

# 『續一切經音義』からみる 漢文文献の TEI マークアップの課題

王 一凡（東京大学大学院 教育学研究科）

渡邊 要一郎（東京大学 史料編纂所）

永崎 研宣（人文情報学研究所）

下田 正弘（東京大学大学院 人文社会系研究科）

**概要：**本稿では大正蔵 54 卷所収の希麟撰『續一切經音義』テキストデータの TEI 適合化を行うにあたり、原資料の特殊な文献学的またはテキスト的性格によって生じた、考慮すべき諸論点を検討する。これに基づき、東アジア資料の TEI エンコーディングのための若干の提案を行う。

**キーワード：**TEI, 仏典, 漢籍, Unicode

## *Xu Yiqiejing Yinyi* and Issues on TEI Markup of Chinese Literature

Wang Yifan (Graduate School of Education, The University of Tokyo)

Watanabe Yoichiro (Historiographical Institute, The University of Tokyo)

Nagasaki Kiyonori (International Institute for Digital Humanities)

Shimoda Masahiro (Graduate School of Humanities and Sociology, The University of Tokyo)

**Abstract:** This paper aims to provide some useful suggestions for better TEI encoding of East Asian literature, while discussing several aspects of consideration that emerged from the interaction between the standards and special textual/philological characteristics of *Xu Yiqiejing Yinyi* by Xilin, a scripture included in the Taisho Tripitaka, vol. 54, during TEI conversion from our former text format.

**Keywords:** TEI, Buddhist scripture, premodern Chinese, Unicode

### 1. 背景

文献研究においてテキストデータを構造的に記述することは、当該研究に際して用いられた解釈や知見を共有してその後の研究発展の礎とする上で有用であり、それは機械可読性が高ければ高いほど効果的なものとなる。TEI (Text Encoding Initiative) ガイドラインは、まさにそれを実現するための枠組みの一つであり、近年は東アジア・日本資料への適用も徐々に広がりつつある。SAT 大蔵経テキストデータベース研究会では、大正蔵の電子化テキストを独自形式で入力し、ウェブ上でデータベースとして公開しているが、標準化推進の一環として、またオープンな人文学テキスト利用の普及を図るため、既存テキストの TEI 化を試みている。

『大正新脩大蔵経』（大正蔵）は、漢文仏典を網羅的に集成した大蔵経の一種であり、大正 13 (1924) 年から昭和 9 (1934) 年にかけて日本で編纂されたものである。底本として高麗版大蔵経などの善本とされる典拠を採用したことに加え、伝統的な大蔵経の形式を離れ、近代仏教学の知見を取り入れた学術的な校訂を施しており、仏教研究における標準テキストとして広く利用されている。

対象となる大正蔵 54 卷所収の『續一切經音義』とは、遼代に成立した文献で、当時存在していた漢文仏典中にみられる術語、難語などを經典内での出現順に集成した「音義書」と呼ばれる種類の辞典である。辞典として同時代に並んで一般的であった「字書」「韻書」の形式とは異なり、見出し語には通常、単字ではなく、二字語や句が立てられている。語を構成する字に対する解説は、その積文の中で適宜施されている [1]。

また、『續一切經音義』の配列順は、収録する經典の巻分けに従っており、以前の巻もしくは経で既に解説された語が重出することを厭わない（「已釋」などの標示がなされていることもあるが、それにかかわらず再度解説されることが多い）。ここから、想定している用途は經典を読む際に、同書を脇に置きながら逐次参照する形であったと推察される。

『續一切經音義』自身は全 10 巻の構成であり、その全内容に照らして概ね等分である。巻の区切りは収録する經典の形式的区切り（巻など）と一致するよう設けられており、経巻の途中で分巻することはない。『續一切經音義』の巻は元来卷子本時代の物理的な切れ目であったと思われるが、後世の刊本に収録される際に必ずしもそれが物理的に踏襲されたわけではなく、今回の底本であ

る大正蔵（活版印刷）では10巻を連続して配置しているため、同書のページ区切りとも一致しない。

『續一切經音義』の現存が確認されている完本は高麗大蔵經（高麗蔵）再雕本をはじめ、獅谷白蓮社本、大日本校訂縮刷大蔵經（縮刷蔵）本、頻伽精舎本、大正蔵本、中華大蔵經本があるが、すべて高麗蔵を祖とする系統に連なる。特に形態上は、高麗蔵を含めすべて版本（整版ないし活版）であることが特徴的だが、版式は諸本で特に一貫しているわけではない。高麗蔵の版木は現在まで残っているが、長年の使用による摩耗により後世の刷りには補刻の跡があるとされる [2]。諸本に出現する字形も、覆刻が行われた時代・地域の影響でしばしば改変されていることがある [3]。大正蔵は概ね近代までに分岐した各本を比較対校したものとされているが、それも完全ではない [4]。

