

木簡画像データベース 「木簡字典」の公開

本年2月、奈文研では木簡に関する2つめのデータベース「木簡字典」のホームページ上での公開を始めました(<http://jiten.nabunken.go.jp>)。

例えば、「簡単検索」で「国」という文字で検索してみてください。いろいろな木簡に用いられた「国」の文字を一覧することができ、同じように「国」と読んでいる文字にも、実にさまざまな字形があることがわかります。選んだ文字を拡大して観察したり、別の撮影方法による写真に切り替えたり、あるいはその文字を含む木簡全体の画像を見たりすることもできます。また、地域による特徴はないか、年代による変化はあるのかなど、さまざまな条件で検索結果を絞り込んでいくこともできます。

「木簡字典」には、従来の木簡データベース(以下、「木簡DB」と略記)と比べると、次のような特徴があります。①「木簡DB」が木簡1点ごとのテキストデータを基本としていたのとは違い、木簡の文字一文字ごとの、しかも画像のデータベースであること、②モノクロ写真だけでなく、カラー写真、赤外線デジタル写真など、異なる撮影方法による画像が複数見られること、③写真だけでなく研究員が木簡を解読した記録である「記帳ノート」も公開していること、④その文字画像が木簡のどういう文脈で用いられているかなど、その文字を含む木簡のデータを随時参照できるようになっていること、⑤専門家だけでなく、一般のユーザの利用をも想定した「簡単検索」を設けるなど、データベースとしてのユーザインタフェースに配慮していること、⑥「木簡DB」では不可能な木簡の大きさの範囲指定を可能にするなど、検索の便が大幅に向上したこと。

これまでにも中国の法書や碑文などを素材として漢字のさまざまな字体を例示する字典はありましたが、日本の資料、それも木簡のような生の史料によって字体を例示する字典、しかもデータベースとしての公開はこの「木簡字典」が初めてです。現在はまだ奈文研で調査したいわば自前の資料を主な対象としていますが、幸い九州歴史資料館の全面的なご協力により、大宰府跡出土木簡のデータも一部収録することができました。データに時代や地域を越えた広がりをもたせるためにも、今後各地の調査機関の



木簡字典トップページ

ご協力を仰ぎながら各地出土のさまざまな時代の木簡を収録できればと考えています。

木簡に使われている文字の種類は1500種余り。今回公開できたのは600種余りですから、データの拡充は今後の大きな課題です。また、熟語など複数文字検索表示も是非実現させたいと思います。ユーザの皆様の声を生かして、よりよいデータベースに育てていきたいと考えていますので、「木簡DB」と同様に、「木簡字典」もご支援ください。

なお、「木簡字典」は、科学研究費補助金(基盤研究(S))による研究「推論機能を有する木簡など出土文字資料の文字自動認識システムの開発(2003年度から2007年度 予定)の成果の一部です。所内のメンバーの他、国際日本文化研究センターの山田奨治(情報学)、東京農工大学の中川正樹(情報科学)、京都大学の柴山守(情報工学)、総合研究大学院大学の及川昭文(数理考古学)、法政大学の小口雅史(日本古代史)、国立歴史民俗博物館の鈴木卓治(博物館情報システム)の各氏に研究分担をお願いし、現在総勢13名で課題に取り組んでいます。標記の課題は言い換えれば出土文字資料解読のOCRシステムの開発ということですが、私たちの発想は解読を機械任せにしようというわけではありません。むしろ機械の力を借りながら、出土文字資料解読の客観化、普遍化、ひいては効率化を図ろうというもので、これまでの奈文研における木簡解読作業の延長上に位置付けられるものです。それは1961年に特別史跡平城宮跡で初めて木簡を発掘して以降、これまで40年間以上にわたって20万点以上に及ぶ木簡を調査しながら培ってきた私たちの木簡解読のノウハウを、形のあるものにしていこうという試みでもあります。「木簡字典」はいわばその第一歩です。

(平城宮跡発掘調査部 渡辺 晃宏)