

chapter.10

慧琳撰『一切経音義』の符号化をめぐる

王一凡

1. はじめに

本稿では、SAT 大蔵経テキストデータベースプロジェクトによる『大正新脩大蔵経』デジタル化の取り組みの中で、特に慧琳撰『一切経音義』という特殊な文献と、その符号化にまつわる課題と知見を紹介する。

慧琳撰『一切経音義』とは漢文諸仏典の用語を解説する音義書として大正新脩大蔵経に収録されている文献だが、その出自や内容、性格ともにほかの大多数の収録経典とは一線を画している。次に、符号化とは文字をコンピューターが一般に処理できるよう特定の数値データに変換、あるいは表現することであるが、これはわれわれが多くの場合何ら考えることなく文字を入力できるように、通常は文字に関する一般常識的観念があれば問題なく達成できるはずの行為である。しかし慧琳撰『一切経音義』の場合、活字本の大正新脩大蔵経にありながら膨大な異体字数と字種が集中する部分であり、歴史的文献であることもあいまって、相互によく似た字形をどう取り扱うべきか否か判断に困ることも少なくない。一方で符号化のためには文字の意味と同一性を吟味しなければならないため、ここに問題の複雑さが現出する。本稿ではまず当該文献の背景と収録に至る経緯を概観し、特筆すべき性質やその重要性について述べる。そしてこれを例にとり、古文献の符号化にあたって、紙資料側から何を取り出し、デジタル側へどのように送り込むべきかという問題を考えていく。

2. 資料としての慧琳撰『一切経音義』

2-1. 慧琳撰『一切経音義』という著作

慧琳撰『一切経音義』は、その名の通り一切経の音義、すなわち大蔵經に収録されている仏典の語句を解説する文献である。唐・長安の西明寺において学僧慧琳により8世紀末から9世紀初にかけてあらわされたと考えられている。

ここで「慧琳撰『一切経音義』」と呼んだのは、「一切経音義」という名で知られる文献はほかにも存在するからである。より有名な『一切経音義』は慧琳の書に先立つこと一世紀あまり前に成立した、玄奘げんじょうの同時代人でもある玄奘げんのうによる著作である。玄奘の書は元来『大唐衆経音義』と題されており、これを「一切経音義」と称するのは一種の俗用と考えられるが広く呼び習わされている（慧琳のものは『新収一切経音義』とも題す）。玄奘の書は当時存在または新訳されたうちの400以上の經典についての注釈を計25巻にまとめたものである。これに対し慧琳は玄奘を含む先人の音義書を集成しつつ、開元釈教録の順に1200部あまりへの注釈を100巻にまとめた。従って形式的には慧琳の著作は玄奘などの著作に対する増補と捉えられるが、内容的にはそれらの部分も慧琳の方針に従って再編集がなされており、単純な引用ではない。

なお、これらのほかに遼の希麟きりんにより10世紀頃にあらわされた、慧琳までに未収の經典を集めた『統一切経音義』や、可洪かこうによって後晋の頃あらわされた『蔵経音義随函録』が一般に「一切経音義」と呼ばれる。以下、これらを撰者の名を冠して玄奘音義、慧琳音義などと呼ぶ。

慧琳音義の一つの大きな特徴は、その音注おんちゅうの独自性である。音義書は辞書であるから、掲出語について読み方を記載しているが、玄奘音義が中国語の発音に関して『切韻せついん』と一致度の高い体系をなしているのに対し、慧琳は玄奘の体系を拒否して慧琳自身の時代に現地で話されていたと思われる長安音（秦音）を一貫して採用している。中国音韻学では『切韻』の音系が重視されるが、中国国外に広まった、日本語の漢音を含む域外漢字音は慧琳音義の体系により近いことが明らかになっており、漢字音の研究上重要な資料に位置づけられる。加えて、西域の疏勒（現カシュガル）出身と伝えられる慧琳によって、インドのみならず西域の語彙・文物についても詳細な注釈が施されており、一級の対

音資料としても扱われる。

慧琳音義はまた、字句の解説のために当時存在していた辞書類や古典籍（儒教・道教経典なども含む）から膨大な数の引用を行っている。これは先行する玄応音義でも同様であるが、慧琳音義の方がその内容量に比例してはるかに資料として豊富である上、玄応が利用していない系統の文献を引用している（秦音系の韻書など）。これらの引用群は、現代では逸失した多数の文献の姿を明らかにする貴重な（多くの場合唯一の）証拠として高い価値を有する。特に著名な例としては、フランスの東洋学者ペリオが、敦煌^{とんこう}から持ち帰った原題不明の文献を慧琳音義との内容の一致をもとに慧超^{えいちょう}『往五天竺国伝^{おうごてんじくこくでん}』に同定したことがあげられる。現存する文献であっても、その異本や古態を知る手がかりとして文献学的に極めて有用である。特に中国国内に残存する文献は、成立が古い資料であっても後世の改訂を経た新しい版であることが多いからである。

漢文仏典の語法には、いわゆる漢文として知られる伝統的な文語のみならず、当時使われていた口語表現が数多く含まれている。これを解説する音義書は必然的に当時の中国語に関する言語学的資料としての性質を持つ。さらに、音義書は当時主流であった韻書とは異なり、字書というよりも辞書、すなわち語句単位で注釈を加える形式のため、見出し語はほとんどが二字以上からなる語である。慧琳音義の3万項目にわたる掲出語からは現代中国語につながる複音節語の発展過程を如実に見て取ることができ、中古中国語の実態を映し出す包括的な現存資料としては類を見ないものである。

また、これは文字の面でも同様である。伝世文献からは『干禄字書^{かんろくじしょ}』をはじめとする唐代の字様書以前の用字環境はほとんど確認できず、古写本や敦煌文献、碑文などから垣間見るほかない。しかし慧琳音義からは、本文または引用した経文を通して、当時かなり多様な異体字が通用していたことがうかがいられる。音義書はしばしば経文の実例をあげながら規範的な字形に修正しており、そこから当時の用字意識と実態を系統的に対照することができるが、慧琳音義は宋以来の標準に基づく改版の影響をほとんど受けていないと考えられるため、より生の形の文字が豊富に残されていると考えられる。

