等，總量在百萬件以上。此外各地已經公開出版的民間文書的分布就更為廣泛。<sup>12</sup>

5 〈發刊詞〉，陶孟和、湯象龍主編，《中國近代經濟史研究集刊》1：1（1932.11）：2。

6 申斌，〈明代地方官府賦役核算體系的早期發展〉，《中國經濟史研究》1（2020.1）：41-54；申斌，〈明代中葉賦役經制冊籍纂修的擴展機制——財政管理技術傳播之一例〉，《史林》1（2021.2）：80-89。

7 傅衣凌，〈明清時代永安農村的社會經濟關係——以黃曆鄉所發現各項契約為根據的一個研究〉，《明清農村社會經濟》（北京：三聯書店，1939），頁20。

8 陳支平，〈《傅衣凌著作集》與中國社會經濟史學派（代序）〉，《傅衣凌著作集》，頁2。

9 科大衛，〈人類學與中國近代社會史：影響與前景〉，《東吳歷史學報》14（2005.12）：28-29。

10 楊培娜、申斌，〈走向民間歷史文獻學——20世紀民間文獻搜集整理方法的演進歷程〉，《中山大學學報（社科版）》54：5（2014.9）：71-80；賴駿楠，〈融民情於國法：民初法政人對土地權習慣的繼承與改造〉，《近代史研究》2（2022.3）：107-123。

11 曹樹基，〈石倉契約的發現、搜集與整理（代序）〉，《石倉契約》第1輯（杭州：浙江大學出版社，2011）。

12 臺灣古地契，參見「臺灣數位歷史圖書館」（THDL）。徽州文書，參見周紹泉，〈徽州文書與徽學〉，《歷史研究》1（2000.2）：51-60；劉伯山，〈徽州文書的遺存及特點〉，《歷史檔案》1（2004.2）：122-126；李琳琦主編，《安徽師範大學館藏千年徽州契約文書集萃》（蕪湖：安徽師範大學出版社，2014）。清水江文書資料，參見吳才茂，〈近五十年來清水江文書的發現與研究〉，《中國史研究動態》1（2014.2）：39-52；張應強，〈方法與路徑：清水江文書整理研究的實踐與反思〉，《貴州大學學報（社科版）》36:1（2018.2）：36-41；黃志繁，〈經濟史研究的新材料與新趨向——《客家珍稀文書叢刊》簡評〉，《光明日報》（2020.1.13）：14。浙南文書資料，參見《石倉契約》（杭州：浙江大學出版社，2011-2017）；李義敏，〈浙江師範大學藏民間文書的基本情況與史料價值〉，《西華師範大學學報（哲學社會科學版）》3（2018.5）：25-30；馮筱才主編，《浙江余族文書集成》（杭州：浙江大學出版社，2019）；「上海交通大學圖書館藏浙南文書」。太行山文書數量，參見馮小紅，〈古文書學視域下的明清民國時期太行山文書整理芻議〉，《邯鄲學院學報》26:3（2016.9）：37-42。閩東文書數量參見林麗金，〈閩東民間文書的收集整理和研究利用〉，《福建文博》4（2021.12）：49-52。

就明清社會經濟史而言，一些新的地權結構研究已經改變了學界對於明清時期鄉村經濟，尤其是金融與市場流通的理解。<sup>13</sup> 這些研究開啟了對於近代早期中國「市場經濟」與私法領域的重新審視。<sup>14</sup> 賦役制度與市場的互動關係以及王朝體制如何與一定的市場秩序、社會結構結合起來，也在近年重新成為研究熱點。<sup>15</sup> 這些研究主要仍是通過逐件閱讀契約文書，或者由研究者個人的眼光與功力，勾連不同文獻之間的有機聯繫而完成。就工作方法而言，這與更早期的社會經濟史研究並無實質區別。

面對既存的上百萬件的民間文書，整理、影印、錄文、出版可能仍將是主流的利用方式。另一方面，我們總是期待：這樣空前大規模的「新史料」，能為史學的研究議題與方法帶來什麼樣的推動呢？至少，我們直接面臨這樣一個問題：如此大規模的資料如何去「讀」？除了歷史學的文本分析，我們是否還有別的方法發現史料內在的研究線索？由這些問題促成的一個顯著趨勢是，資料庫成為了學界共識的解決之道。

