

人文学資料としてのテキスト構造化の意義を再考する

大正新脩大藏經における脚注の解析と Linked Data 化をめぐる

永崎研宣

一般財団法人人文情報学研究所

三宅真紀

大阪大学言語文化研究科

苔米地等流

一般財団法人人文情報学研究所

A. Charles Muller

東京大学人文社会系研究科

下田正弘

東京大学人文社会系研究科

本稿は、大正新脩大藏經テキストデータベースにおける脚注の校勘情報を通じた人文学資料としてのテキスト構造化に関する研究を扱う。これはこのデータベースを構築・運用する SAT プロジェクトの次のフェーズの方針の策定に役立てることを目指すものである。大正新脩大藏經の校勘情報は、各一次資料の歴史的な関係からある程度想定可能だが、具体的に研究されてきたわけではなかった。ここでは、デジタル化された校勘情報を用いた分析を行ったが、結果は一般的な想定を裏付けするものであった。いくつかの例外的なものがあったことが、今後の方針を策定する上で有益かもしれない。また、Linked Data という形で校勘情報を扱うことで、一次資料公開に関する権利所有者との合意形成が容易になるかもしれない。

Reconsideration of the Significance of the Text Encoding

Kiyonori Nagasaki

International Institute
for Digital Humanities

Maki Miyake

Graduate School of
Language and Culture
Osaka University

Toru Tomabechi

International Institute
for Digital Humanities

A. Charles Muller

Graduate School of
Humanities and Sociology
University of Tokyo

Masahiro Shimoda

Graduate School of
Humanities and Sociology
University of Tokyo

This paper describes a preliminary survey of data of critical apparatus of the digitized version of the Taishō Shinshū Daizō kyō in order to assist a decision regarding the next phase of the SAT project. This project aims to enhance the scholarly reliability canonical collection of the scriptures. The relationship between witnesses referred in the Taishō Shinshū Daizōkyō had been assumed based on the historical process of composing of the witnesses, but it had not been thoroughly analyzed. In spite of various difficulties, the results of the analysis seem to be almost within the scope of the general assumption. In this case, several isolated examples might be useful for creating a policy. In the next phase, it would be useful for the project to adopt the idea of Linked Data in order to make consensus with contents holders. It is also needed to develop any new implementation for treating the data..

1. まえがき

本稿は、大正新脩大藏經¹⁾における校勘情報とそのデジタル化手法の研究に関する報告で

あり、それを通じて仏典をはじめとする東洋古典籍の校勘情報のデジタル化の可能性について検討するものである。

2. 大正新脩大蔵經データベースの成り立ち

大正新脩大蔵經は大正時代末期から昭和初期にかけて編纂された、仏典の叢書である。大蔵經、もしくは一切經を編纂することは、特に東アジアにおいては広く行われてきており、当初は書写によって伝えられていたものの、宋代の木版大蔵經の開版以降、木版での印刷が行われるようになり、広く流布するようになった。我国においても、17世紀に鉄眼禪師が開版した黄檗版大蔵經は広く流布し仏典の普及に大きく貢献したとされる。詳しくは後述するが、大正新脩大蔵經はこのような伝統の中で編纂された大蔵經の一つであり、当時の最新の研究成果が盛り込まれた活版の洋装本として、国内のみならず世界中の大学図書館に配架され、以来、中国語仏典のスタンダードとして仏教学のみならず仏典を扱う様々な分野において世界中で広く用いられてきたものである。デジタル化時代の到来に伴い、この大正新脩大蔵經の本文をすべてデジタル翻刻しようとするプロジェクトが始まったのが1994年のことであった。その後、情報技術の変遷にあわせ、2005年にはWebコラボレーションシステムが導入され²⁾、6億円ほどの費用と200人以上の主に研究者の手によって、85巻、約600万行、約1万種のいわゆる外字を含む1億字超のテキストのデジタル翻刻が完了したのが2007年のことであった。2008年には全文一括検索機能や電子佛教辞典との動的リンク機能を備えたWebサービスとしても公開され、その後、さらなる校正と、英訳大蔵經とのパラレル・コーパスや他のデータベースとの連携機能の追加等が行われた新版が2012年に公開され³⁴⁾、現在に至っている。以下では、この大規模仏典テキストのデジタル翻刻における校勘情報の扱いについて、いくつかの角度から検討してみたい。