このようにして、慧琳音義はその時代性と網羅性とあいまって、当時の言語文化を知るためのミッシングリンクとして、大蔵経の解説書という枠を大きく

超えた現代的な価値を持つに至るのである。

■ 2-2. 成立と伝承

慧琳音義の伝承経路は、かなり紆余曲折を経たことが知られている。

慧琳音義の成立年には諸説あるが、慧琳音義冒頭に付された景審の序によれば唐の元和2年(807)、『宋高僧伝』によれば元和5年(810)に完成し、西明寺に収められたとある。『仏祖統紀』には元和2年に慧琳が「一切経音義百三卷」を朝廷に献上し入蔵を認められ、紫衣などの褒賞を賜ったとの記事がある。ただし現行本の慧琳音義には開成5年(840)の顧斉之の序が加えられており、さらに『宋高僧伝』は大中5年(851)に入蔵して後に流行したとのみ記載している。これが単なる『宋高僧伝』『仏祖統紀』間の史料上の齟齬か、会昌の廃仏(845頃)の影響によるものかは不明だが、成立当初より評価の高い著作であり、9世紀中頃までは少なくとも西明寺のある長安で利用可能な文献であったことがうかがえる。

慧琳音義はその高い評価にもかかわらず、成書後唐末の動乱に見舞われて広く流通せず、長い間中国大陆では幻の書となっていた。その間の経緯は高田(1994)^{*1}に詳しい。

陳垣は、廣明元年(880)黄巢が長安に侵入したとき、京師にあった慧琳音は滅んだとしている〔中略〕續く唐末の戦亂は慧琳音の流行を著しく阻害したに違いない。のち可洪が『隨函錄』の撰述を行ったのは十世紀半ばの河中府においてであったが、それでも慧琳を利用し得るに至らなかったのである。周の顯徳年間に高麗使節が江南の地に慧琳音を求めても、得られなかったことは蓋し當然のことと言えよう。同じ頃越地において、慧琳音の傳わらないことが行瑠の音義撰述の動機になったというのも、慧琳がすでに極めて希覯の書であった證左である(『宋高僧傳』卷二十五行瑠傳)。江南に傳わらなかったばかりではない。慧琳音は敦煌遺書のうちにも見られない。

實際、後世の書にしても慧琳を引くものはほとんど皆無である。ただ契丹には燕雲十六州の歸屬とともに慧琳音が傳わったとされる。〔中略〕慧

琳音が今日に傳わるのは、唯一高麗藏本によってであるが、それは契丹から高麗に傳わたものと考えられている。[中略] 現在、世上に流布する慧琳音義はほとんど全てがこの本に由來する。

高麗藏は希麟音義も収録しており、同じく現在に伝わる唯一の祖本となっている。この高麗藏（再雕^{さいちよう}本と呼ばれる）は木版本であり、さらに版本は廃棄されることなく現在に至るまで保存されている。後世求めに応じて重刷することが可能であったため、日本にも多数の刷本が伝來した。江戸時代に獅谷^{ししがたに}（鹿ヶ谷）法然院^{にんちよう}の忍^{にん}激^{きやう}上人（1645–1711）が、建仁寺所藏の高麗藏中に中国に伝わらない慧琳・希麟の音義があることを発見し、これを増上寺の高麗藏とも対校しながら翻刻した。翻刻は忍^{にん}激^{きやう}の弟子らの手により元文3年（1738）に出版された。希麟音義も延享2年（1745）に高野山の僧^{ぎやうしやう}堯昌によって翻刻された。両本ともに獅谷白蓮社の名を取り白蓮社本と呼ばれるが、後者は高野山北室院版との印記を持つものもある。この白蓮社本は清末光緒年間に楊^{よう}守^{しゅ}敬^{けい}が中国に持ち帰り、それによって慧琳音義が再度中国に伝わったという経緯がある。

近代になって刊行された慧琳音義は、初の活字版藏経として明治時代に開板された大日本校訂縮刷大藏經（縮刷藏）に収録されたものと、民国初期に上海で開板され、縮刷藏を底本として校訂した頻伽^{びんが}精舎校訂大藏經（頻伽精舎本）のもの、大正から昭和にかけて新規の活字で刊行された大正新脩大藏經（大正藏）のもの、また最近中国において諸本を集成して校訂した中華大藏經のもの、計4種類の版が存在する。このうち、大正藏は仏教学者らによって当時最新の研究成果を盛り込んだ学術版として世界の研究者に普及し、現在に至るまで標準テキストとしての重要な地位を占めている。

大正藏は、当時最善本とされた高麗藏を底本に使用したと謳っているが、実際の作業では入手が容易な頻伽精舎本から版を起こしながら、高麗藏と対校させていたようである。大正藏が下敷きにした頻伽精舎本は、おおむね縮刷藏の影印に等しいのだが、こと慧琳・希麟音義のみは白蓮社本を底本にした経緯が、頻伽精舎本『一切經音義』の序文に記されている。

本精舎重印大藏即據弘教本爲式音義一部舊有六種夾注縮印每遇繁密處手

民苦其朦混謝弗能任宗公與予謀琳麟二書別取澁師刻本裁爲兩層縮以石印大小如今本用備藏經之數餘四種無單本可得即姑置之

「本精舎は大藏經復刻にあたり弘教本〔縮刷蔵のこと〕を基準とした。原本には音義書が6種入っていたが、割注の縮印で込み入った箇所になるたび、植字工が繁雑さを厭い仕事を引き受けてくれないので、宗公と私は、慧琳・希麟の両書を別に忍澁師の刻本〔白蓮社本のこと〕から取り、裁断して二段にし、リトグラフで本書のサイズに縮小して蔵經の数に含めようと考えた。残り4種は単行本が得られなかったのでひとまず差し置いた。」
(訳は筆者)

