

正倉院文書のXML化による構造化と復原の検討

後藤 真[†], 高山 典史[†], 柴山 守[‡]

大阪市立大学 大学院文学研究科[†]

大阪市立大学 学術情報総合センター[‡]

正倉院文書は、東大寺の正倉院に伝来した8世紀の文書群の総称である。本文書は、背面再利用と19世紀初頭からの「整理」作業のため、その奈良時代の帳簿の形態より著しく損なわれ、論理構造と物理構造の差異という特徴をもつ。この奈良時代の帳簿形態の復原作業をコンピュータで支援できれば文書研究に貢献する。筆者らは、XML(eXtensible Markup Language)を用いて、奈良時代の帳簿形態を復原する支援ツールを開発し、構造化ルールに関する検討を行った。支援ツールの利用によって、関連する複数資料に基づいて構築された正倉院文書データベースは、文書構造をもつXML表現に変換される。このXML文書は、論理構造を復原するXSL(eXtensible Stylesheet Language)に基づいてWebブラウザ上に表示され、またWordマクロの定義により、物理構造が反映された「短冊」として容易に復原することができる。また、RDB形式で表現される正倉院文書データベースからXMLでの構造化を行う構造化ルール定義について検討している。

Structuring and Restoration of “Syoso-in Monjyo” using the XML

Makoto GOTO Norifumi TAKAYAMA Mamoru SHIBAYAMA

Graduate School of Literature, Osaka City University

Media Center, Osaka City University

The “Syoso-in Monjyo” is a generic name of a set of documents that were transcribed in the Syoso-in of Todaiji Temple in the 8th century. The “Syoso-in Monjyo” has been remarkably deformed and ruined due to their fact that transcriptions were imprinted on either side of the paper and also “Rearrangement” in the 19th Century. The “Syoso-in Monjyo” exhibits an unique dissimilarity between their logical layout and physical structure. This research attempts to apply some computer-assisted restoration to the “Syoso-in Monjyo” and develops techniques that could contribute towards a better understanding and preservation of similar historical documents.

The authors have developed a support tool that restores the logical form of the documents using the XML(eXtensible Markup Language) and also examined procedures for formulating a rule-based logic. The “Syoso-in Monjyo” structure is based on relating plural material that can be converted into XML expressions using the structured support tool. This XML document can be displayed on a Web browser using the XSL(eXtensible Stylesheet Language) which restores the original logical form of the document and the physical structure is restored by defining a suitable MS-Word macro as “Tanzaku”. Further, various aspects concerning rule-based definition that helps in structuring a database and subsequent conversion to XML is also evaluated.

1. はじめに

本報告は日本古代史における貴重な一次史料である正倉院文書をXML(eXtensible Markup Language)によって構造化し、研究・分析をよりいっそう深化させる事を目的としたものである。正倉院文書は、奈良時代の帳簿の形態より著しく損なわれている。そのため、正倉院文書データベースを構築し、復原の支援ツールを作成する。また、正倉院文書の背面再利用と19世紀初頭からの「整理」作業から、後述する論理構造と物理構造の差異という特徴をもつ。これをXMLによって構造化し、正倉院文書研究に新しい知見を得ることを目標とする。本報告ではその展望にたち、XML化と支援ツールの実現について、そのねらいと進展状況を報告するものである。

歴史資料としても、特殊な形態をもつ正倉院文書の構造と表記を、XMLによってWeb上や個人環境のコンピュータの上で表現することを可能とし、XML生成の構造化ルールを作成することによって正倉院文書研究のモデル化を図る。また、復原作業の効率化を目指す。