「臺灣數位歷史圖書館」是較早開發也較為成熟的契約文書資料庫。這一資料庫不僅提供全文檢索，更著

力於為使用者提供分析工具。例如實現上下手契關聯，契約文書的空間分布，以及契約文書內人物相關性的分析，這都為此後史料資料庫開發提供了範本。2010年前後，黃山學院也以館藏徽州文書為基礎探索了契約文書資料庫建設方案。<sup>16</sup> 「中國地方歷史文獻數據庫」為契約文書提供了基於文書結構的資源分類法、後設數據方案，及文書間關係的分析工具。<sup>17</sup> 此外，中山大學、廈門大學都已經提出了民間文書資料庫建設的實施方案，一些文書收集數量較大的館藏單位，也都已經有資料庫建設。<sup>18</sup>

## 二、數位工具能夠「讀」史料嗎？

那麼，資料庫建設能夠說明我們用新的方式「讀」史料嗎？資料庫僅僅是將我們慣常閱讀的史料放入一個「數位倉庫」，還是能夠帶來史學研究的視角或方法論的轉變？對此，歷史學界一直存在著質疑、批評的聲音。更有學者擔心，資料庫造成了史料運用的誤區，從而損害了「歷史學家的技藝」。

資料庫或者任何其他數位人文工具都不能替代史學

- 
- 13 曹樹基，《傳統中國地權結構及其演變》（上海：上海交通大學出版社，2014）；龍登高，《中國傳統地權制度及其變遷》（北京：中國社會科學出版社，2018）；杜正貞，《近代山區社會的習慣、契約和權利》（北京：中華書局，2018）；劉詩古，《資源、產權與秩序：明清鄱陽湖區的漁課制度與水域社會》（北京：社會科學文獻出版社，2018）；仲偉民、王正華，〈契約文書對中國歷史研究的重要意義——從契約文書看中國文化的統一性與多樣性〉，《史學月刊》5（2018.5）：5-15。
- 14 賴駿楠，〈重新認識中國傳統地權——從龍登高、曹樹基與劉詩古有關傳統地權的三部著作談起〉，《廈門大學法律評論》1（2022.4）：1-8。
- 15 劉志偉，《貢賦體制與市場：明清社會經濟史論稿》（北京：中華書局，2019）；丁亮，《明代浙直地方財政結構變遷研究》（北京：中國社會科學出版社，2020）；邱永志，〈「白銀時代」的落地：明代貨幣白銀化與銀錢並行格局的形成〉（北京：社會科學文獻出版社，2018）；趙思淵、劉志偉，〈在戶籍賦稅制度與地權市場運作中認識明清土地制度〉，《中國社會科學》1（2020.2）：195-203；申斌，〈賦役全書與明清法定財政集中管理體制的形成——兼論明清國家財政治理焦點之轉移〉，《中國經濟史研究》1（2021.1）：54-72；黃忠鑫，〈明代前期里甲賦役制度下的徽州社會——祁門縣文書《百戶三代總圖》考析〉，《中山大學學報（社科版）》58:1（2018.1）：92-104；郭永欽，〈明清賦稅核算技術變革與賦稅折畝數字的製造〉，《清華大學學報（哲社科版）》34:4（2019.7）：21-36。
- 16 俞乃華、何廣龍、汪柏樹，〈徽州文書整理及特色數據庫建設研究〉，《大學圖書館情報學刊》28:1（2010.1）：54-56。
- 17 趙思淵、湯萌，〈上海交通大學新藏地方歷史文獻的分類法及其依據〉，《上海交通大學學報（哲社科版）》22:3（2014.5）：76-87；張潔、李芳、湯萌，〈契約文書描述元數據規範設計與應用〉，《圖書情報工作》61:8（2017.6）：106-111頁；趙思淵，〈地方歷史文獻的數字化、數據化與文本挖掘：以《中國地方歷史文獻數據庫》為例〉，《清史研究》4（2016.11）：26-35。
- 18 王蕾、薛玉、肖鵬、申斌，〈民間歷史文獻數字人文圖書館構建——以徽州文書數字人文圖書館實踐反思為例〉，《圖書館論壇》38:3（2017.10）：30-36；黃向春，〈民間文獻、數據庫與作為方法的總體史〉，《光明日報》（2020.2.17）：14；楊存林，〈清水江文書數據庫功能的設計與實現——以凱里學院清水江文書數據庫為例〉，《凱里學院學報》36:4（2018.7）：37-40。