3. 校勘情報とそのデジタル化手法

ここで言う「校勘情報」とは、校異、critical apparatus 等とも呼ばれるものであり、聖書や仏典をはじめとする様々なテキストに関する現存している複数の資料から、何らかの基準に基づいて一つのまとまりのあるテキスト(いわゆる校訂テキスト、あるいは学術編集版(scholarly edition)などと呼ばれる)を再構成する際に、照合した結果異なっている部分(異文、variant)について記述される注記のことである。多くの場合は、脚注という形で本文と同じペー

ジか周辺のページに記載され、その記述ルールは分野や時代、あるいは学派などによっても様々である。特に、「テキストの元々の姿を再構成すること」を目指したものは「校訂テキスト」と呼ばれる事が多い。有名な校訂テキストとしてしばしば挙げられるのは、新約聖書のNetsle-Alandの『Novum Testamentum Graece』である。これは、現存する様々な新約聖書の写本に基づいてギリシャ語の新訳聖書を再構成する試みであり、新発見写本の校勘情報を随時追記しつつ場合によっては本文の再検討も行った上で改版を重ねてきており、現在では、第28版となっている¹⁾。

仏典の場合にも、現存する資料には様々なものが存在しており、それらを照合することで、本来のテキストがどのようなものであったか、あるいはある時点でのテキストがどのようなものとして読まれていたか、といったことを解明するべく、校勘が行われてきている。インド発祥の宗教であるため、パーリ語やサンスクリット語と言ったインド系の言語によるテキストが基本となるが、必ずしも現存しているとは限らず、現存していたとしても点数が少なかったり不完全であったりする場合もあり、さらに、残っていたとしても書写に書写を重ねたことによる誤記もあり得る。その一方で、多くの仏典が中国とチベットにおいてそれぞれ翻訳されたことがあり、中国語・チベット語の「大蔵經」として各文化伝統の中で継承されてきたという経緯があり、なかにはインド系の言語の写本が未発見のテキストも少なからず含まれている。すなわち、翻訳時点でのテキストの内容を解釈・記録した上で、その後、異なる言語世界において伝承してきたということであり、それらのテキストもまた、仏教テキストの過去の姿を再構成していく上で欠かせない資料となっている。それゆえ、中国語やチベット語の翻訳經典の一節もまた、しばしば校勘情報の一部として脚注に記載されることとなっているのである。

宗教テキストだけでなく、複数の版が流通し現存している文学作品や哲学書は、そのような形で校勘情報を含めて編集され、出版されることが学術的手続きとして一般的となっており、

¹⁾ 『Novum Testamentum Graece』第28版については以下のURLを参照されたい。
<http://www.nestle-aland.com/en/the-28-edition/>

これが学術編集版と呼ばれるものである。

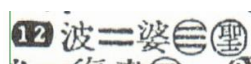
学術編集版をデジタル化しようとする試みにはすでに 20 年以上の歴史があり、その試行錯誤の成果の多くは TEI (Text Encoding Initiative) ガイドラインに反映されている。なかでも、校勘情報に関しては深い議論が展開されてきており、近年までの成果はこのガイドラインの第 12 章 Critical Apparatus¹において詳述されている。校勘情報のデジタル化の手法として現在ガイドラインに採用されているのは三つであり、校勘情報の正確な対応箇所を機械処理することは難しいが紙媒体での記述方法との親和性が高く記述がしやすい Location-referenced Method、任意の位置にアンカーエレメントとして始点/終点を入れた上で XML ID を付与してそれを他の箇所から参照することで機械処理での効率的利用を可能としたがマニュアルでの入力作業がやや煩雑な Double End-Point Attachment Method、本文テキストの校勘情報の対応箇所に直接校勘情報を埋め込むことで機械処理の可能性を高めるとともに入力作業の工数を減らした Parallel Segmentation Method、が現在用いられている代表的な手法である²。さらに、Stand-off マークアップに基づくものなど、いくつかの別の手法も提案され議論されているところであり、いずれはそれらもガイドラインに取り込まれるかもしれない。ただ、いずれにしても、XML タグを人が直接手入力する作業を前提とするべきではないという空気が醸成されつつあり、何らかの方法で入力をするソフトウェア側で自動的に上記のいずれかの手法に準拠する形で記述されるか、あるいは、最終的にそのようにして出力されるようにするという流れになりつつあり、各地で専用・汎用 TEI エディタの開発が進められつつある³。近年新たに提案されている校勘情報記述手法はいずれも、マニュアルでのタグ付けを原則としないという前提をとることにより、コンピュータに処理しやすい記述手法と人にとって操作しやすいイン