2. 正倉院文書

2-1. 正倉院文書とその研究の概要

正倉院文書とは、東大寺の正倉院に伝来した文書群の総称である〔1〕。宮内庁正倉院事務所管理の正倉院宝物のうち、広義には正倉院に納められていた文書すべてをさし、北倉文書17巻、東大寺献納図書10点、正倉院古文書667巻と5冊、宝物付属文書、木簡（題籤）がある。狭義には正倉院中倉の櫃（ひつ）に納められていた8世紀の写経・造宮関係文書（帳簿）群（正倉院古文書）を指し、一般的には正倉院古文書667巻5冊と木簡の一部のことを呼称する。狭義の正倉院文書は東大寺写経所とよばれる古代の一官司の事務管理のために作成された帳簿、写経所と造東大寺司および他官司・寺院・貴族諸家との往復文書で、律令制下の律令官司の活動実態を伝える貴重な記録である。また、その料紙に大量の廃棄公文書が再利用されており、なかでも各国から中央政府に提出された戸籍・計帳・正税帳などは律令支配

の実態、奈良時代の村落・社会構造、経済状況を解明するための基礎史料である。（本論文では狭義の意で「正倉院文書」の語をもちいる事とする）一般的には、再利用される前の廃棄文書（公文書を含む）を一次面、再利用によって書き込まれた文書（帳簿）面を二次面と呼ぶ。天保年間（1833～1836）に穂井田忠友が一次面に注目して主として公文書を抜き取り、成巻文書を編成して以来（のちに正集と称される）、明治に統修・統修後集・統修別集・塵芥（じんかい）・続々修に編成された。その結果、写経所文書群の形態は変容した。（正集45巻、統修50巻、統修後集43巻、統修別集50巻、続々修440巻2冊、塵芥39巻3冊）東大史料編纂所『大日本古文書』1巻～25巻（1901～1940）に翻刻が、宮内庁正倉院事務所『正倉院古文書影印集成』（1988～）に写真版が、東大史料編纂所『正倉院文書目録』（1988～）に目録が掲載されている。また、原文書はマイクロフィルムで見るとも可能である。庫外流出文書は国立歴史民俗博物館『正倉院文書拾遺』（1992）に集められている。また松嶋順正氏編『正倉院宝物銘文集』（1978）に題籤（だいせん）が収められている。

1950年代より正倉院文書研究は著しく増加した。初期における、正倉院文書研究の主眼はもっぱら一次面に向けられたもので、この公文書類が日本古代史研究にはたした役割は非常に巨大なものであった。1980年代に入って、正倉院文書研究は二次面の写経所文書にかかわる研究へと転換が起こる〔2〕。正倉院古文書に関わる詳細な目録、写真などが、出版されはじめた。そして、「新しい研究動向が始まって約十五年を経過した時点で、われわれは何をすべきか。まずどうしても行わなければならない基礎作業は、関係史料の「復原」である」として、まずは二次面の復原が肝要である、との提案がなされている〔3〕。われわれはこの視点にたち、復原の支援ツールを作成することを一つの主眼とする。