研究本身，就這一意義上來說，這些批評是有道理的。例如陳爽批評資料庫「窮盡史料」並不能達到「史料大發現」，「數位時代史學研究所面臨的主要困惑之一，是史料的處理和運用。」<sup>19</sup> 推動史學研究發展的根本還是史家的內在功力。此後包偉民也指出，研究者依據自己的主觀目標由機器檢索獲得史料，可能存在「懸問題以覓材料」的問題，理想的情況是「讀書得間」，也就是「通過分析存世歷史文本，在字裡行間的細節中發現前人所忽略的問題。」<sup>20</sup>

人文學科的優勢在於面對並解釋有關人類文明的不確定性以及模糊而複雜的問題。通過細膩的文本分析解釋人性、社會、世界，這是人文學特別是歷史學的根本職責，是任何數位工具乃至人工智慧都不可替代的。否則，歷史學家這個職業本身就會消失。在世界範圍內，也廣泛存在著人文學者對於運用數位工具作大規模文本處理與資料分析的批評。<sup>21</sup>

避免資料庫或任何其他數位工具變成「懸問題以覓材料」的誤用，其解決途徑也許已包含於包偉民教授所指出的：「如何在符合學科特點的基礎之上，更有效地利用資訊技術，以推動歷史學研究的深化發展，對史學從業人員來說，的確是一個新課題、新挑戰。在這裡歷史學與計算機學科等領域的攜手合作非常重要。……同時，更試圖為計算機等學科提供參考意見，期待有關學者改進技術，幫助史學研究更好地融入數字化時代。」<sup>22</sup>

歷史學者參與到資料庫建設與數位人文工具開發中，本就是題中應有之義。以現代中國歷史學的發展來看，編纂文獻目錄、史料索引、史料工具書本就是重要

的學術傳統。如敦煌文書目錄、二十四史地名索引、內閣大庫目錄，今天已視之為典範工作，本身就是重要的研究成果。將這些工作轉換到當下的研究場景中，其實正是歷史學者如何在數位化環境中製造分析工具的問題。申斌和楊培娜的研究就發現，經典的紙本史料整理方式、工具書，今天都已經有其對應的資料庫類型。<sup>23</sup>

為文獻編制索引、目錄、注解，也可以理解為一種資訊存儲和檢索方法，當下的資料庫建設事實上是這一套工作方法的自然延伸。<sup>24</sup> 史睿更將索引(index)視作關鍵的方法論概念，以其聯結起傳統的文獻學、史料學工作與當下的資料庫建設。在人文學傳統中，為史料編製索引的方法，正對應了數位環境下「神經網絡式的知識管理體系」。<sup>25</sup> 20世紀30年代洪業所主持的〈引得〉編纂工作，可以視作當下史料資料庫建設的開端與必經之路。<sup>26</sup> 數位人文研究者內部也已有這樣的共識：數位人文是傳統知識技能和方法的延伸，而不是替代者。<sup>27</sup> 數位人文的意義是幫助人文學科進一步強化其研究理念與優勢，例如對文本作精細的批判性閱讀以及有關經典問題的複雜論辯。

就明清社會經濟史而言，專題研究與史料數位化整理相結合，已經有相當長的發展歷程。這其中最早的典範工作應當是王業鍵主持的「清代糧價資料庫」。清代自乾隆元年開始，各省按月向皇帝奏報糧價，也即「糧價清單」。20世紀90年代之前，已有不少研究專門利用糧價資料，討論區域經濟或市場流通。<sup>28</sup>

1994年王業鍵開始建設清代糧價資料庫，到2008年初步完成。王業鍵及其團隊在這一過程中形成了一系列

19 陳爽，〈回歸傳統：淺談數字化時代的史料處理與運用〉，《史學月刊》1（2015.1）：15。

20 包偉民，〈數字人文及其對歷史學的新挑戰〉，《史學月刊》9（2018.9）：9。

21 安妮·伯迪克等，《數字人文：改變知識創新與分享的遊戲規則》（北京：中國人民大學出版社，2018），頁106。

22 包偉民，〈數字人文及其對歷史學的新挑戰〉：11-12。

23 申斌、楊培娜，〈數字技術與史學觀念——中國歷史數據庫與史學理念方法關係探析〉，《史學理論研究》2（2017.4）：87-95。

24 安妮·伯迪克等，《數字人文：改變知識創新與分享的遊戲規則》，頁21。

25 史睿，〈數位人文憂思錄〉，《數字人文》2（2020.6）：157-160。

26 〈史睿談數字人文與現代文獻學研究〉，澎湃新聞·上海書評（2020.8.9）。

27 安妮·伯迪克等，《數字人文：改變知識創新與分享的遊戲規則》，頁16。

28 最詳盡的綜述可參見彭凱翔，《清代以來的糧價：歷史學的解釋與再解釋》（上海：上海人民出版社，2006），頁17-23；羅暢，〈兩套清代糧價數據資料的比較與使用〉，《近代史研究》5（2012.9）：142-156。