実際に頻伽精舎本の版面を見ると、両音義の部分のみは白蓮社本の二張を1ページに並べて縮小製版したような体裁となっている。従って、大正蔵は全体として縮刷蔵系統のテキストを引き継いでいるにもかかわらず、この部分のみは白蓮社本に由来していることになる。さらに大正蔵は頻伽精舎本で取り除かれた頭注を原本である縮刷蔵によって補っているが、慧琳・希麟音義は本文が縮刷蔵に由来しないため、校訂情報と本文が対応しないという問題がある【図1】。

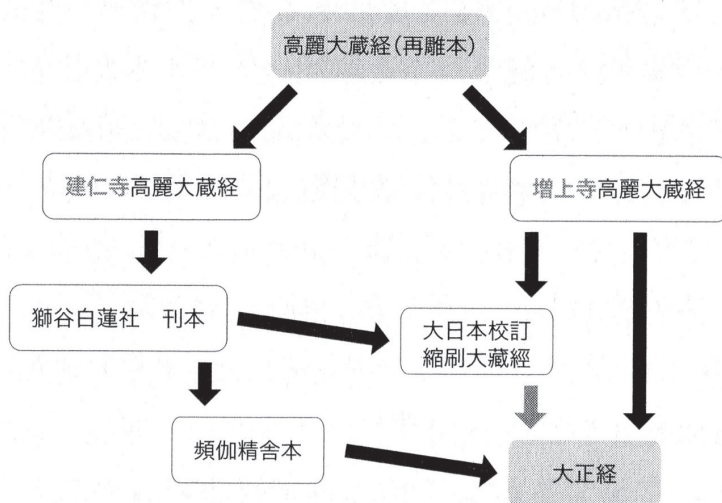


図1 大正蔵慧琳撰『一切經音義』の来歴 (永崎 2016) ²

大正蔵は当初和紙・線装の和装本と洋紙・ハードカバー製本の洋装本という二形態で刊行され、これによって諸外国にも広く流通した。しかし、この初版本は戦災により紙型と在庫本の一切を喪失したため、資料が残っていない。戦後大正蔵を写真製版により再刊した際に、校訂の「正誤表」を作成しているという記述が垣間見られるが、その実物は管見の限り確認できない。そのため、現在利用できる大正蔵の編集資料は、若干の関係者の回想に基づくものを除いて極めて乏しい。

2-3. 諸本の形式

以上に述べた通り、現在伝わる慧琳音義の諸本には高麗蔵本・白蓮社本・縮刷蔵本・頻伽精舎本・大正蔵本・中華大蔵経本が存在するが、版式からみると頻伽精舎本は白蓮社本の影印、中華大蔵経本は高麗蔵本の影印をもとにしているため、高麗蔵本・白蓮社本・縮刷蔵本・大正蔵本の4種類に分類できる。以上の諸本はすべて整版または活版による刊本であり、確認できる限り慧琳音義に写本は残っていない。これらの本にはそれぞれ形式上特徴的な点が見受けられる。

高麗蔵には、玄応・^{え おん}慧苑（華嚴経音義）・慧琳・希麟・可洪による音義書が収録されているが、玄応・慧苑・希麟音義はいずれも一張が横長（20～25行）で版心のない体裁をなしているのに対し、慧琳・可洪音義は中央に版心のある版式となっており、これらの底本が卷子本ではなく冊子本であったことを推測させる。さらに、可洪音義は一張の幅が前三者と同等なのに対し、慧琳音義だけがいちじるしく短く、3分の2程度という特異な形状をしている【図2】。

高麗蔵を翻刻した白蓮社本は、版式をまったく改め、慧琳・希麟音義ともに明の^{ばんれき}万暦版（というよりもその日本における覆刻である^{おうばく}黄檗版）に似た縦長の明朝体のような字形で、まれに頭注が付されている。黄檗版は当時普及していたとみられ、それとの統一性を念頭に置いていたと思われる。整版ではあるが、行数・字数がよく整っており、題字・注記・掲出字・割注の別もわかりやすい【図3】。

縮刷蔵は近代に導入された活版印刷を用いており、一字のサイズが従前の木版本と比べ極めて小ぶりになっているばかりではなく、行間が密に組まれてい



図2 高麗蔵所収の玄応・慧苑・慧琳・希麟・可洪音義の版面

るのが特徴である。この版式は意図されたものであり、発起人である島田蕃根^{ばんこん}の回想によれば、活版で印刷されているキリスト教の聖書と同等の普及性と携帯性を意識したとのことである。印刷に使用した活版所は、協力者のひとりがある所有していたものを借り切って使ったとのこと、活字は独特の風格を持っている。版面はすべて五号・六号相当の寸法の活字で組まれている【図4】。

大正蔵も活版印刷であるが、全体を3段組としているのが特徴であり、これは先行する^{まんじ}卍字蔵の影響かもしれない。慧琳音義部分については掲出字ごとに行を改めるなど、余裕を持った組版を行っている。使用活字の寸法は四号（題のみ）・五号（本文）・六号（割注）であり、音義書の大半を占める割注は縮刷蔵よりも判読しやすい行組になっている。字形も現代に通じるスタンダードなものであり、字のほとんどは秀英舎の活字を使用していると思しい【図5】。