また、個々の写経事業の分析に基づく、写経所帳簿研究も必要であり、この研究から古代帳簿論

を構築するべきであるとの提案もされている。この写経所帳簿を構造化することによって、その特徴を見出し、古代帳簿論の構築に関わる新知見を

求めることが可能であるとおもわれる。

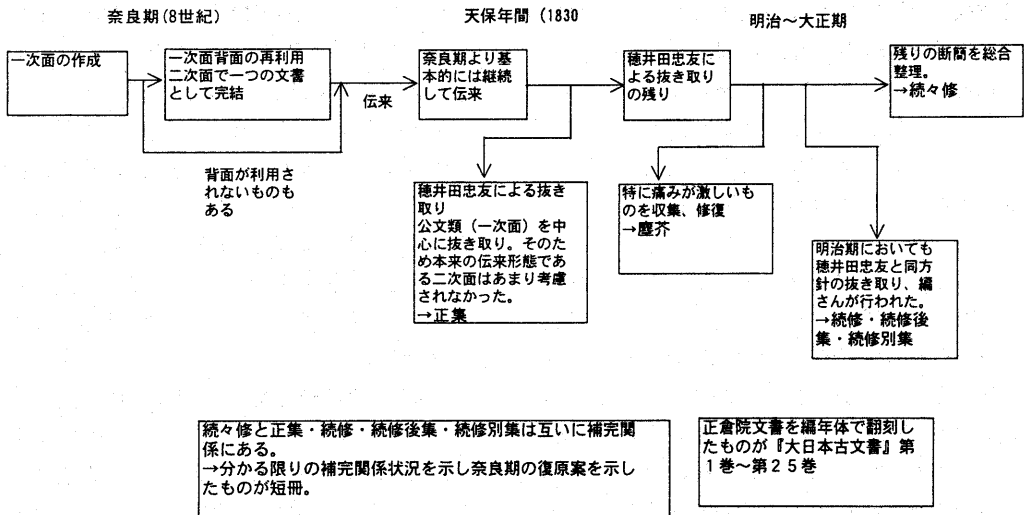


図1 正倉院文書伝来の過程

14/275	14/284~285	14/287		
首飾職所録 …日五廿月二十廿二	…月七廿三 …天 …百二十二册合	…百十廿三 …所 …舎人出勤 …右依太師去 …伝所請如件	…日五廿四 …次	可奉写一切経…
東寺奉写経所 …天平宝字…	東寺奉写経 返抄	東寺奉写経 返抄	東寺下任安都	
14/369	14/367~368	14/367	14/366~367	14/365~366
5	4	3	2	続々修18-61

図2 「短冊」による表裏関係図

2-2. 正倉院文書の「復原」

そもそも、二次面の復原はなぜ行われなければならないのか。天保年間に穂井田忠友の「整理」、明治年間にはいつの「整理」が一次面に注目して抜き取りをしたことは、前述したが、正集、続修などでは基本的に、優品の集成が行われたため、8世紀以来、伝来した文書・帳簿のはりつぎや、括りが失われた。また、一次面を意識した「整理」であるため、料紙の表裏の関係が混乱した部分もある(図1)。その失われた形態を復原する必要がある。

それでは、正倉院文書の「復原」(以下では原則として二次面の復原を第一義とする)はいかにして行われるのか。基本的には「短冊」とよばれる特殊な表を用いる(図2)。「短冊」は一次面、二次面の表裏関係、失われた伝来時形態の復原案としての接続関係、料紙の状況、翻刻(『大日本古文書』)の該当部分を一見して確認できるようにしたものである。

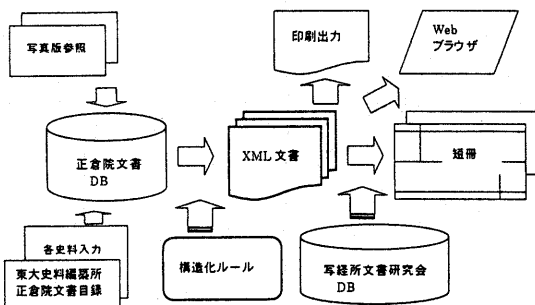


図3 XML化による復原支援

この「短冊」の作成には、先述のマイクロフィルム、『正倉院文書目録』、『正倉院古文書影印集成』などの諸情報もちいる。また、翻刻本である『大日本古文書』の基本方針は、編年形態をとっているため、成巻文書と状況が異なる。また、一部では復原が試みられているものの誤りもある。したがって『大日本古文書』との対応関係も一見して分かる必要がある。

3. XML化による正倉院文書復原の検討

正倉院文書にはXMLによって表現しやすい部分と、それが、困難な部分とがある。正倉院文書は大きくは正倉院文書、成巻単位、紙、行と分類され、これらが入れ子構造になっている。また、その一方で、『大日本古文書』の巻ページとのずれ、また論理的な文書構造とのずれがみられるため(図2)、その表現をすることは一つの課題となる(後述)。短冊作成においては初行、終行を表現し料紙の内部での文字の位置関係を表記すること、第x紙目に続々修xx-xが来るといった状況を示すこと、その接続関係を明確にすること、大日本古文書との位置関係などが一見できるようにしなければならない。