ターフェイスを両立しようとするものである。

また、特に Parallel Segmentation Method をターゲットにした学術編集版閲覧用ソフトウェアとして有名なものにメリーランド大学にて開発・公開されている Versioning Machine がある³。これは校勘情報における各版のテキスト断片から、各版を再構成し、Web ブラウザ上で並置して読めるようにするというものであり、校勘情報をデジタル化する意義を目の当たりにさせてくれるものである。

4. 大正新脩大藏經の校勘情報とデジタル化

では、同じく学術編集版である大正新脩大藏經の校勘情報のデジタル化はどのように行われているのか。このデジタル化は 90 年代半ばに計画されたものであり、校勘情報の記述方法については TEI や XML を念頭に置かずに設計されたものである。もちろん、現状でも Location-referenced Method であれば比較的短期間でほぼ自動的に TEI/XML 形式文書の生成が可能だが、次のフェーズとしては、機械処理しやすい形式での作成・提供を念頭に置いているため、現在のところは既存の手法のままである。典型的には、脚注としての位置情報を付与された上で、以下のような形で記述されている。

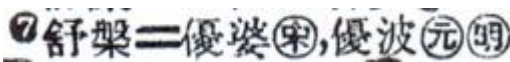


(大正新脩大藏經脚注での校勘情報の例 1)

<#12/>波=婆<三><聖>

(上記の校勘情報のデジタル化の例 1)

ここでは、大正新脩大藏經当該頁の本文における脚注 12 が付されたテキスト「波」に対して校勘対象として照合された三つの木版大藏經「宋本・元本・明本」(<三>) (詳しくは後述する) と正倉院聖語藏 (<聖>) の同一箇所ではいずれも「婆」となっているということを意味している。もう一つ例を見てみると、



(大正新脩大藏經脚注での校勘情報の例 2)

<#7/>舒槃=優婆<宋>, 優波<元><明>

(上記の校勘情報のデジタル化の例 2)

ここでは、本文中の脚注 7 が付された「舒槃」

¹ TEI P5 ガイドラインの第 12 章については以下の URL を参照されたい。

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>

² これらの詳細については、以下の Web ページにも詳しい。TEI by Example:

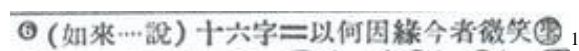
<http://teibyexample.org/modules/TBED07v00.htm?target=appConnect>

³ Versioning Machine: <http://v-machine.org/>

が宋本では「優婆」、元本と明本では「優波」となっていたということになる。なお、この場合、聖語蔵への言及はないため、聖語蔵は底本と同じ、ということになる。

以上の2つの校勘情報については、正確にデジタル化できていれば TEI における3つのメソッドのいずれについても機械的に変換が可能はずである。しかし、原稿時点での誤植が存在する上にデジタル翻刻時の誤植もあり得るため、完璧な処理というわけにはいかないだろう。

こういったものとは別に、以下のように字数を指定するものもある。



(大正新脩大藏經脚注での校勘情報の例3)

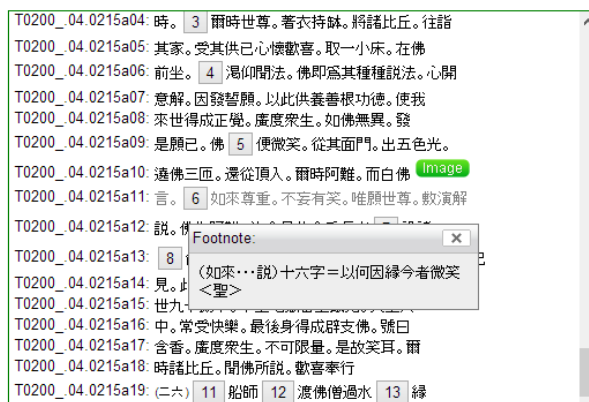
<#6/> (如來…説) 十六字=以何因縁今者微笑 <聖>

(上記の校勘情報のデジタル化の例3)