糧價與市場的專題論文。<sup>29</sup>今天的經濟史研究中，這一資料庫業已成為數位基礎設施。基於這一資料庫，學者們對清代市場發育與市場整合的認識獲得長足發展。<sup>30</sup>值得一提的是，「清代糧價資料庫」極大促動了發展經濟學對中國經濟史的研究興趣，儘管由此帶動的經濟學與歷史學的對話至今仍存在很多爭議與分歧。

「清代糧價資料庫」主要仍是一種「檢索——獲得」型資料庫。研究者根據自己的研究經驗與研究興趣，提出需求，形成資料庫的基本結構。參與開發的研究者對於史料的理解，的確會影響資料庫所呈現的形態。王業鍵本人對清代經濟史的精深見解，保證了資料庫的品質。我們在開發「中國地方歷史文獻數據庫」時也感受到，資料庫建設的第一步並非數位技術相關的討論，而是對史料的文獻學研究以及對相關研究的總體把握。<sup>31</sup>這個過程中，研究者應當提出明確的研究需求，這些研究需求還應當轉換為資料庫開發中可操作的「工程圖」。簡言之，歷史學家的術語要能轉換為工程師的術語，歷史學者需要學會與工程師打交道。

由學者深度參與開發的資料庫能夠解決相應的研究需求。但是這種資料庫往往是封閉的，僅能針對特定的研究資料。這時，數位人文真正重要的意義才凸顯出來：資料庫不僅僅是為研究者提供隨手可得資料的「資料倉庫」，而是搭建一種不同於以往的研究環境。

數位人文方法幫助歷史研究建立新的研究環境，這種取徑與歷史學本身追求對史料的考辨與細讀，事實上

是同構的。數位人文研究的先行者項潔就曾經說：「一個以研究為導向的系統，它必須提供觀察史料間多種脈絡的可能。」這意味著，資料庫建設與數位工具開發，都是幫助研究者以檢索、統計、結構化、關係網絡等方式把握史料。這些方法既可能是回應研究者所預設的問題，也可能是啟發研究者發現新問題，但這與人文學的核心研究方法之間並不是更新換代的關係。「檔案數位化不是消極的收藏，而是有更積極的目的，要打造一個適合研究者使用的觀察環境，並進一步找出更精密的史料分析方式，乃至開拓一種新的研究方法，建立起檔案間的新脈絡。而這也才是數位人文研究的意旨所在。」<sup>32</sup>

項潔開發的 THDL 以及 DocuSky 都可以算是應用這種理念的數位人文工具。這一類數位人文工具最重要的功能是提供使用者「觀察」史料的工具。「觀察」史料，或者是解決研究者本就懸而未決的問題，或者是發現研究者意料之外的問題。<sup>33</sup>

魏希德（Hilde De Weerd）團隊所開發的 MARKUS 也在這一方法論的脈絡之中。MARKUS 最初開發的目的就是為人文學者提供一種文本標記工具。就研究方法而言，這其實就是在數位環境下進行人文學者所熟悉的文獻批註、校注工作。<sup>34</sup>數位工具給這類傳統研究方式所帶來的變化是，以標籤（tag）形式記錄下來的批註、注解可以用於後續的量化分析。當研究者所作的標籤積累到一定規模後，就會發現這些標籤所呈現的結構特徵，並說明研究者建立對文本的解釋與分析，從而產生

29 王業鍵，《清代經濟史論文集（一）》（臺北：稻鄉出版社，2003）。

30 顏色、劉叢，〈18世紀中國南北方市場整合程度的比較——利用清代糧價數據的研究〉，《經濟研究》46:12（2011.12）：124-137；Shuo Chen, James Kai-sing Kung, "Of maize and men: the effect of a New World crop on population and economic growth in China," *Journal of Economic Growth* 21:1(2016.3):71-99.