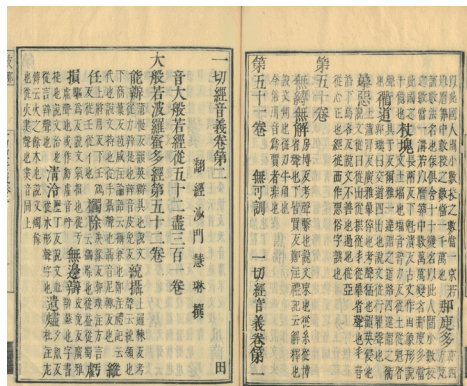


図3 白蓮社本慧琳音義の版面

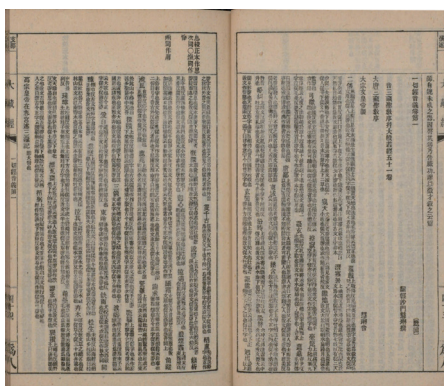


図4 縮刷蔵本慧琳音義の版面

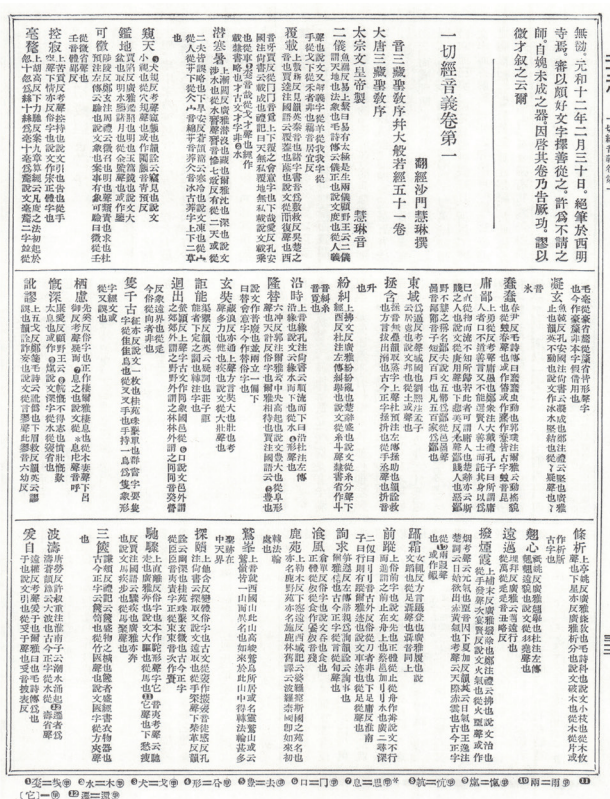


図5 大正蔵本慧琳音義の版面

3. デジタルテキストにおける文字

紙に書かれた文字をデジタルデータに変換するというとき、具体的には何を何に変換すべきなのだろうか。

まず考えられるのは、文字の外形、つまり紙面に印されたインクの形を写し取ることである。計算機は色を数値に置き換えて処理できるため、色情報を持った点の配列を整列して表示させれば、その点が十分精細である限り、網点印刷と同様の原理により視覚情報を再現できる。このようにして紙面の色彩を点の集合に置き換えたデータを人間に見せて同等に見えれば、変換できたと見なせるだろう。この場合、紙媒体に残る立体構造が一切捨象されるのを除けば、極めて現物に忠実に再現することも可能であるし、手法次第では肉眼で現物を観察する以上に文字情報が判読できることもあり得るので、保全が必要な原本に代わって公開するのにも役立つだろう。デジタルデータは原媒体と比べてはるかに物理的なサイズが小さく、現代の環境であれば複製、持ち運びが容易というメリットがある。一方で、本文の読解、分析に関する利便性は当然ながら原本と同等にとどまる。このような手法はデジタル技術を利用した影印または覆刻の進化形といえる。

一方で、文字が内容の主体である文献においては、文字の抽象的な単位をデジタルデータに変換するという手法も考えられる。われわれは字のイメージを思い浮かべながらそれを紙の上に転写する。手書きでは物理的にまったく違うの字はほぼ書けないが、それでも一文字ごとに違う字のつもりで書いているわけではなく、限られた種類の字を繰り返し投影しているはずである。そこで、区別する必要があるイメージの数を調べ上げ、すべてに識別可能な名前（典型的には特定の数値）を与えることで、紙の資料と同等の内容の区別が実現できると期待される。これは読者と計算機がある意味で共通の字の認識手段を得たことを意味するので、「機械可読」すなわち検索や照合、索引づけその他の処理を計算機が高速で代行できるという恩恵を受けやすい。つまりより進化した媒体への翻刻と同等である。ところでこの手法は活字印刷の原理に通じるものである。それがあくまで紙上の図形に還元する前提で、活字自体は中間段階にすぎないのに対し、デジタルデータはいうなれば活字の鋳型の状態で流

通させ、表示のたびに計算機の内部で鋳造し、画面に印字させようというものである。そのため理想的にはすべての機器に同一規格の活字棚を備え付けさせるのが望ましい。こうした背景から、デジタル世界の文字はそれを用いる各種業界、国、さらには国際的な標準によって規定されてきた部分が多い。学術的な文字の利用と流通のためには、それらを最大限利用しながら、個別用途に応じた変更と補完を考えなければならない。

4. 紙本からみた符号化

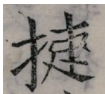
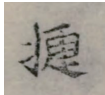
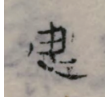
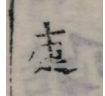
4-1. 字形の問題

紙の資料を符号化するとした際、前提的ではあるが無視してはならない問題は、その文書の作成者の意図が正しく紙の上に現れていると考えられるかという点である。慧琳音義に関していえば、現存する諸本がいずれも刊本の漢籍、さらには辞書に類するものであるという性質が特に問題となる。刊本であるということは、一面では転記による変異を抑えられることを意味するが、もう一方では機械的な複製によって生成されるために、作成者のあずかり知らない変異が混入しがちということである。これは具体的にいえば印刷のかすれ、にじみ、版の摩耗などがあげられる。写本であるならば必ず作成者の制御の下で生成されるので、不適切な形であることに気づけば修正や注記などの手段をとりうるが、刊本では版を介した間接的な関与となる（完成した版が関係者の死後にさえ複製される）ため、この要素の影響が顕著になる。また、欧文などの表音文字で記述された資料であれば、限られたレパートリーの制約を利用して、不完全な字形からも本来あるべき内容を復元しやすいが、漢文資料では出現する字形の制約が立てにくく、読み取った字形の真正性が担保しにくい。さらに慧琳音義はいわゆる古辞書にあたるため、登場する字は佚文ないし伝世本には存在しない記述を引用しているなど、とにかく並行する文献が得にくいことが多い。そのため単純な誤脱かどうかの判定すらままならないケースがある。