この意味するところは、脚注6が付された「如來」という文字列に始まり「説」で終わる16文字の箇所が聖語蔵では等号の右辺の文字列になっている、ということである。この書式自体は大正新脩大藏經の校合規則において規定されているが⁷、「字」がテキストそのものかメタ情報としての「字」なのかという区別をはじめ、読者の判断に委ねられる部分があるため、機械処理はやや困難な状況にある。この種の問題が解決できたなら、校勘情報に限って言えば、少なくとも Location-referenced Method に従って機械的に変換することは可能だろう。

大正新脩大藏經の校勘情報には、他にもいくつかの規則があり、上記の例3と同様、機械処理にはやや難しいものが散見される。データ量としても、おおよそ600万行、全体で1億字、脚注数約75万という状況にあっては、当初の計画そのものを変更することは難しく、また、紙媒体ですでに書かれたものをデジタル化するにあたって Double End-Point Attachment Method や Parallel Segmentation Method のような形でマークアップしていくことは作業量的に難しいということもあり、現状に至っている。ちなみに、Web サービス上ではこれは以下のよ

うな形で、脚注番号をクリックするとポップアップ表示されるようになっており、公開当初よりユーザからは好評である。



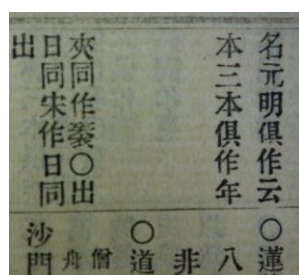
また、ここに端的に見られるように、学術編集版の校勘情報を分析するには、それがどのような出自の資料を用いてどのように遂行されたのかという状況についての理解が欠かせない。そこで、次に、この編纂と出版の事情についてみていこう。

5. 大正新脩大藏經の校勘の歴史的経緯

大正新脩大藏經の校勘の歴史的経緯と問題点についてはすでにいくつかの研究があるので詳しくは参考文献⁸⁾⁹⁾に譲るとして、ここでは簡単に触れておきたい。先述のように、我国においては鉄眼禅師が開版した木版大藏經が、大藏經としては特に流布していた。これは明代に開版され輸入された嘉興蔵と呼ばれる大藏經に基づくものであったが、その後、明治時代に入り活版印刷が広まる中で、活版を用いた携帯性のある大藏經を出版しようという機運が持ち上がる。その結果、篤志家の協力の下、約400巻に上る縮刷版の和装本、『大日本校訂大藏經』(縮刷蔵とも呼ばれる)が実現したのである。ここで照合作業に用いられたのは、増上寺に所蔵されていた木版大藏經である、高麗版大藏經、宋本(宋代に開版された思溪蔵)、元本(元代に開版された普寧蔵)であり、これに加えて、明代の大藏經として、上述の黄檗版が用いられた。増上寺所蔵の高麗版は一度戦火で版木を失った後、13世紀に再度開版された再彫本と呼ばれるものだが、10世紀宋代に蜀の成都にて開版された『開宝蔵』の復刻版として11世紀に高麗にて作られた初彫本をただ引き写しただけでなく、唐の長安の中原写経を淵源

¹ 大正新脩大藏經テキストデータベース
http://21dzk.l.u-tokyo.ac.jp/SAT2012/T0200_04.0215a11:0215a11.cit

とするとされる『契丹藏』をも照合して編纂されたものであり、伝統的にはもっとも良い大藏經であるとされてきたものである。これを底本としつつ、他の三版を照合して校勘を行っていたのであるが、実際の作業としては、一人が高麗本を読み上げ、他の三人がそれぞれの版を確認して必要に応じて校勘情報の注記を行っていたということである。この時は、校勘情報は以下の図のように頭注として組版された。右側の頭注のみ説明すると、「本文中の「名」は元本と明本では共に「云」となっており、「本」は宋本・元本・明本の三本共に「年」となっている」と記載されている。これをそのままデジタル化したとしても機械処理はかなり困難だろう。



（『大日本校訂大藏經』の校勘情報）

その後、仏教学の発展とともに聖語藏や敦煌写本の内実が明らかになるにつれ、新しい知見を反映させた新しい校訂大藏經の出版の機運が高まり、着手されるに至ったのが大正新脩大藏經である。高楠順次郎・渡辺海旭を都監に、仏教学者の力を結集して行われたこの出版事業は、費用の工面や印刷工の手配をはじめ様々な苦労があったとのことだが¹、まずは、縮刷版ゆえに活字が小さく読みにくかった大日本校訂大藏經を少し大きな活字で印刷し直したものをベースとして作業が進められていった。それゆえ、基本的に、大日本校訂大藏經の校勘情報が引き継がれることも少なくなかったようである。また、何よりも、1000 頁の本をほぼ月一冊のペースで刊行していき、活字も 2000 面、すなわち、当月の分を校正する間に翌月の分を組み上げていくという慌ただしい流れの中、芝の増上寺では国宝の三種の漢訳大藏經及び明代大藏經の照合、上野の帝室博物館で東京に移送された正倉院聖語藏との校合、宮内省図書寮にて宋本一切經の校合という地理的にも