31 趙思淵，〈地方歷史文獻的數字化、數據化與文本挖掘：以中國地方歷史文獻數據庫為例〉，《清史研究》4（2016.11）：26-35；趙思淵、湯萌，〈上海交通大學館藏地方歷史文獻的整理與數字化〉，吳佩林、申斌主編，《地方檔案與文獻研究》第三輯（北京：國家圖書館出版社，2017）。

32 項潔、翁稷安，〈數位人文和歷史研究〉，項潔編，《數位人文在歷史學研究的應用》（臺北：臺灣大學出版中心，2011），頁18-19。

33 塗豐恩等，〈當資訊科技遇到史料——臺灣歷史數位圖書館中的未解問題〉，項潔主編，《數位人文研究的新視野：基礎與想像》（臺北：臺灣大學出版中心，2011），頁23-24。

34 魏希德等，〈魏希德：數字人文研究的問題意識〉，《數字人文》3（2020.9）：1-12。

新的研究。<sup>35</sup> 另一方面，機器可以學習這些人工標籤，此後分析工具就可以對其他文本作自動標記，從而說明研究者更高效地做史料分析。同樣的方法，也可用於文本比勘，以提高效率。

### 三、從「檢索——獲取」到「建置觀察史料的數位環境」

以下將以運用 DocuSky 整理契約文書的實例說明如何以「建置史料觀察環境」的思路運用數位工具於社會經濟史研究。如前所述，民間文書是近年來迅速擴增的史料群。學界對於民間文書的研究興趣，則很大程度上來自於明清社會經濟史有關地權結構的討論。基於大量新發掘材料的研究顯示出，明清時期的鄉村土地市場發達而活躍，形成了形式多樣的地權交易結構，由此也形成了一定程度的信用金融體系。<sup>36</sup> 由此值得進一步探討的是，土地市場及其信用體系，需要依託一整套契約格式規範以保障其運轉。明清時期契約格式規範的一致與穩定，本身就顯示了土地市場背後的社會秩序與信用機制的內在特徵。<sup>37</sup> 契約如何書寫，使用怎樣的用語，顯示出稅契制度、民間地權觀念之間的互動關係。<sup>38</sup>

資料庫首先可以便利學者對契約書寫形式作量化及歷時性的考察。我曾經利用「中國地方歷史文獻數據庫」考察清代以降徽州歙縣的田面權買賣契約形式的演

變。<sup>39</sup> 這個研究在資料庫中的工作方式是，首先限定於清代的歙縣，之後檢索「退批」、「退業」、「退小買」、「杜退」、「退大小買」、「頂首」、「頂頭」、「頂約」、「交業」、「賣小買」等關鍵字，得到一組資料統計。控制時間、地域以及檢索到這些關鍵字，都依靠資料庫開發時製作的中繼資料資訊。這需要資料庫建設時，研究者參與到中繼資料設計與製作，研究者基於既有的民間歷史文獻的文獻性質研究而設定，包含了歸戶、文獻分類、地域、時間等必要元素，從而實現交叉檢索。

DocuSky 則可以提供一種新的工作方式。<sup>40</sup> 以「歙縣二十六都七圖陸氏文書」為例，這是來自上海交通大學館藏的一包歸戶性的契約文書，共有 87 件。將這批資料全文錄入後，不需要進行中繼資料編目，直接導入 DocuSky。DocuSky 提供的擷詞工具可自動標記文本中需要研究的內容。首先，資料庫提供的「種子詞彙巨集」包含了一組文史研究中的常用標記，如「中曆」、「年月日」、「干支」、「數字」、「數量或金額」、「蘇州碼」、「人名 CBDB」等。這些詞彙巨集可以說明研究者先將研究資料中的相關資訊提取聚集起來，做整體的觀察。我們對文本處理的第一步是用「種子詞彙巨集」提供的年份標記所有文本，從而初步觀察契約文書的時間分布，由此可以注意到 19 世紀 50 年代和 90 年代是契約交易比較活躍的時期（圖 1）。

35 魏希德、陳大龍〈中文與韓文數據集的創建、關聯與分析——用 MARKUS 與 COMPARATIVUS 進行數字文本標注〉，《數字人文》3 (2021.9)：28-36。

36 曹樹基，〈傳統中國地權結構及其演變〉（上海：上海交通大學出版社，2014）；龍登高，〈中國傳統地權制度及其變遷〉（北京：中國社會科學出版社，2018）；曹樹基、楊啟明，〈清代東南地方小市場中的憑票與信用〉，《中國經濟史研究》152 (2020.11)：102-117；婁敏，〈近代中國長江上游地區的商業信用票據：以巴縣與江津為例〉，《中國經濟史研究》6 (2020.11)：118-132；馬秀英、曹樹基，〈清代後期吐魯番的葡萄園典當與金融〉，《清史研究》6 (2021.11)：58-73。

