

# 關鍵詞叢與文本意義挖掘的嚐試： 以《清季外交史料》為例\*

邱偉雲\*\*

王爾敏先生曾分析研究先秦文獻中「中國」一詞的使用類型，勾劃出早期「中國」觀念的形成。筆者分析統計了「二十四史」中二千多次「中國」的意義，對王爾敏先生的結論進行檢驗。

筆者於上述研究基礎上，進而欲考察歷史上形成之傳統「中國」觀念，在近現代之演變。通過含有一億兩千萬字的《中國近現代思想史專業數據庫 1830-1930》的檢索發現，共有 19 萬條相關論述。19 萬筆相關資料，已不是個別研究者用傳統方法可以處理的了，此即呈現出需借助數位人文研究法之迫切性。

本文研究目的，在於構設出一套可搭配程式，進行龐大史料處理之數位人文研究法。透過此方法的操作，協助研究者快速建立宏觀歷史發展脈絡，即為本文關懷所在。本文的研究方法如下：1，先以數據庫中的《清季外交史料》(約 400 萬字)為研究範圍，分析「中國」論述中提及的關鍵詞類型。2，仿科學實驗的抽樣小試，設計程序對所有提及「中國」之內容，進行觀念類型提取並統計結果。3，形成一個特定樣本後檢驗並調整在小範圍史料分析的有效性。4，再以所歸納出的近現代「中國」認同下「主權」觀念類型發展趨勢圖，去與當前相關研究成果對話，為研究者提供一個宏觀的參考。將來冀能在此基礎上，處理《中國近現代思想史專業數據庫 1830-1930》中 19 萬筆相關資料，完整呈現出多元面向之「中國」認同展演，此即為本文之推廣前景所在。

**關鍵字：認同形成、認同重構、清季外交史料、數據庫方法、觀念史**

---

\* 本研究中之《清季外交史料》「中國」關鍵詞用例部分，取自於「中國近現代思想史專業數據庫」(香港中文大學中國文化研究所當代中國文化研究中心研究開發，劉青峰主編)，謹致謝意。本文研究構想之萌發，乃立基於參與由政治大學鄭文惠教授主持，由政治大學講座教授金觀濤先生，與香港中文大學中國文化研究所名譽研究員劉青峰教授，在政治大學所帶領之「中國現代認同的形成和演變研究群」。筆者參與研究群一年以來，獲益良多。在每隔周二舉辦的研究群中，諸位成員皆對自己所負責之關鍵詞研究進行報告，再由老師們進行評點。在 20 幾次的聚會中，透過成員們各自負責的主題報告，如「天下」、「萬國」、「個人」、「國家」、「民族」、「主義」、「中華」、「中國」等，在上述近現代政治範疇下，最重要的幾個觀念彼此討論與互動中，成員們可不斷的檢視著自己的研究，也共享著彼此的成果，從中迸生出諸多新的研究方向。在不斷地以數據庫進行觀念史研究下，也讓每個成員對觀念史研究應如何與數位密切結合產生莫大興趣，而本文正是立基於此，提出觀念史研究與數位研究結合之新取徑。這個新取徑也是本研究群在一年多的人文與數位結合研究試錯過程中的體驗結果，亟欲與人文研究者分享與討論之新方法。本文更期許將來有更多的研究者，可以投入研究群中，一起來思考觀念史研究與數位研究結合的各種可能，為數位人文研究開啓更多的嘗試，此為本文之殷切期盼所在。

\*\* 國立政治大學中文研究所博士生。

# **A Discussion of the Keywords Cluster and Meaning of the Text Mining: Focusing on the “Qing Diplomatic Historical Data”**

Chiu, Wei-Yun\*

Prof. Wang Ermin analyzed the term of ‘China’ in pre-Qin literature, and outlined how the idea ‘China’ had been forged. By observing and analyzing the term of ‘China’ in the "Twenty-Four Histories", more than 2000 times of the usage ‘China’ can be used as a quantitative data in this research to testify Mr. Wang Ermin’s conclusion.

In order to research the formation and evolution of the idea of ‘China’, the first step is to investigate modern evolution of the traditional idea ‘China’. With the great amount of primary sources in the 'the Database on History of Chinese Modern Thoughts (1830-1930)', we found 190,000 related discussions. Therefore it is evidently imperative that we need to develop a new methodology of digital research on humanities to deal with enormous data.

The method is presented as following : 1.) The primary source of the research will be confined to the "Qing diplomatic historical data" (About 400 million words), in order to analyze the types of discourses ‘China’. 2.) The sampling process of scientific experiments is designed to process and extract the ideal type of "China", and the results will be analyzed. 3.) A certain mode is established to test the sample and to exam its validity with a small amount of historical data within a restrained range, 4.) Finally, we can gather the picture of the development and tendency how the idea “sovereignty” had been evolved in the light of modern Chinese identity. In this sense, more dialogue among current relevant research would be encouraged, and a macro-historical reference is thus established. In the future, hopefully, this method can be applied to deal with the whole 190,000 data, and to demonstrate a multi-oriented performances of “Chinese” identity. This is the prospects of this study.

**Keywords: Identity Formation, Identity Reconstruction, “Qing Diplomatic Historical Data”, Database Approach, History of Ideas**

---

\* Ph.D Student, Department of Chinese, National Cheng-chi University