決して便利とは言えない状況で、さらに、印刷工の賃金を含む印刷関連の諸費用をはじめとする様々な経費のための金策に困り果て、ついには最初の会社を畳んで新会社を設立して債務整理を行い、高利貸しに紙型を差し押さえられないよう身を以て拒むといったことを行いながらの作業であり、膨大な資料と向き合うという本来の仕事だけでも大変であったにも関わらず、事業の全体としても困難を極めたようである。そのようなことから、誤植の混入は無理からぬ事であり、実際のところ、紙型が失われた後の 1960 年の再刊にあたっては、大規模な字句の修正が行われたということである。それでも誤植を少なからず含んでいることについての指摘が近年においてもなされている¹⁰⁾。しかしながら、底本及び主な参照資料となった高麗藏・宋本・元本の各木版大藏經については、重要文化財であり、容易にアクセスできるものではなく、明版については、大正新脩大藏經の略号表では「Ming Edition, A. D. 1601」となっていることから万暦版大藏經（嘉興藏）であると想定されるが、黄檗版ではなかったかとみる向きもあり、実際にどれを参照したのかは確認が必要という状況である。このようなことから、大正新脩大藏經の校勘情報の正確性について検証することは、やや難しい状況であると言わざるを得ない。

その一方で、近年、奈良・平安時代から我国に残されてきた古写經についての調査がさらに進展してきており、特に、金剛寺一切經については、すでにデジタル画像として国際仏教学大学院大学の図書館で閲覧できるようになっている。古写經の対照目録も同大学の古写經研究プロジェクトによって整備されてきている¹¹⁾。大藏經テキストデータベースの今後の発展を考慮するなら、これらの古写經についても校勘を行い記述していくことが必要になってくるだろう。上述のような事情により、既存の校勘情報の検証ももちろん必要であるが、同時に、新たな研究成果に基づく校勘情報を追加していくという仕事もまた、大正新脩大藏經テキストデータベースを学術編集版としてより精度の高いものにしていくためには避けては通れない。そして、そこでは適切な計画と着実な実行が必要となる。これにあたって、既存の校勘情報について事前に分析しておくことが有益であると考えられることから、以下で、デジタル化された大正新脩大藏經の校勘情報の分析を試みた。

¹ これに関する詳細は、以下の URL を参照されたい。 <http://togetter.com/li/521547>

6. 大正新脩大藏經における校勘情報の調査

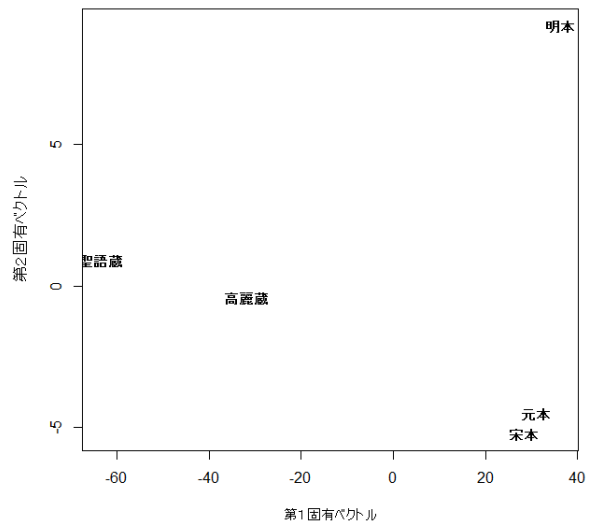
ここまで述べてきたように、大正新脩大藏經の校勘情報は改めての調査が必要とされているが、すでに現在の校勘情報はデジタル化されている。そこで、ここでは、再調査の指針を立てる上で、この校勘情報のデータの分析結果を利用することを検討してみたい。