37 岸本美緒，〈明清契約文書〉，王亞新、梁治平、趙晶編，〈明清時期的民事審判與民間契約〉（北京：法律出版社，2022），頁 302-339；趙思淵，〈歙縣田面權買賣契約形式的演變（1650—1949）〉，《清華大學學報（哲學社會科學版）》32:6 (2017.11)：67-76。

38 阿風，〈明末清初田宅交易稅契制度的演變——以徽州文書為中心〉，《江海學刊》4 (2021.7)：237-246；王正華，〈合與分：清代鄉村土地交易中的典與活賣〉，《中國經濟史研究》5 (2022.9)：96-111。

39 趙思淵，〈歙縣田面權買賣契約形式的演變（1650-1949）〉：67-76。

40 Hsieh-Chang Tu, Jieh Hsiang, I-Mei Hung, Chijui Hu, "DocuSky, A Personal Digital Humanities Platform for Scholars," *Journal of Chinese History* 4:2(2020.7):564-580.

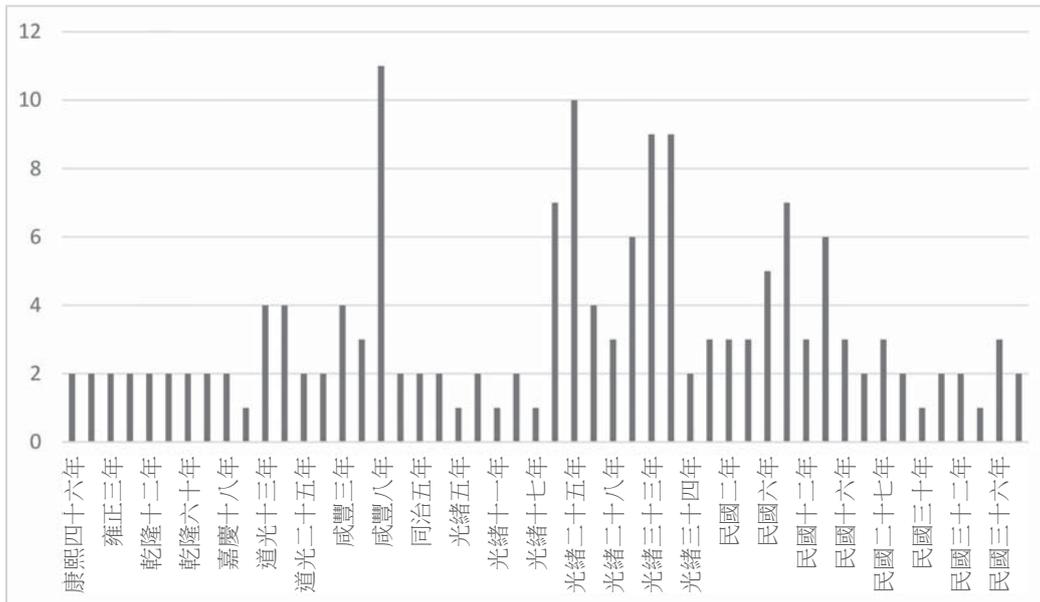


圖 1 「歙縣二十六都七圖陸氏文書」時間分布

DocuSky 的擷詞工具還提供了與 CHGIS 及 CBDB 關聯的種子詞彙巨集，可以對文本中的人名、地名進行自動識別標記。同樣是對以上文本，利用 DocuSky 關聯 CHGIS 地名標記可以得到以下結果（圖 2）。這樣的結果當然不是完善的史料分析，但的確可以為我們分析這組分數提供初步線索。在此基礎上進行人工判定，排除

江南、安徽、徽州府等高等級行政區劃名稱，可以將當地的土地交易聚焦於二十六都的青山、小田、大洋等地名。這些資訊還需要進行人工校訂。如自動抓取的地名「青山」，核對契約文書，實際上應為「青山下」。經過校訂的地名可以作為標籤，在文本中進行標記，從而形成有關當地的地名集合。