次の表は同一項目中に出現する4箇所の変遷を古い順に示している（大正蔵54巻p.385）。これらは現代でいう「捷」「走」として読めるが、もとの高麗蔵においても同じはずの字が同じ項目中にもかかわらず明らかに異なる

形で現れている。これを解釈した白蓮社本や大正蔵はそれぞれ別様に異なる字形を宛てており混乱を増している。後半の「走」は高麗蔵では全体的に不鮮明であり、白蓮社本の字形は根拠が薄弱に見えるが、現存しない古い建仁寺の刷りを参照した可能性も考えられ、本来の字形の推定を難しくしている。【表1】

表1 慧琳音義卷第十三「捷慧」項の「捷」「建」の字形

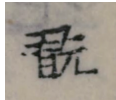
高麗	白蓮	大正
		
		
		
		

また、ごくまれに印刷の状態に起因して、原本とはかけはなれた字形に変化してしまうことがある。極端な例を挙げると、大正蔵 54 巻 p. 697 の「烏翫反」の例がある。高麗蔵では「翫」の字として読めたものが、白蓮社本の細長い字形に転写され、頻伽精舎本で縮小された結果、筆画の構成が不明瞭になってしまった。恐らくこの状態の字を参照したのでであろう大正蔵の採字担当者は、字形を正しく解釈できず、図に見えるように上辺を「止」に統合し、「月」「凡」からなる存在しない字を作ってしまったものと推測される。この場合、高麗蔵の字形は明瞭と思われるため、対校をどの段階で行ったのかなど、編集過程の復元に関して示唆するものがある。【表2】

4-2. 字種の問題

前節で述べた字形の認定に疑いがないような場合であっても、どの程度の字

表2 慧琳音義卷第五十八「驚惋」項の「翫」の字形

高麗	頻伽	大正
		

形差を「有意な」ものとして扱うかは、なお検討を要することがある。人が書き、また物理的な模様として印字される以上、個々の字（字形）は多種多様な面で異なりを見せるが、作成者にとってはあくまで少数の意図した字（字種）の実現形であるはずである。版下から起こす整版では、同一の文字を複数書こうとしても自然と微妙に異なる形にならざるを得ないのはもちろんだが、慧琳音義では字の構造が異なったり、（一つの釈文の中でも）筆画が増減したりということもしばしばである。単体の著作であっても、慧琳音義のように大部になると最初から最後まで完全に整合性がないことも考えられる上、ある変異が成立当時は無視すべき差異であったのか、後世の訛伝による偶発的な変異なのか判然としない面もある。慧琳音義の成立当初は写本によって伝わったと考えられ、また敦煌文献に現れる字体を見るに、原文献がどれだけ楷書的であったかも疑問が残る。

これに対し、大正蔵のような活字母型^{かつじ ぽけい}レベルでの同一性が保証できる版本を字種の区別の根拠と見なすことはできるだろうか。大正蔵の慧琳音義を調べたところ、一部の活字について意味上の違いが明白ではないにもかかわらず、活字の形状ないし字の結構が異なる場合があることが判明した。

図では、「傳」および「並」「竝」の活字それぞれ数種の分布を示している。「傳」の場合、字の構造には違いがないが明らかに印影が異なる字形が相補的に交替しながら出現している。また「並」「竝」の二字は音義ともに同一の異体字だが、同様に使用される区間に相補的な偏りがあることがわかる。このように常用の字であっても形態が変わりうる以上、ある出現頻度の低い字形群の同一性を機械的に判断することは難しい。この字形の交替について、大正蔵本の内部比較からのみでは全体に通底する規則を見出すのが難しく、他本を含めた字形コーパスの構築を試みている【図6】。

また、以上の議論では慧琳音義諸本に通底する字種の識別や認定ができると

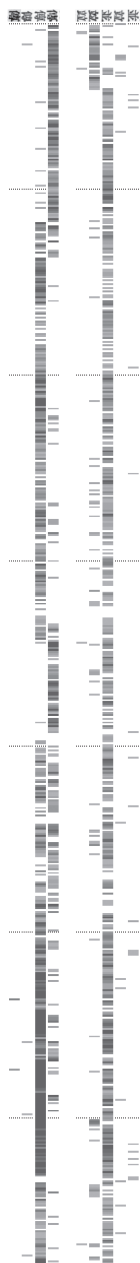


図6 大正蔵本慧琳音義における「傳」および「並」「竝」の活字形分布

いう前提に立っているが、実際にわれわれが目にすることができる慧琳音義のバージョンは、それぞれ時代（13世紀～20世紀）・地域（中国～朝鮮～日本）・手法（木版～活版）・背景を異にする刊本の集合であって、しかもそれらは連続的に伝承されてきた著作の現存する一部というよりも、ごく個別的な状況で作成された独自性の高い編集版の寄せ集めである。そのため、個々の本が立脚している字種の認識のしかたの違いも無視できない要素となる。慧琳音義は大正蔵においても膨大な異体字を包含しているが、それは底本にある字形をすべて忠実に写す方針だからではなく、通常は元の高麗蔵に対して異体字は整理されており、しかも当時「普通」と思われていたであろう字形に置換している。図7に大正蔵・白蓮社本・高麗蔵からそれぞれ同一箇所の画像を並べたが、掲出字である「胃（絹）索」をはじめ、𠂔（あみがしら）に部品の形状の規則的な異なりが現れているのがわかる。このように対応が綺麗に取れるものは、字種を抽出する手がかりとなる重要な情報である。【図7】