脚注における校勘情報はすでに上述のような形式でデジタル化されているため、「現在の大正新脩大藏經において校勘情報として記載されている各版同士の相違が何文字ずつあるか」というデータ（以下、これを編集距離と呼ぶ）を取り出すことができるように思われる。しかし、実際のところ、大正新脩大藏經そのものの記述方式が機械処理という観点からはあいまいな面があることや誤植があること、それに加えて、デジタル翻刻に際しての誤植があり得ることを考慮に入れると、大正新脩大藏經において意図された校勘情報を正確に反映した編集距離を導き出すことは現時点では困難であると言えるが、それに近い数字を取り出すことはできるだろう。

また、大正新脩大藏經に収録されたテキストには様々なものがあり、すべてが一様に同じ一次資料群に基づくものではない。宋・元・明の各木版大藏經を参照しているものは多いが、それらに加えてさらに正倉院聖語藏を校勘対象としたテキストとなると、現在の脚注データを確認した限りでは、全体で190点であった。そこで、まずはこの190点全体での編集距離を計測し、Rで多次元尺度法を用いてプロットしてみたのが以下の図1である。ここでは、相違・増加・欠如をすべて1とした上で、最大値を100とした相対値を用いている。

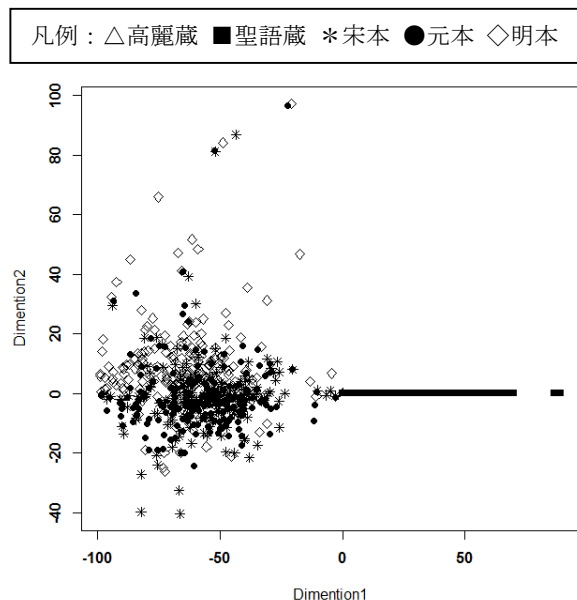
全体としては、高麗藏はどれからもそれなりに離れており、聖語藏が他の諸版から離れたところにあること、加えて、縦軸は横軸に比べて目盛の数値差が小さいことから、宋本と元本との違いは極めて小さく、明本も図表上は離れているが実際にはそれほどかけ離れていないということがわかる。この図からは、木版大藏經に関してはその成り立ちを大体反映していると言ってよいのではないかとと思われるが、一方で、聖語藏の位置については、各版のすべてから離れているという点は理解し得るものの、その成り立ちを考慮するなら、高麗藏や宋・元・明本のグループの距離よりももっと離れてい

てもおかしくないことから、やはり聖語藏の校勘情報については再照合の余地があるのではないかと考えられる。



(図1: 多次元尺度法を用いて計測した各版の関係)

また、実際にテキスト毎にどのようなになっているかということについて確認することで、大正新脩大藏經の校勘情報の特徴や今後の再照合の指針についてのより踏み込んだ可能性が出てくるかもしれないので、それも確認してみよう。

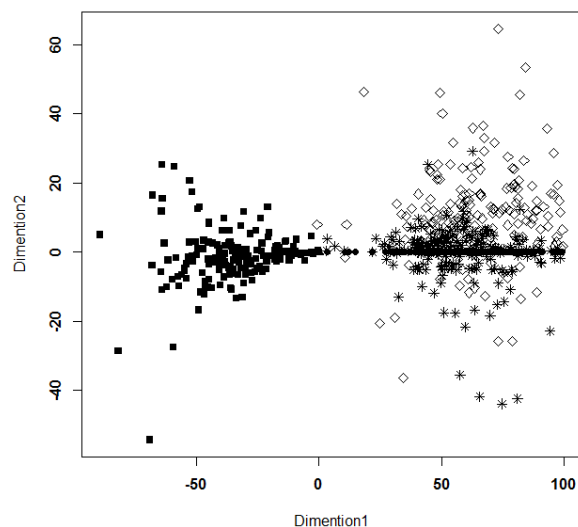


(図2: 多次元尺度法を用いて計測したテキスト毎の各版の関係1)