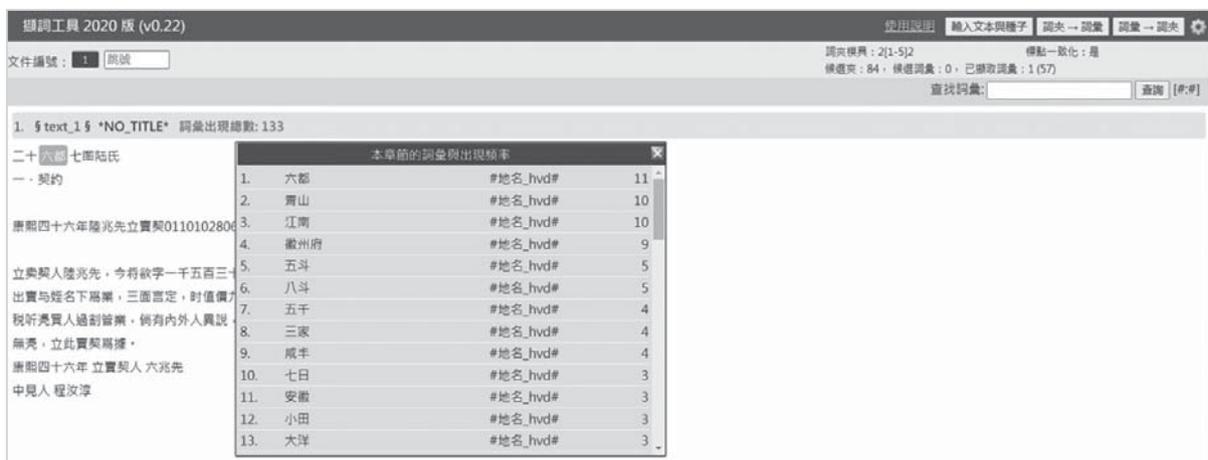


圖 2 「歙縣二十六都七圖陸氏文書」地名詞彙標記統計（種子詞彙集自動標記）

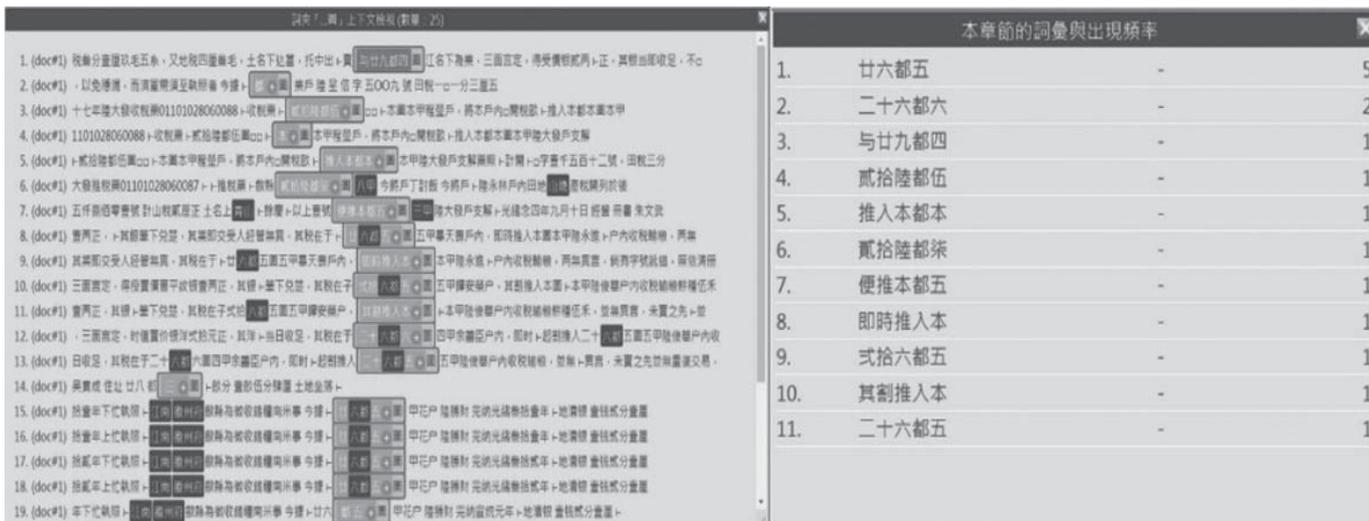


圖 3 「歙縣二十六都七圖陸氏文書」地名詞彙標記統計（人工修訂）

這批文書在收集時是有歸戶線索的，所以我們已經知道這批文書主要應來自歙縣二十六都七圖陸氏。因此，以上檢索中的六都，其實是擷詞工具自動提取時將「二十六都」或「廿六都」誤切分為「六都」造成的。這是否說明我們不能依靠這樣的工具做進一步的研究呢？並非如此。擷詞工具提供的前置詞與後置詞的提取工具（夾詞）就可以修正這個問題。我們將「立賣……」與「推入……」、「……圖」分別設為前置詞、後置詞，這樣就可以將契約文書中標識地域資訊的文字夾取出來，從而得到一組新的都圖資訊（圖 3）。這個檢索即可幫我們觀察到多數交易都在二十六都六圖、五圖的範圍內達成，最遠則有二十九都四圖業戶參與的交易。而以都、圖標記業戶資訊的契約亦僅占全部契約的少數。