一方で、諸本（の編者）間の認識の衝突によると思われる意図しない現象も散見される。例えば、慧琳音義卷第三十九の「贅薬」「蓮薬」両項目（大正蔵54巻 pp.563, 565）中に「蕤」の異体字が出現する。これを大正蔵では「豕」を「麥」に置換した字形（蕤）で示すが、他本ではいずれも「麦」であった。つまり「豕」が草体化したために「麦」の形に変化したものを、大正蔵の編纂当時には「麦」という形状を見て当然「麥」の略字体であろう（現代でも同じく「麥」はムギを表す「麦」の旧字体とされる）との連想が働き、「正字」に修正してしまった結果、むしろ原字とはかけはなれた形となって現れたものと考えられる。

また、白蓮社本についても類似した例を徐（2002, p.153）³

が挙げている。慧琳音義卷第九十七の「方冊」の項（大正蔵 54 卷 p. 910）の中に、高麗蔵では「象其礼」と書く箇所があり、これを白蓮社本では「象其禮」と変えている（これも現代の新字・旧字関係にあたる）。しかしこの一節は説文解字からの引用であり、本来は「象其札」としなければ意味が通らない部分だったため、字形に対する規範意識がかえって原文を遠ざけてしまったことになる。なお、大正蔵でも白蓮社本を引き継いでこの「正字」を採用している。

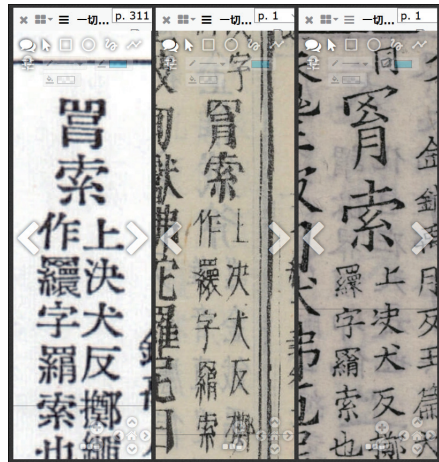


図7 慧琳音義卷第二十五「胃索」項の諸本対照

4-3. 校合の問題

船山（2008）⁴ が指摘するように、大正蔵全体では縮刷蔵の影響が大きくみられる。『一切経音義』について言えば、少なくとも管見の限りでは文献内に付してある脚注の多くが縮刷蔵の頭注の引き写しであるように見える。しかし図 1.2 に示したように、縮刷蔵が行ったのは白蓮社本と増上寺蔵高麗蔵との対校である（白蓮社本の底本となった建仁寺の高麗蔵は焼失していた）。一方、頻伽精舎本は白蓮社本をもとに独自に本文の修正を行った。そのため、増上寺蔵経と頻伽精舎本を見比べても、同じく増上寺蔵経と白蓮社本を対照させた縮刷蔵の校勘記とは噛み合わないことがある。はなはだしくは本文に現れる字形と脚注に示された字形が矛盾していることもある。本文に果たして縮刷蔵の影響がどの程度見られるかはまだ判然としないが、現在までに、ごく少数の字について大正蔵の本文字形が高麗蔵や白蓮社本系統には見られず、縮刷蔵とのみ一致するという例が認められる。

また、大正蔵『一切経音義』は本文中でも目録でも、原本が高麗蔵、対校本（甲本）が頻伽精舎本であるとししか記載されていないはずが、三カ所の脚注にのみ「乙本」の注記が見られた。この注記箇所をみると、すべて縮刷蔵の頭注と同一の内容であり、「乙本」とは縮刷蔵を指している可能性がある。「乙本」注記

は非常にまれであるが、この注記の前後にある脚注が独自の校勘であることを示唆しているかもしれない。

いずれにしても、脚注と本文に相関がなければ脚注の字形を無視することも可能だが、現状はそれほど単純ではないように見える。大正蔵における脚注字形と諸本との対応については、区間による傾向性がないかどうかを含め、引き続き調査していく必要がある。

5. 技術からみた符号化

5-1. 文字集合の適用

符号化した文字を取り扱うには、その文字を表す符号をほかと区別する手段と、そのリストないし対応表（集合）を設定しなければならない。漢字のように字形の組み合わせが有限ではない体系では、定められた汎用的な文字集合に存在しない文字と遭遇する可能性が常に存在し、一般に「外字」として知られている（欧文などでも歴史的文献には珍しくないことである）。

SAT プロジェクト初期の1990年代において、日本にはすでにいくつかの公的な規格が存在したが、それらのカバーする集合は大正蔵に出現する字種と比較するとはるかに小規模であった。それらを補うため、字典や特殊用途文字を集成した大規模文字集合として先駆的であった今昔文字鏡の符号表を用いて表現し、さらに足りないものはSATが独自に符号を付与していた。

2000年代に入ると、今昔文字鏡の許諾条件が改定されて従来のように自由に利用できなくなったため、外字の表現に今昔文字鏡の代わりに利用しやすいGT書体コードを採用することとなった。このような規格の裏付けがない文字集合を用いてデータを作成することはそれ自体簡単な手段ではあるが、符号化テキストの解釈や処理に特別な設定と配慮が必要になってしまうこと、特定の環境に依存するために相互利用性を損なうこと、また外部の定めた集合に依拠するとその安定性に左右されやすいことといった問題点がある。

現在、SATではUCS/Unicode⁵と呼ばれる符号化文字の国際規格に、直接従来の外字を追加していくことにしている。これは1990年代初めより存在しているが、近年普及が進み他の規格をほぼ代替しつつあること、従来の公的規格

をほぼ完全にカバーし終えてなお符号空間に余裕があり、大量の追加に懸念がないこと、そしてそれを背景として、多方面からの符号化要求を積極的に受け入れはじめていることから、SAT 外字の解決先として適していると考えている。UCS/Unicode は業界における統一規格を志向しており、収録された文字は現在利用できるコンピューターでの標準対応がほぼ保証される。したがって、上述のアプローチとは異なり、そこへの追加は「外字」の駆逐を意味することになる。SAT は 2012 年に最初の主に『一切経音義』用字以外からなる漢字集合を提出し、2,800 字余りが独立の符号を与えられて 2017 年に収録された。残る『一切経音義』用字については引き続き調査のうえ符号化を進めており、それも執筆時現在、一部が実質的な審議を終えて規格化を待っている段階である。