4. XMLによる研究支援ツールの開発(4)

この短冊作成、論理文書の表記を同時に実現させ、また、歴史研究者が、訓練の必要なXMLタグを用いることのない支援ツールとして(図3)のようなものを構想した。歴史研究者は、ある一定のルールにおいて各史料入力を行って、正倉院文書データベースを作成する。このデータベースは、リレーショナル形式である。この後、RDB形式からXML化のための「構造」を持たせる変換が必要になる。これを構造化ルールと呼称する。正倉院文書データベースは、知識ベースである構造化ルールをもとにXMLを生成する。その結果を図4にしめした。史料入力の際には参照データとして、画像データである写真版、マイクロフィルム、また先述の『正倉院文書目録』を使用する。構造化ルールに関しては、現在は検討中であるが、一定の法則性のもとに、展開することは可能であると思われる。作成されたXML文書において、基本的にはタグ<正倉院文書>以下、行単位に至るまで、物理的なタグにおいて構成されるが、その一方で論理IDを用いることによって、物理構造である紙単位をこえた、論理構造の表現を行うこととした。論理表現としてはWebブラウザでの表現とし、物理表現としては短冊での表記を目指すものとした。まずは論理単位の表現から説明したい。

```

<正倉院文書>
<成巻><成巻番号>続々修 18-6</成巻番号>
<成巻本文>
<成巻題目>御願奉写等雜文案</成巻題目>
<紙>
<大日本古文書本文><大日本古文書巻頁>14-365</大日本古文書巻頁>
<本文><論理文書 id="0001-1">
<line_s>可奉写一切經<del>律并</del><ins>一部經律并</ins>三千四百卅四<ins>三</ins><a>巻</a><wari>之
中卅九</wari><ins>八</ins><wari>巻今所</wari></line_s><br />
<line>合可用紙六万七千<del>八百四</del><ins>九百三十五</ins>張</line><br />
<line> 六万<del>三千九百五十一張</del><ins><del>四千八十二張</del><ins>四千六百六十六張
</ins></ins>経紙</line>
<line> 五万九千<del>七百卅九</del><ins>九百<gi set="mojikyo" name="002712" />九</ins>張見写料
</line>
</論理文書></本文>
</大日本古文書本文>
<大日本古文書本文><大日本古文書巻頁>14-366</大日本古文書巻頁>
<本文><論理文書 id="0001-2">
<line> 一千七百十二<ins>七</ins>張<f>はなだ</f>紙料</line>
<line> 三千張儲料</line>
<line> 凡紙三千八百五十三張<wari>一千六百二張経師六万四千八十張</wari><ka>端</ka><wari>
継料以一張着</wari><gi set="mojikyo" name="002712" /><wari>張</wari></line>
<line> 兔毛筆三百<del>七十三</del><ins>九十七</ins>管 墨<del>一百九十九廷</del><ins>二百廷
</ins></line>
<line> 鹿毛筆一百管 綺五百<del>七十九丈七尺</del><ins>八十一丈二尺</ins></line>
<line> 軸三千四百卅四<ins>三</ins>枚 狸毛筆廿五管</line>
<line>以前応奉写一切經<ins>一部</ins>調度如件</line>
<line> 天平宝字四年二月十九日</line>
<line> 右於承前奉写経律用度、加今渡坐経巻</line>
<line> 勘注進送如件</line>
<line_e> 東寺下任安都</line_e>
</論理文書></本文>
</大日本古文書本文></紙>
<切断 /><白紙 />
<紙>
<大日本古文書本文><大日本古文書巻頁>14-366</大日本古文書巻頁>
<本文><論理文書 id="0002-1">
<line_s>次官従五位下高麗</line_s>
<line>造東大寺司移散位寮</line>
</論理文書></本文>
</大日本古文書本文>
<大日本古文書本文><大日本古文書巻頁>14-367</大日本古文書巻頁>
<本文><論理文書 id="0002-2">
<line>散位正八位上秦忌寸太菜</line>
<line> 右依太師去正月十一日宣為令奉写 出願経</line>
<line> 件如以移</line>
<line> 天平宝字四年二月</line>
</論理文書></本文>

  中略

</大日本古文書本文></紙>
</成巻本文>
</成巻>
</正倉院文書>