針對我們所感興趣的土地交易形式的問題，將「立……契」、「立……字」、「立……批」作為夾詞，又可以得到以下結果（圖 4）。這裡的資料當然還需要進一步清洗、精煉，才能得到有效的史料分析。但這一資料處理過程，已經可以幫助我們在研究初期快速觀察

資料中所可能存在的研究線索。例如，這裡所觀察到的「當」與「退」的交易形式頻率相當，這與此前研究所觀察到的 18 世紀之後「退」型契約占主導的情況有所不同，可能成為新的研究突破點。



圖 4 「歙縣二十六都七圖陸氏文書」交易形式詞彙統計

Docusky 平臺中的「文獻集字頻統計」工具還能夠幫我們得到一些新的觀察。這一家戶留存的所有契約文書所包含的不重複單字是 3,334 個，這個數字其實遠超過我們通常所認為的鄉村生活中的常用字的數量。<sup>41</sup>當然，這個數字可能還要剔除賦役冊籍中的刊刻文字以及同一字的大量異體寫法，但這的確可以對爭論已久的鄉村社會識字率問題提供新的解答路徑。<sup>42</sup>

從明清社會經濟史研究的經驗來看，新的數位人文方法與數位工具所帶來的變化，最重要的可能就是從「檢索—獲得」的工作方法轉變為「建置觀察史料的數位環境」。近年來，明清社會經濟史研究越來越注重發現史料之間的「關係之網」與「意義之網」。<sup>43</sup>建置觀察史料的數位環境，正能夠與這樣的研究需求相適應。

在數位環境下處理與解讀史料意味著，研究者將更容易發現一個史料群的結構性特徵。例如上文所處理的契約文書群，其交易形式、空間坐落、人物關係等方面的特徵，很容易通過數位分析工具呈現出來。這種分析方法與傳統的諸字逐件精細閱讀的工作方法是並行不悖的。通過精細閱讀，研究者才能夠理解史料文本的內涵與語境。這種理解常常幫助研究者解決由數位分析工具「鳥瞰」史料時所發現的問題。

#### 四、新挑戰：數位時代「歷史學家的技藝」

數位人文方法對於歷史學更深遠的挑戰恐怕在於，未來的歷史學者將面臨比我們更複雜的史料環境。我們只要想想看，自媒體、短視頻、虛擬實境等等要如何作為史料來運用？這將是未來歷史學者真正要面臨的問題。這就意味著，我們必須要為這些真正意義上的「數位原生」史料準備新的「歷史學家的技藝」，這可能才是數位人文帶給歷史學真正深遠且亟待探索的問題。

至少對於 20 世紀 90 年代之後的全球社會來說，網路原生資料已經成為理解社會變遷的必不可少的部分。如美國歷史學家所已經討論的，對於任何一個將會研究川普時期美國政治史的歷史學家來說，推特、臉書都會是核心史料，而非僅僅是檔案、報紙等傳統政治史材料的補充。<sup>44</sup>也就是說，未來的歷史學家也必須認真思考如何處理這些史料。逐字逐句的文本細讀仍將是歷史學的核心工作方式。但是，在面對這些「新材料」時，歷史學家將首先需要習得合適的數位工具以處理史料。如何使用數位工具，也將相當程度形塑所獲取史料的形態。從社群媒體中抓取資料，對影音視頻作標記與分析，以程式設計處理檔案，都將可能是未來歷史學家的必備技能。在那種情景中，「數位歷史」或「數位人文」都將是不必再討論的議題。

41 李伯重，〈什麼是『中國』？——經濟史中的『微觀』研究〉，《清史研究》6（2020.11）：131-134。

42 有關明清民眾識字問題爭論的最新研究參見：溫海波，〈雜字讀物與明清識字問題研究〉，《安徽史學》4（2021.7）：26-33。

43 杜正貞，〈訴訟實踐中的意義之網與關係之網：社會關係網絡視角下的清末民初龍泉司法案例〉，《民俗研究》2（2020.3）：56-69。

44 Daniel J. Story, etc., "History's Future in the Age of the Internet," *American Historical Review* 125: 4 (2020.10): 1337-1346.