本研究においてこのテキストの構造を記述するために採用した TEI とは、人文学において利用する電子テキスト資料の構造を記述し効率的効果的に共有するために 1987 年に策定が開始されたガイドラインである [5]。主に欧州と米国の研究者らによって開始された TEI ガイドラインは、汎用的なものを目指しているものの、すべてのタイプの人文学資料に対応するにはそれぞれに即した検討と必要に応じた拡張が必要である。また、人文学資料とそれに対する観点の多様性から、規格ではなくガイドラインという位置づけになっており、利用者は必要に応じて要素や属性等の構造を選択して利用することになっている。このため、東アジア資料に TEI ガイドラインを適用するためには、まずは東アジア資料に独特な文書構造が TEI ガイドラインで適用可能かどうかを検討する必要があり、その上で、自らの研究に必要な構造を選択する必要がある。この選択に共通性がないと他の同種のテキストとの相互運用性が低くなってしまうため、同種のテキスト同士では可能な限り同じルールで構造化できるように、共有可能な TEI ガイドラインのサブセットを作成する必要がある。本研究はそれを実現するためのプロトタイプングとしても位置づけ得る。

## 2. 構造化

TEI は XML に従い、テキストの各部分を要素とみなして、ツリー状に構造化するが、この時にテキストのアウトラインを構成する意味上の段落をツリー上の要素として分割し、ページや分冊などの形式上の区切りは空要素を挿入して示すのが慣例である [6]。この原則を大正蔵所収の『續一切經音義』に適用しようとする、その重層的な構造のためにいくつものマークアップ戦略が必要となる。

第一に、本文献における「意味上の単位」が指すものの解釈に幅がある。

- 原經典の単位
- 原經典の巻の単位
- 原經典の意味的な段落の単位
- 『續一切經音義』の巻の単位

通常の經典であれば、読解を通して意味的な単位を理解できるが、本文献の性格はなかば集注的なものであって、参照先の經典の段落に関する情報は一切表面から読み取れず、經典の巻相当の分節（これも現行版では名目上のものであることが多い）を表す題とともに、見出し語が延々と羅列されているにすぎない。そのため見方しだいでは『續一切經音義』のアウトライン自体がその目的に即した意味的分節ともいえるし、逆に同文献に意味上の下位分節は存在しないととらえることもできる。

一方で形式上の区切りについても、先述のもの以外に

- 大正蔵におけるページ、行区切り

が考慮の対象として加わる。このような複層的な構造をどのように扱い、混在させてマークアップするのが最善かを検討しなければならない。

ただし大正蔵における形式情報は、SAT が保有する元データの段階で、ページ、段（上・中・下）、行番号が完全に記述されている。そのため、これらを<lb>（行開始）タグに集約して以下のような形式で1行ごとに記述し、全体の構造にはあまり影響を与えず挿入することができた。

```
<lb n="T2129_,54,0934a02"/>
```

以上を踏まえたうえで、『續一切經音義』では次の模式図のような構造でマークアップすることとした。

**階層関係：**『續一切經音義』の本文で解説されている各經典の区間を内容の単位とみなし、主たるアウトラインを構成するものとした。この階層は本文全体を<div1>、各經典を<div2>で囲んでいる。また、經典に分巻がある場合は、その巻ごとに<div3>で区別した。經典自体の内容的区分は、『續一切經音義』内の配列から明らかに推測できる場合もあるが、ほとんどはそうではない。また上述の通り『續一切經音義』はしばしば前出の単語を再掲するが、この時の繰り返しは巻単位でなされているように観察された。したがって、原經典においては分巻は形式的な切れ目であるが、音義中での章分けとしては機能的な切れ目であると判断した。そして『續一切經音義』の本文中ではその他の意味的な段落を示唆する標識に乏しいため、このような分け方でも支障は生じないと思われる。

**章題と内容：**『續一切經音義』では、分巻のない經典は經典名の見出しを置き、分巻された經典については必ず巻ごとの見出し（○○經卷 X）を置いている（つまり經典全体の見出しはない）。これを反映して、不分巻の經典は<div2>直下に、

分巻のあるものは<div3>の中に見出しを配置した。見出しに続いて各掲出語の羅列が続く、これらは TEI の辞書モジュールを用いて詳細に構造化することも可能であるが、現段階の目標は Best Practices Level 3 [7] 相当のマークアップを達成することであるため、行と割注を標示する程度に留め（後述）、今後の課題としている。また、一部の経典は見出しを挙げるのみで、割注に「無字可音訓」とのみ記されているものがあり、その場合内容部分は存在しない。

