

「共現」詞頻分析及其運用

——以「華人」觀念起源為例¹

金觀濤*、邱偉雲**、劉昭麟***

摘要

本文研究目的旨在透過《清季外交史料》(1875-1911年)當中對於「華工事件」的歷時性描述，形構出當時官方對於「華工事件」的認知轉移過程，再從事件中諸多「關鍵詞」內涵的歷時性變化中，觀察「華人觀念」是如何隨著「華工事件」認知焦點的轉移而逐漸形成之脈絡。本文之研究範圍以四百萬字之《清季外交史料》(1875-1911年)為底本，挑選曾出現過「華工」此一關鍵詞之文獻共有109篇，總字數為118,899字，旁及目前「華工事件」與「華人意識」之相關研究。本文主要以劉昭麟博士所研發之「共現詞頻分析法」為數位輔助程式，透過此程式，可協助研究者快速在所欲研究的龐大文本中，切割出眾多的待選詞，再由研究者確定重要關鍵詞後，進一步將關鍵詞兩兩配對成共現詞組，觀察各共現詞組的共現頻度及歷時性分布。而研究者即可從高共現頻度之共現詞組中，架構出以事件為核心之重要關鍵詞叢，進而觀察事件與觀念之間的互動過程。期待透過本文的小試，能將數位人文學推廣給更多人文學界研究者使用，藉由數位方法能讓人文研究者處理龐大的文獻底本，進行更為宏觀的研究，此為本文結合數位方法與人文研究之推廣前景所在。

關鍵字：華工、共現詞頻分析法、事件研究、觀念史、數位人文學

¹本文所使用之共現詞頻分析法，為劉昭麟博士所研發，其中有關程式撰寫和理論基礎，見於劉昭麟、金觀濤、劉青峰、邱偉雲、姚育松。2011。〈自然語言處理技術於中文史學文獻分析之初步應用〉。「2011第三屆數位典藏與數位人文國際研討會」發表之論文，國立臺灣大學。

* 國立政治大學講座教授，E-mail: gtqf1908@gmail.com。

** 國立政治大學中國文學系博士班研究生，E-mail: acwu0523@gmail.com。

*** 國立政治大學資訊科學系教授，E-mail: chaolin@nccu.edu.tw。

Frequency Analysis and Application of “Co-occurrence” Phrases: The Origin of the “Hua-ren” Concept as an Example

Guan-tao Jin^{*}、Wei-yun Chiu^{**}、Chao-lin Liu^{***}

Abstract

This article aims to demonstrate the official conceptual change in the “Hua-gong Event” (華工事件 the Chinese worker event) using the chronological records of the “*Qingji waijiao shiliao*” (清季外交史料) (1875–1911) and to illustrate the context of the concept “Hua-ren” (華人, the Chinese) developed with the attention shift during the events and the chronological change in the content of the keywords. The primary source is the all-text data from the *Qingji waijiao shiliao*, which contains over 4,000,000 characters of official records from 1875 to 1911. There are 109 articles (the overall word count is 118,899), which include the phrase “Hua-gong” (華工), regarding relevant research on the Hua-gong Event and Hua-ren consciousness. The current research utilizes the “analysis of co-occurrence frequency” (共現詞頻分析法) developed by Dr. Liu as the chief digital approach. With this program, various candidate phrases are selected among an enormous quantity of text. After a collection of the keywords from the candidates has been chosen, the program compares each keyword and chronologically shows their correlation with frequency of co-occurrence (共現頻度). Those with a high frequency of coexistence establish a keyword cluster that demonstrates the attention of the events. Thus, the interaction of concepts and events could be observed. The research could encourage more humanists to apply the approach to digital humanities. Through this digital approach, humanities research can be conducted on immense amounts of data and accordingly depict a panorama of the past, which is the aim of the current research.

Keywords: Hua-Gong, analysis of co-occurrence frequency, event studies, history of concepts, digital humanities

^{*} Chair Professor, National Chengchi University. E-mail: guantao@nccu.edu.tw

^{**} Ph.D. student, Department of Chinese Literature, National Chengchi University. E-mail: acwu0523@gmail.com

^{***} Professor, Department of Computer Science, National Chengchi University. E-mail: chaolin@nccu.edu.tw