図2では、190点のテキストのそれぞれについて編集距離を測って相対値をとった上で、高麗藏を中心点としつつ高麗藏と聖語藏を正の

方向で同一線上に揃えつつプロットしてきている。

この図2を、高麗蔵を中心点としつつ高麗蔵と元本を正の方向で同一線上に揃えてプロットした図3と比較すると、その特徴が一層明らかになる。すなわち、聖語蔵に関しては、どのテキストにおいても、高麗蔵が最も近く、宋・元・明本はそれよりも遠いということになる。



(図3：多次元尺度法を用いて計測したテキスト毎の各版の関係2)

このことは、他の木版大藏經との関係を考慮した時、大変興味深いものとなる。すなわち、高麗蔵及び宋・元・明本は、いずれも10世紀、蜀において開版された開宝蔵と呼ばれる大藏經から派生したものとされているが、元本、明本はいずれもこの宋本を継承するものである一方、高麗蔵は同じく開宝蔵を継承するものではあるものの、聖語蔵と系統が近い可能性があることとされる唐の長安で開版された契丹蔵を参照しつつ編纂されたとされている¹²⁾。したがって、各版の出自を考えるなら、聖語蔵が宋・元・明本と比べて高麗蔵と近い関係にあるということは十分に考えられることであり、図2がそのことを端的に示しているということが可能かもしれない。

しかしながら、これはあくまでも現段階での作業仮説であることに注意しなければならない。上述のように、大正新脩大藏經の校勘情報そのものに対して問題が指摘されている上に、それを必ずしも正確にデジタル化できたとは限らず、さらに、機械処理においてあいまいとなる要素を完全には排除できていないために、今回の試行においては、大正新脩大藏經校勘者の意図に正しく沿った形で計測できたかどうか

かが定かではない。また、上述のように、作業工程上、聖語蔵と直接対比されたのは高麗蔵を底本とする縮刷蔵の拡大印刷版のみであり、宋・元・明本と直接対比されなかった可能性が高いということもこのことに影響している可能性もある。これを踏まえた時、やはり、近年各方面から指摘されるように、大正新脩大藏經と各版との再照合を避けることは困難だろう。

また、図3において特に差異が大きいことが確認できるものとしては、『出生無邊門陀羅尼經』(大正 No. 1018)『過去莊嚴劫千佛名經』(大正 No. 446)『開元釋教錄』(大正 No. 2154)『顯揚聖教論頌』(大正 No. 1603)『東方最勝燈王如來經』(大正 No. 1354)が挙げられる。大正新脩大藏經の各版との再照合は膨大な作業量となるため、担当者の業績や動機付けから作業進捗の具合に至るまで、作業に取りかかるにあたっては様々な配慮と入念な準備が必要となる。既存の諸研究を踏まえることは当然だが、それに加えて、全体の再照合の方針を策定する前に、予備調査として、まずこれらのテキストを調査してみることで、より効果的な作業方針を策定することに貢献できる可能性があり、今後、まずはこれらを手がかりとして検討を進めてみたい。

7. 今後の展開と Linked Data の可能性

すでに大正新脩大藏經校訂の頃には我国に現存する古写經の存在は注目されつつあったが、近年の飛躍的な研究の進展により、上述のように、さらに古写經が注目を浴びようになっている。実際に、金剛寺一切經写本については国際仏教学大学院大学の図書館にて、聖語蔵は CD-ROM・DVD-ROM でのデジタル画像の刊行が進められつつある。これらを実際に照合・再照合していくことで、確認できる部分は多いだろう。この作業に取りかかる際にも、この大正新脩大藏經の脚注における校勘情報の分析は作業の方針を立てる上で欠かせない有益なものとなるだろう。

また、校勘情報を機械処理に適した形式で正確に記述していくことは、Versioning Machine において実現されているように、デジタルテキストの場合には一次資料を丸ごと簡単に再構成できてしまうことでもある。すべての一次資料を全文入力することに比べればかなりの省力化が見込めるが、一方で、入力形式の統一や適切なアプリケーションの開発が必要である。

また、校勘情報の扱いに際しての Versioning Machine の現状の問題点として、critical apparatus の後に必ず空白が置かれるという実装になっていることが挙げられる。日本語や中国語で本格的に用いる場合には修正が必要なのは確かだが、ここで興味深いのは、言語文化圏によるテキストの様態の相違がデジタル技術の活用を通じて意図しないままに視覚化され明らかになってしまっている点である。小さなことだが、しかし、このことから、今後一層展開していくであろうデジタル技術を通じたグローバルな協働が人文学に与え得るインパクトの大きさを予感せざるを得ない。