5-2. 制度とのかかわり

現在 SAT は UCS/Unicode に外字を追加する方針に舵を切ったが、それは国際規格の世界に独特の動向や制度と整合しなければならないことを意味する。特に UCS/Unicode は JIS などと同様、電気通信分野の工業規格という位置づけであるがゆえに、単に「学術上必要な文字を追加する」では済まされないような異質な諸概念を勘案しなければならない。

まず UCS/Unicode での漢字の取り扱いを説明する。各文字はそれが属する文字体系ごとに分類されて収録されるが、漢字は通常 CJK 統合漢字 (CJK Unified Ideographs) ^{*6} と呼ばれるブロックに配当される。ほかに CJK 互換漢字 (CJK Compatibility Ideographs) という、各国の既存規格との技術的な整合性を取るために存在する特殊な領域がある。一般に「漢字」と呼ばれるものはほぼすべてがこれらにおさまる。執筆時時点で CJK 統合漢字の総数は 87,887 文字を数え、UCS/Unicode 全収録符号の 3 分の 2 を単独で占めているが、まもなく慧琳・希麟音義用字を含む次の約 5,000 字が追加される予定である。

表 3 UCS/Unicode における漢字の収録状況

ブロック	字数	収録時期
CJK 統合漢字	20,902	Unicode 1.0.1 (1992)
…の追補分	74	不定
CJK 統合漢字拡張 A	6,582	Unicode 3.0 (1999)

CJK 統合漢字拡張 B	42,711	Unicode 3.1 (2001)
CJK 統合漢字拡張 C	4,149	Unicode 5.2 (2009)
CJK 統合漢字拡張 D	222	Unicode 6.0 (2010)
CJK 統合漢字拡張 E	5,762	Unicode 8.0 (2015)
CJK 統合漢字拡張 F	7,473	Unicode 10.0 (2017)
CJK 統合漢字拡張 G (予定)	4,938	未定
CJK 互換漢字	472	未定
(…のうち統合漢字)	(12)	Unicode 1.0.1 (1992)
CJK 互換漢字補助	542	Unicode 3.1 (2001)

通常、新しい文字体系が登録される際は、はじめに判明しているほぼ完全なレパートリーが一度に入り、その後もしあれば必要に応じて小規模な追加が行われる程度であるが、漢字は原理上無制限に字を合成できる性質があるため、膨大な字種が存在しており、既存字の概数すら定かではない。そこで、漢字に限っては専門部会が設けられており、定期的に募集と審議を行って、通過したものがその都度追加される運用となっている。特に審議が必要な理由の一つは UCS/Unicode が採用している「包摂」と「統合」のメカニズムである。多数の代表から提案⁷される多数の字を統一的に処理するため、国・地域を問わず同一視する字形の差異の基準を設け、同じと見なされる字の重複登録を避ける仕組みが整備されている。一方でこれによって地域固有の字形が損なわれるという問題が長らく存在し、現在では任意で字形選択子 (Variation Selector)⁸を付加して個別の表示形を指示するという手段が導入されている。この方式で特定の漢字 (字形) を表すことを特に IVS と呼び、公式に管理されたリストが存在する⁹。IVS は現在のところ主に印刷業界や戸籍・登記などの行政用途への利用が多く、これによって区別される字形を算入するならば、UCS/Unicode の漢字総数は少なくともすでに 10 万字を超えていることになる。

現在 UCS/Unicode に漢字を登録するためには、事実上常設の担当部会である IRG (Ideographic Research Group) に参加するか、文書で要望を提出する必要がある。IRG は本来 ISO の傘下にあるため、公式に字を提案できるのは会員資格のある主体、多くの場合国の代表に限られる。個人などからの提案でも妥当と判断されれば、通常は会員に準ずる米国 Unicode 委員会が代行して追加提案を行うが、独自の名義で提案資格を得る道もある。SAT は 2016 年より上位委員会 (ISO/IEC JTC 1/SC 2/WG 2) でのリエゾン資格を承認されたため、現在

は独自主体として追加提案を行えるようになった¹⁰。IRG は年 2 回会議を招集し、その場で議論と決定を行い、勧告を発する。IRG で審議される字の種類や数は各代表の持ち込み次第という面があったが、過去の教訓と改善により近年では一定数の字がコンスタントに送り出されるようになりつつある。例えば現行の規則では各提案主体は一度に 1,000 字までしか提案できないようになった。この IRG の審議は決して会期中にのみなされるものではない。毎回の会議を踏まえて、各提案主体が次回会議までに分担して提出字の精査そのほかの業務を行うことが、IRG の最も重要な機能といえる。SAT は参加機関の一つとして、提出字の審査および自身の提出分への指摘の対応、議論への参加、およびそれに付随する諸業務という形でこのワークフローにかかわっているが、一つの集合が確定するまでに平均 2 ～ 3 年程度の審議が必要となり、さらにその前後の手続きを含めると、ある外字の提案から登録までは 5 年程度の長丁場を想定しなければならない。単純な時間だけではなく、議論への参加や調査に割り振るための労力も必要となる。

IRG に漢字を提案する際、明朝体で表された字形のみならず、出典画像、文字属性、異体字情報などを定められた形式で提出しなければならない。審議ではこれらに基づいて提出された漢字が確かに存在するか、原典と齟齬がないか、部首や画数が基準に従っているかなどを検査する。これらの規定は現在までに相当程度詳細に明文化されているが、細部は都度変更が加えられるため、内容については割愛する。審議の結果、真正性に疑義がある、ほかの既存字に包摂可能であるなどの問題が認められた場合は、新規登録の候補から外されるが、追加証拠の提出などによって回復することもでき、最終的に問題がないと認められた字の集合が上位の委員会に引き渡される。提案された文字も、必ずしも原提案の通りに承認されるのではない。証拠に疑義があっても却下される場合はもちろんであるが、例えば SAT の内部では別字と見なすに足る違いがあるとして提案しても、一般の情報通信目的では既存字と同等と判断され、新たな符号位置が与えられないことがある。この場合、SAT では主に前述の IVS を用いて区別を維持するよう再提案する方針である。