```

図 4 生成された XML 文書

4-1. 論理表現

論理構造文書の表示を行うために XSL (eXtensible Stylesheet Language) によるスタイルシートを作成し、XML を web 上で表現できる機能を作成している。論理構造で表示する手続きに XML を使い (図 5-1)、図 5-2 中の "shosoin-main.xml" にあたる。この "shosoin-main.xml" で指定したスタイルシート shosoin1.xsl は、XML 化された正倉院文書を読み込み、論理構造を構成して Web ブラウザに表示する [5]。その結果を図 6 に示す。

```
<?xml version="1.0" encoding = "Shift_JIS"?>
<?xml-stylesheet type="text/xsl" href = "shosoin1.xsl"?>
<正倉院文書>
<volume no="1">正倉院文書 成巻18-6 </volume>
</正倉院文書>
```

(1) 主 XML 表現

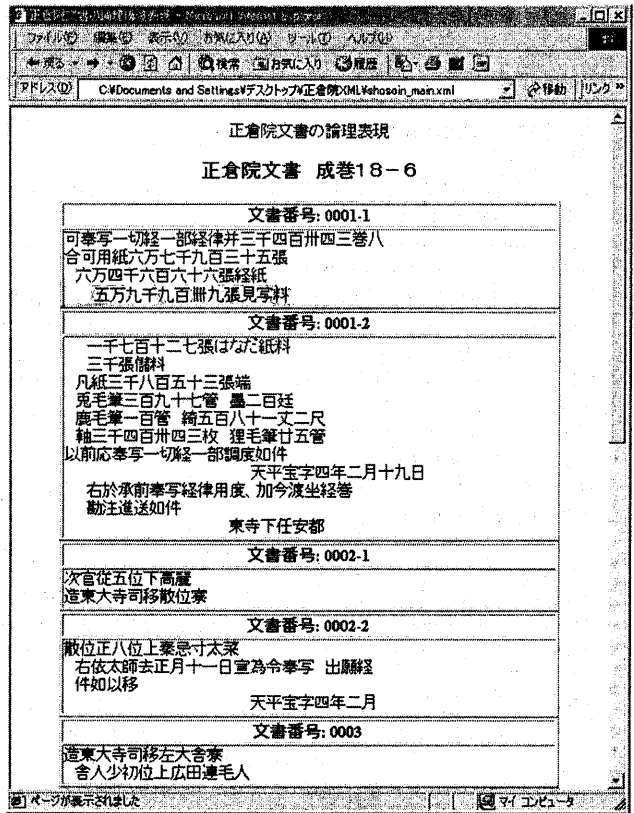
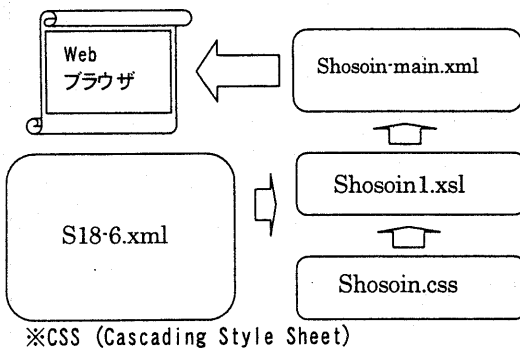


図 6 論理構造の Web ブラウザ表示



(2) XML 文書の構成

図 5 XML 文書からの論理構造表示

4-2. 短冊

物理表現である短冊を表現するためには、Word マクロを使用した。必要なタグ (紙、大日本古文書巻ページ、初行、終行) を検索して、該当部分

に、挿入していくことによって、今まで非常に手間のかかっていた、短冊作成の作業の支援ツールとなる (図 7)。その方式としては、ウィンドウで大日本古文書の巻ページを参照しながら、位置を決定して挿入していく。また、該当文書の内容を確認できるように、別のウィンドウを作成した。これらによって、短冊作成の作業時間の短縮と、正確性の向上がのぞまれ、研究効率の向上がのぞまれる。

今後の検討課題としては、写経所文書研究会 [6] が、作成しているデータベースを用いることによって、詳細な接続情報や、文書情報を今後充実させ、料紙情報をより正確に表現することが、まず、あげられる。また、研究の対象となる文書の「背面」は通常、天地逆転して表示されるため、この表記方法を作り出すことも今後の課題となる。

