大規模佛教文獻群に對する確率統計的分析の試み

師茂樹
（花園大學、京都）

文獻研究には、何らかの假説が必要である。例えば、ある研究者が無數にある文獻の中からどれかを選び出し讀み始めるとき、その選擇の背景には文獻に對する知識や先行研究、研究者の關心などに基づいた何らかの假説が形成されているはずである。電子テキストに對する檢索もまた、テキストの内容や表現をあらかじめ知っていなければ、有意な結果が出にくい場合が多い。

近年、佛教文獻をはじめとする漢字文獻の電子化が急速に進み、研究環境は大幅に進歩したように思われる。しかし、研究者の知識の範圍を大きく超える大規模な文獻群を扱うことができる現在の環境においては、假説を前提とする從來の方法が通用しなくなってきている。大規模な電子文獻群を扱うための方法論の檢討が、大きな課題となっている。

情報科學の分野では、大規模なデータベースの中から、確率・統計やパターンマッチなどの技術によってパターンを發見し、發想や假説生成を支援する技術であるデータマイニングが注目されている。古典テキストの計量的な分析はこれまでも行われてきたが、文獻學者の假説形成のために用いられることはほとんどされてこなかった。

本發表では、大規模な佛教文獻データベースに對して、確率・統計的手法による分類を通じて、文獻研究につながるような假説の形成を試みることで、大規模電子文獻を扱うための方法論について考えてみたい。その際、具體的には玄奘譯とされるすべての漢譯佛典を對象とし、コンピュータによる分類と從來の文獻學的研究の比較を行いたい。

師茂樹　MORO Shigeki
1972 年生
花園大學文學部講師
主要著作　「データベースがもたらすもの―データベースがもたらすもの」　「思想史としての文字情報処理 問題提起として」　「清辨比量の東アジアにおける受容」ほか多數