IRG では公式には提案主体らが業務の負担を最も多く引き受けることになっているが、実質的には漢文字符号化に関心のある一般参加者らも大きく貢献して

いる。特にプロセスの効率化に関しては、川幡^{かわばた}太一^{たいち}氏による漢字構造の自動照合による重複検出や、近年の Henry Chan 氏によるレビュー入力のウェブ化といった抜本的な改善はもっぱら個人参加者によってもたらされたものである。この点は筆者も IRG の公開データの再利用性の低さが審議の正確性に及ぼす悪影響に懸念を抱いており、IRG のリポジトリをデータベース化する取り組みを行っている。

5-3. 字形の利用

ここまで符号化は文字に番号を割り当てる作業であると説明してきたが、これは同時に文字の実体を図形的情報から切り離すことを意味する。その状態でも、データとしての機械可読性に基づくさまざまな恩恵を受けることには変わらないが、それだけでは人間がいわゆる「文字」として、すなわち目で見て利用できる状態であることを意味しない。符号化によって得られるのは符号位置、つまり仮想的な活字棚の特定の箱を予約するということにすぎず、それを使用するためには活字を入れる、つまり表示するための字形情報を与えなければならないからである。

コンピューターでは文字符号から字形を呼び出すためのフォントファイルが用いられるが、フォントは符号化によって自動的に発生するものではなく、誰かが別に供給しなければならない。字形の表示はユーザーにとって基本的な要求のため、ほとんどのパソコン用 OS は一般ユーザーが困らないだけのフォントを標準で搭載しているが、慧琳音義を出典とするようなまれな学術用漢字はこのような対応を期待しにくい。過去に作られた今昔文字鏡や GT 漢字といった大規模文字集合は、いずれも公式の専用フォントを配布していた。このように、符号化文字を「実用的」にするためにはフォントの作成が必要¹¹であるが、しばしば膨大な字数となる漢字フォントを一から作成するには多大なコストと労力が必要となり、普及の障碍となっていた。

幸いなことに、2007 年頃に現大東文化大学准教授の^{かみち こういち}上地宏一氏によって漢字字形自動生成システムを利用したウェブサービスである GlyphWiki¹²が開発され、簡便な操作で利用制約のない高品質な明朝体フォントが作成可能となった。SAT はこれを用いて UCS/Unicode 規格に掲載する例示字形を作成すると

ともに、既収録字を表示するための「SAT 大正蔵漢字フォント」^{*13}を提供している。

6. まとめ

これまで、符号化という作業が、片や紙の文献に現れる文字の解説と解釈に否応なく判断を迫る場面があること、もう片方ではデジタル技術が提供する環境に適合するための方法論の確立を求めることの一端を、慧琳音義における経験を通じて述べたつもりである。情報媒体の発達の歴史は、簡単にいえばより知識の「本質」以外を削ぎ落とし、「本質」を際立たせる手法を探る道のりにとらえられる。その意味では、デジタル媒体による情報表現は離散的であるがゆえに、表現しようとするその試み自体にモデル化の必要を内包するという点で、かつてないほどわれわれが「本質」をどうとらえているかを浮き彫りにするものであり、かつそれ自体に研究性が要求されるものといえよう。同時に、新しい技術で何かを実現しようとする時、単に技術の「成果」を採用することにとどまらず、技術を支える人々との協働という側面も大切であるというのも、お伝えしたいことの一つである。拙文がこの分野でデジタル化を考える方々にとって、わずかながら参考になれば幸いである。

注

- 1 高田時雄(1994)「可洪隨函錄と行珙隨函音疏」高田時雄(編)『中國語史の資料と方法』京都大学人文科学研究所:109-156.
- 2 永崎研宣(2016)「SAT 大蔵経データベースをめぐる漢字情報」高田智和・馬場基・横山詔一・石塚晴通(編)『字体と漢字情報』勉誠出版:265-280.
- 3 徐時儀(2012)「一切経音義三種校本合刊緒論」『一切経音義三種校本合刊』1. 上海:上海古籍出版社.
- 4 船山徹(2008)「漢語仏典——その初期の成立状況をめぐって」京都大学人文科学研究所附属漢字情報研究センター(編)『漢籍はおもしろい』研文出版:71-118.
- 5 歴史的経緯により国際規格である UCS とそれを補完する業界規格の Unicode になる。本稿では詳細に立ち入らず総称してこのように呼ぶこととする。
- 6 JIS X 0221 の訳語に従うが、直訳すると「中・日・韓(朝)語統一表意文字」。この用語は現在から見るとやや不正確なきらいがあるが、制定初期から存在する術語として使われ続けている。正確にはベトナム語 Vietnamese からの寄与(チュノム)

も大きい。これらを合わせた CJKV はもっぱら東アジア言語処理の代名詞的な呼称となっている。

7 UCS/Unicode に新規字の追加や変更を要望することをこう呼ぶ。

8 訳は「異体字セレクト」とも。

9 <https://unicode.org/ivd/>

10 SAT 提出字は執筆時現在、Unicode 委員会そのほかと識別名を共有しているが、取り扱いに不都合があることが判明したため、近々新しい識別名を与えられる見込みである。

11 実務的に言えば、UCS/Unicode に文字を登録する際に審査段階でフォントが必要とされるため、この文脈ではむしろ符号化を可能にするためにフォントの作成が求められることになる。

12 <https://glyphwiki.org/>

13 <http://21dzk.l.u-tokyo.ac.jp/SAT/unicode/satunicode.html>