**音義の巻・目次**：経典の区切りにかかわらず、『續一切経音義』自体の巻ごとに巻題と収録経典の目次が付されている。これらは視覚的にも比較的大型の要素であるが、上述の通り本文のアウトラインの一部とはみなさなかつた。その代わり、空要素である<milestone>を音義の巻首に置き、それに続けて、補助要素であることを示すために、巻題と目次をそれぞれ<note>要素に入れて示した。これらの要素は内容の階層いかんを問わず挿入される可能性があるが、本文に割り込んでまで挿入されることはなく、内容面から見ると何かしら一段落した箇所に差し込まれる形となる。そのため、マークアップ上は直近の要素を閉じ、次の同階層の要素が始まる前にこれを置くこととした。すなわち巻要素の切れ目に挟まれたものは<div3>の間、経典要素の切れ目に挟まれたものは<div2>の間に置かれる。

目次は、経名の羅列なので<list>でマークアップした。まれに目次中出现する経典と実際にその（音義）巻中に収録されている内容や順序に齟齬がある場合がある。その場合は必ず本文を基準とし、目次の側に訂正情報を標示した。例えば次のようなものである。

```
<item>大<choice><sic>樂</sic><corr>藥</corr></choice>叉女歡喜母并愛子成就法一卷</item>
```

なお、古典籍によく見られる形式として、音義の巻の最後にも尾題として巻名を改めて記すことがあるが、現在のところこれはプロジェクトの他の文書と合わせて<ab>で記述している。ところが、TEI の制約により、先頭要素を除き<div>の同階層に<ab>などの段落要素を置くことができないため、現時点ではガイドラインへの準拠性を保つために前の子要素に繰り入れている。この扱いに関してはなお検討を要する。

### 3. 符号化文字

『續一切経音義』は成立後中国大陸から永らく失われ、高麗にのみ伝存したため、後世の歴代中国王朝で刊行された大藏経に収録された文献と異なり、字種の系統立った正規化がなされていない。それは一方で伝写による訛誤を多く孕むということであるが、同時に後の字体規範の影響を免

```
<div1>——本文
<milestone/>——音義巻開始
<note><title>音義巻題</title></note>
<note>
  <list>音義目次</list>
</note>
```

```
<div2>——経典（分巻）分
```

```
<div3>——経典巻／品分
```

```
<header>
  <title>経典巻題</title>
</header>
```

```
<p>
  <lb/>内容
  <lb/>……
</p>
```

```
</div3>
```

```
<div3>
  内容……
```

```
<ab>音義尾題</ab>
```

```
</div3>
```

```
<milestone/>——音義巻開始
<note>音義巻題・目次</note>
```

```
<div3>……</div3>
```

```
<div3>……</div3>
```

```
</div2>
```

```
<div2>——経典（不分巻）分
```

```
<header>
  <title>経典題</title>
</header>
```

```
<p>内容……</p>
```

```
</div2>
```

```
<div2>
  内容……
```

```
<ab>音義尾題</ab>
```

```
</div2>
```

```
<milestone/>——音義巻開始
<note>音義巻題・目次</note>
```

```
<div2>……</div2>
```

```
<div2>……</div2>
```

```
</div1>
```

れているという面もある。同じ遼代には、仏典に出現する異体字を多く収集し、稀字を多数含むことで知られる『龍龕手鏡』という字書が編纂されているが、実際に『續一切經音義』にはこれと相照らす字形など、興味深い用例が数多く現れることから、当時の書記習慣の実態が反映されているものと考えられる。従って、文献中に出現する異体字の情報はできるだけ保存する価値があるものである。

ところで、TEIはXML標準を基盤としており、XMLはソースをUnicode符号で記述することを求めている。TEIの現有の枠組みでも外字（ここでは、XML標準の準拠するUnicode標準に未収録の字種または字体）を記述するための基本的なモジュールが定義されており、TEI文書内で外字を注釈することが可能である [8]。

```
<TEI>
  <teiHeader>
    <encodingDesc>
      <charDecl>
        <glyph xml:id="per">
          <localProp name="Name"
value="LATIN ABBREVIATION PER"/>
          <figure>
            <graphic url="per.png"/>
          </figure>
        </glyph>
      </charDecl>
    </encodingDesc>
  </teiHeader>
  <text>
    <body>
      <p> ... <abbr><g ref="#per">per
</g></abbr> ardua ...</p>
    </body>
  </text>
</TEI>
```

しかし、東アジア（特に前近代）文献における外字の立ち位置は、少数の理論的に閉じた文字集合を前提とする書記体系と比べ次のような差異がある。

- (1) 既存の符号化文字に還元できない外字が無視できない頻度で出現する
- (2) 外字の出現は偶発的ではなく構造的であり、Unicode標準の実務的制約によって表現することができない「待機文字」の比率が多い
- (3) 外字を既に収録している非Unicode系の外部文字集合が存在することがある

