

中国古典資料の電子化について (概要)

関西大学・文学部 二階堂 善弘 (Yoshihiro NIKAIDO)

Faculty of Letters,
Kansai University

近年のコンピュータやインターネット技術の発展はめざましいものがあり、これを受けて人文学の諸分野でも大きな変化を蒙っている。中国学においても、大規模な漢籍データベースを利用して研究することは、もはや当たり前になりつつある。

パーソナルコンピュータにおいて扱える漢字の数は、当初は JIS X 0208 が 6 千字程度、また中国の GB2312 がやはり 6 千字くらいであり、古典を表記するには大きな問題があった。例えば「八佾」の「佾」や「墨翟」の「翟」のような基本的な漢字が JIS コードでは表示できなかったことは広く知られている。

その後 Unicode が発展し、扱える漢字数は当初の 2 万字前後から、約 9 万字にまで拡張された。いまや『大漢和辞典』を凌駕する数の漢字が MS Word・メモ帳やジャストシステムの一太郎といった、一般的なアプリケーションで使用できるようになっている。

また漢籍データについては、中国大陸や台湾を中心に大規模なものが続々と作られている。その先駆けと言えるのが台湾中央研究院の漢籍電子文献である。これは『史記』『漢書』などを含む『二十五史』をはじめ、『論語』『老子』などの先秦諸子の膨大なデータをインターネット上で公開したものであり、全世界の中国研究の現場に大きな影響を与えた。

さらに中国では『四部叢刊』(約 1 億字) や『四庫全書』(約 8 億字) といった大部の叢書をデータ化し、CD-ROM に収録して発売した。これらは Unicode のテキストデータのみならず、画像データも併用しており、原典を画像として確認できるのが特色である。

その後も、CD-ROM などの媒体やインターネットを通じて、『古今圖書集成』『全唐詩』『朱子語類』『三国志演義』といった文史哲の主要な資料が続々データ化され、提供された。現在、その最大とされるのは『中国基本古籍庫』であり、数多くの古典籍を網羅しており約 20 億字のデータが収録されている。この中には数学関連の

このような膨大な中国古典籍の電子化が行われ、一般化する一方で、その質も問われるようになってきている。また主要でない版を用いたデータや、有名ではない典籍の電子化も必要とされているが、こういった動きはまだ始まったばかりである。