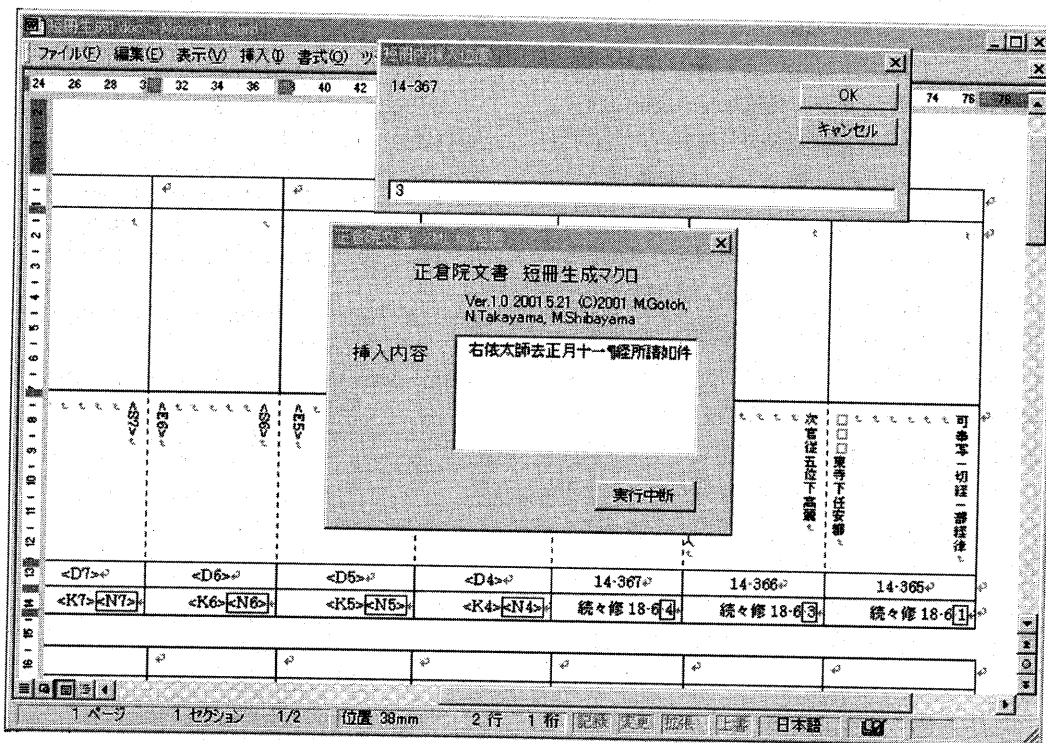


図7 MS-Word マクロによる「短冊」生成

4-3. 正倉院文書研究支援システムの実現

それでは以上のようなツールを具体的には、どのように実現していくのか。以下にその検討状況を示す(図8)。実現のシステムとしては、クライアント/サーバ方式をとる。クライアントである歴史研究者は Web を通して必要な史料を入力する。サーバは、正倉院文書データベース、および構造化ルールを用いて XML 文書を作成する。クライアントは、その XML 文書を論理構造として確認したい場合には、Web ブラウザによって表示される。そして、短冊として確認したい場合には Word マクロと XML 文書をダウンロード可能とする。また、XML 文書そのものとして手に入れることも出来る。以上のようなシステムを検討中である。

