

日本・中国・韓国の仏典で用いられてきた外字 2800 字種超が Unicode に収録されました

2017 年 7 月 14 日

SAT 大蔵経テキストデータベース研究会

代表 下田正弘（東京大学大学院人文社会系研究科教授）

<http://21dzk.l.u-tokyo.ac.jp/SAT/>

漢字を共有する諸国において、実際には使用されながらも世界共通のコードが付与されていない未符号化文字、いわゆる Unicode 外字が多数存在する事実は、人類の文化全体を継承してゆくうえで、大きな障害となっています。加えて、これまで、学术界からこの問題について解決の提案をする道が開かれていなかったのは、さらに問題を大きくしていました。

このたび、SAT 大蔵経テキストデータベース研究会（以下、SAT 研究会）では、情報処理学会情報規格調査会 SC2 委員会の多大な協力のもと、仏典〈大正新脩大蔵經〉のなかに収録された 2800 字超の漢字を、Unicode 10.0 において符号化するに至りました。これは、ISO/IEC JTC1/SC2/WG2/IRG における関連委員会に対する、学术界からの提言がはじめて実を結んだものであり、これによって、今後、学界から Unicode への提言をおこなう道が開かれました。ここにご報告申し上げる次第です。

わが国において、1924 年から 10 年をかけて刊行された〈大正新脩大蔵經〉は、東アジアの仏教伝統に対する文献学的な研究の集大成として、当時の日本が世界に示しえた最高水準の成果であり、公刊以後、広く人類文化の考究に貢献をしてきました。

この仏典の集成を、デジタル知識基盤に適切に継承するため、1994 年に SAT 研究会が設立され、2008 年には 1 億字を超える全文検索サイトを公開するに至りました。

しかしながら、その時点では、テキストデータとして一律に表示しえない約 6000 字種超の Unicode 外字を含んでいたため、データ全体が効率的に扱えないのみならず、世界共通の文字基盤における知識継承という点において、課題を残したままでした。

この過去二十年ほどは、漢字のコード化をめぐる、さまざまな試みが現れ、国際的な環境が変動しつづけた時期に当たります。SAT 研究会では、これらの諸事業と協働しつつも、長期的に安定的な方向に向けて、独自の可能性を模索してきました。

この過程で、6000 字種超の外字全体をデータベース化するとともに、現時点での利便性を可能な限り高めるための方策として、Unicode で利用可能とすることをめざしました。これによって、デジタル媒体上で効率的に利用しうるのみならず、それらの文字が国際的な標準規格のなかに位置づけられ、一つの学界を超えたより大きな枠組みによって維持さ

れてゆく仕組みをつくることができました。

この成果は、世界中の仏教研究者、デジタル・ヒューマニティーズ研究者をはじめ、多くの関係者の共同作業と、(独法)日本学術振興会(科学研究費補助金)、(一財)仏教学術振興財団、(公財)全日本仏教会、大蔵経研究推進会議、(一財)人文情報学研究所をはじめ、さまざまな機関、団体の支援によって、達成しえたものであります。

なかでも、漢字の効率的な探索と比定を可能にする、守岡知彦氏(京都大学人文科学研究所)作成のCHISE、フォントの共同作成を容易にする、上地宏一氏(大東文化大学外国語学部)作成のGlyphWiki、自由に利用可能な花園明朝フォント、これらの先行するプロジェクトの成果は、本事業の推進に不可欠でありました。

文化の基盤を構成する重要な要素である文字を地上から消失せしめないことは、その文字を地上のあらゆるところで利用可能にすることと、実質的に同義であります。これを達成する事業は、単独のプロジェクトによってではなく、同時代に存在する関連するプロジェクトと協働することによって効果的に遂行しうることが、いよいよ明らかになってきています。SAT研究会は、今後、関連事業との連携、協働をいっそう強化し、人文学全体の知識基盤構築に向けて、着実な取り組みを進めてゆく所存です。