

変体仮名字形データベースの構築と公開

間淵洋子, 福井尚子

国立国語研究所では、人間文化研究機構の4機関が協同で実施している「異分野融合による「総合書物学」の構築」プロジェクトの一環として、近世後期の版本である『比翼連理花廻志満台』初編3巻、『春色梅児与美』初編3巻、計6巻分の平仮名延べ約30,000字に対して、変体仮名が収録されたUnicode10.0.0によるコーディングを行い、変体仮名字形のデータベースを構築した（異なり字体数135字）。本データベースは、Unicode10.0.0で規格化された変体仮名コードを実資料へまとまった形で適用した初の試みで、(1)変体仮名使用の計量的研究、(2)形態論情報を用いた変体仮名と言語単位との相関解明、(3)Unicode変体仮名の実用性検証、(4)変体仮名の字体機械認識正解データなど、日本語研究だけでなく情報学等周辺領域でも利活用可能な言語資源の開発を目的に設計・構築したものである。

凸版印刷株式会社と共同で開発を行った、字体認定データと『日本古典籍くずし字データセット』の字形画像データを組み合わせたデータベース閲覧ビューア（右図）は、平仮名・変体仮名を「音価」「字母」「Unicode 字体」の三層により階層的に分類表示し、各Unicodeに対応する実資料の字形画像を一覧提示するものである。くずし字資料の原典画像と翻刻文字列とを対応させて表示する機能を持ち、字形画像との相互参照を実現している点に特長がある。本ビューアは、国立国語研究所サイトより無償公開するもので、変体仮名の学習支援など、教育分野での利活用も期待できる。



図 変体仮名字形データベースビューア