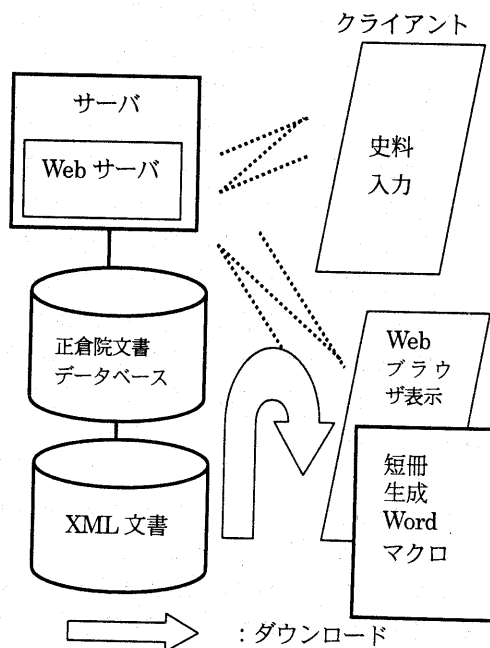


図8 研究支援システムの構成

5. 今後の検討課題と展望

今後の検討課題としては、まず構造化ルールの作成が第1に挙げられる。構造化ルールを作成することによって、歴史研究者が「復原」として現在、行っている作業を一般化、あるいはモデル化することになる。また、例外を見つけだすことにより、これまで研究上で漠然と意識されてきたような事象が明確になる。たとえば、帳簿作成者ごとに微妙に異なると言われる、帳簿作成の方法もより明らかになってくるであろうし、時期による段階差も明白となって来るだろう。また、ルール自体の理論化は、歴史研究者がはっきりと意識せずにおこなっている作業を理論化することでもあり、「古代帳簿論」の一方法として寄与できるようになるのではないかと考えている。

また、本ルールを組み込んだシステムの実現によって、歴史研究者の負担の軽減が、大いに期待できる。

第2に短冊の作成ツールの完成である。短冊作成には、接続情報（ハガシトリ痕や墨痕の有無）や、料紙の状態が不可欠であるが、現状ではそれを表記できていない。技術的な問題なので、比較的早急に対応することができると考えている。

第3は論理的な表現によって、研究に寄与できる環境をより整えることである。たとえば写経所文書研究会データベース参照形態の完成によって、個々の写経事業における当該文書すべてを論理的な表現において見ることも、可能となるだろう。

正倉院文書という、複雑な構造と表記形式を持つテキストにたいしてデジタルデータ化することは、非常に困難と歴史学研究的なかでは思われてきたが、XMLを用いることによって、その形態と表記は可能である。今後のアプローチの仕方が課題となる。

最後に本論文を作成するにあたって大阪市立大学大学院文学研究科の柴原永遠男氏をはじめ、同大学大学院文学研究科日本史研究室の方々、写経所文書研究会の方々に多くのご教示をいただいた。ここに末文ながら謝辞をあらわす次第である。

〔参考文献〕

- 〔1〕 正倉院文書研究に関して多数の文献があるが、概略的なものや正倉院文書研究自体の文献として以下のようなものがある。
柴原永遠男：2000『奈良時代の写経と内裏』
杉本一樹：1999『正倉院文書の原本調査』『古代文書論』
石上英一：1999『正倉院文書目録編纂の成果と古代文書論再検討の視角』『古代文書論』
山下有美：1999『正倉院文書研究の展開と史料へのアプローチ』『歴史科学』156
土田直鎮：1992『正倉院文書について』『奈良平安時代史研究』
吉田孝：1983『律令時代の交易』『律令国家と古代の社会』
正倉院文書一般に関しては適宜これらを参照した
- 〔2〕 柴原永遠男：1993『正倉院文書関係文献目録』『正倉院文書研究』1, 2, 3
- 〔3〕 柴原永遠男〔1〕論文より
- 〔4〕 XMLに関する主な文献には以下のようなものがある。
Khun Yee Fung :2001"XSLT:working with XML and HTML", Addison-Wesley
山田祥寛：2001『XML サンプル集』秀和システム
<http://msdn.microsoft.com/xml/default.asp>
XMLパーサ等に関する情報
<http://www.w3c.org/>
XML全般、XSL及びXSLTに関する説明

<http://www.microsoft.com/japan/developer/workshop/default.asp>
本論文で用いたXMLパーサは、MSXML3、XSLではXSL-Transformを用いた。
- 〔5〕 外字表示に関しては、文字鏡研究会編：2001『パソコン悠悠漢字術』を参照した。
- 〔6〕 写経所文書研究会とは正倉院古文書の中でも二次面に多い写経所文書を中心に分析し、8世紀の写経所と写経の実態を明らかにしようとする研究会であり、この研究会において個々の写経事業ごとに正倉院文書の当該文書を明らかにするというデータベースの蓄積作業を行っている。