このように、学術編集版において欠かせない校勘情報の構造化は、様々な意味と可能性を持っており、TEI ガイドラインをはじめとする国際的なデジタル化の枠組みを通じてこれをグローバルに捉え直していくことは、人文情報学/デジタル・ヒューマニティーズにとってのみならず、人文学そのものにとっても今後さらに深い意義を有していくことになるだろう。

最後に、Linked Data として校勘情報を公開することについて検討してみたい。すでに TEI Conference 2013 のテーマが Linked TEI であったように、人文学デジタル化の国際的な潮流も、徐々に Linked Data の方向に進みつつある。校勘情報に関する脚注を Linked Data 化としたなら、少なくとも二つの可能性が考えられる。一つは、二つの異なる版のテキストをつなぐ情報として Linked Data を作成し公開するものである。もう一つは、校勘情報そのものを Linked Data として記述し公開するものである。

前者に関しては、まだ公開はしていないものの、BDK-SAT パラレル・コーパス（英訳大蔵経と大正新脩大蔵経とのパラレル・コーパス）のデータが Linked Data の形式で出力可能となっている。この場合には、リンクされる2つのテキストが両方ともに公開されて URI で部分指定可能となっている必要があり、複数の別々のプロジェクトでそれぞれにテキストを公開している場合には特に有益な手法だろう。

後者に関しては、権利関係の問題でテキスト全体を公開できず、それでも校勘情報のみが公開可能であるという場合に有効だろう。もちろん、完全に正確なデータであればベースとなったテキストと組み合わせることでそのテキストを再構成できてしまうため、その点も含めた上での合意形成が必要だが、全文テキストの公開に比べると多少敷居が低くなる可能性がある。

り、そこに大きな期待をかけている。

いずれにしても、校勘情報の Linked Data 化は、それを有益に扱えるアプリケーションやサービスまで視野に入れて進めていかねばならず、じっくり腰を据えて取り組んでいきたい。

謝辞 本研究は多くの方々のご協力とご支援の成果としてのデジタルデータを活用しており、深く感謝する。なお、本研究の一部は JSPS 科研費 22242002 の助成を受けたものである。

参考文献

- 1) 高楠順次郎・渡辺海旭編: 大正新脩大蔵経, 大正新脩大蔵経刊行會(1924).
- 2) 永崎研宣, 鈴木隆泰, 下田正弘: 大正新脩大蔵経テキストデータベース構築のためのコラボレーションシステムの開発, 情処研報 CH-70(2006年5月), pp. 33-40.
- 3) Kiyonori Nagasaki, Toru Tomabechi and Masahiro Shimoda: Towards a Digital Research Environment for Buddhist Studies, Literary and Linguistic Computing, Oxford University Press, (2013), pp. 296-300.
- 4) Kiyonori Nagasaki, Toru Tomabechi, A. Charles Muller, and Masahiro Shimoda: A Case Study of Integration of Services and Resources on a Web Service, Digital Humanities 2013, pp. 517-519.
- 5) Geoffrey Rockwell, Susan Brown, et al., CWRC-Writer: An In-Browser XML Editor, Digital Humanities 2012, pp. 508-511.
- 6) Marco Petris, et al.: Crowdsourcing meaning: a hands-on introduction to CLÉA, the Collaborative Literature Exploration and Annotation Environment, Digital Humanities 2013, pp. 24-25.
- 7) 山崎清華: 異字の撰擇に就いて, 現代佛教 1928年11月号, pp. 103-115.
- 8) 梶浦晋: 近代における大蔵経の編纂, 常照, 51, 佛教大学図書館, pp. 10-19.
- 9) 永崎研宣: 大蔵経の歴史と現在, 新アジア仏教史 15, 佼成出版社 (2011年3月) pp. 15-53.
- 10) 船山徹: 漢語仏典—その初期の成立状況をめぐって, 漢籍はおもしろい, 研文出版 (2008年3月) pp. 71-118.
- 11) 林寺正俊: 日本古写経データベースの構築とその意義, 人文科学とコンピュータシンポジウム論文集(2012年11月), pp. 11-16.
- 12) 竺沙雅章: 宋元佛教文化史研究, 汲古書院 (2000年8月).