(1) に関して、ラテン文字圏でも前近代の写本では複雑な合字などが多用され、その中にはUnicodeに未収録の文字も多数存在する [9]。しかしこれらは表記法の情報を除けば基本的に現

代の字母の組み合わせに還元できるものであり、外字は置換してしまうか、あるいは注記情報自体が（左記の例のように）簡易なものであっても事足りる。そしてそのような対応づけが困難なものは、そもそも未判読文字であることが多い。それに対し、東アジア文献においては、特に漢字は部首を組み立てて造字することが可能なため、明確に判読できるにもかかわらず字義が不明であるとか、意味上は既存字に同定できるが異同が甚だしい（古字・俗字など）ために、外字によってしか満足に表現できないケースも多い。

また、(2) を理解する前提として、Unicodeに文字を収録するためには、単なる申請と受理ではなく、担当委員会による審査を経なければならない。漢字については、1字1字の文字情報や既収録字との衝突を検査するため、現時点では約2~3年ごとに5,000字のペースでしか新規収録が行えない（もちろん学術用字のみならず、人名・地名・その他地域・方言専用字などが多く含まれる）。したがって、存在と必要性が明らかに確認されているにもかかわらず、手続き上の理由によって長期間Unicodeへの収録が見込めない待機文字が多く存在する。また、Unicodeの漢字対応は徐々に複雑化しており、収録を望む文字をIVSと呼ばれるしくみによって修飾符号つきの変種（規格上の異体字）として登録することが勧告されることもある。この場合は、基底文字となる符号位置が並行して審査中である場合、完全な収録までに2段階のプロセスを経る必要がある。また、審査は各地域の代表の合議によって行われるため、必ずしも提案者や個別の当事者にとって一貫した時期・一貫した方式で登録されるとは限らない。

TEI文書のエンコーディング作業にとって、できるだけ統一的に記述すべきにもかかわらず、時期と規格の制約によって、個々の字を異なる手段で記述しなければならない、さらに、標準化作業の進展に伴って収録字が徐々に増加していくと、過去に作成した文書、あるいは異なる時期に作成した文書どうしで記述方式が異なることで、文字の同定や比較が難しくなる恐れがある。これを避けて東アジア文献のTEI文書の永続性を保つためには、外字に対するアイデンティティの追跡を可能にする手段を講じなければならない。

(3) について、漢字圏における電子テキストで表現可能な字種の不足は長年の課題であり、特にUnicodeを含め各種規格の収録字種・字形数が既存出版物と比べても顕著に不足していた2000年代中頃までは、学術的な大規模文字集合の構築をめざして多様な試みが行われた。そこでは技術的にもインターフェイス的にも様々なレイヤーによる解決策が模索された。日本国内における代表的な事例として「今昔文字鏡」（画像形式・文字集合・共有システム）「GT書体」（フォント形式・符号化文字集合）「e漢字」（公開終了、文字集合・データベース）などが知られる。前世紀

末頃の文献 ([10] など) ではこれらの技術の得失や適応を盛んに論じており、当時の普及の程度がうかがえる。現在では、量的に Unicode がすべてをカバーしつつあるが、このような非 Unicode 集合を使用するテキストが残存していることに加え、Unicode との符号化方針の差異により、対応関係が必ずしも明確ではなく、変換が自明ではないという問題がある。そのような公共的な集合でなくとも、大正蔵のように、扱う文献の規模が大きければ、管理のために一定規模の外字集合を規定していることもある。

一般化すれば、このような文献における外字の発生は TEI が基盤とする諸技術の情報の空隙に由来するものであり、よりリッチな表現を可能にするための補助的な上位レイヤーという道具立てでは不足することが多い。これに対処するためには、外字のワンオフ的なマークアップ機構のみならず、当該字の同定に必要な文字情報を包含しうるだけの記述力を外字モジュールに持たせなければならない。

執筆時現在の TEI 外字モジュール [8] の機能では、ヘッダー内に書かれる文字定義要素に主に以下の内容を含めることが可能である。

- Unicode の文字属性 (<unicodeProp>)
- Unihan の文字属性 (<unihanProp>)
- カスタムの文字属性 (<localProp>)
- 対応字もしくは類字 (<mapping>)
- 画像など (<graphic>)

以上の道具立てでは、Unicode に準拠した情報の入力には手厚くサポートされているが、外字が有している情報への対応は、未だ貧弱である。特に Unicode 以外はすべて非標準扱いのためか、外部情報と対応づけを行う属性が設けられておらず、基本的には文書内で個別に出現した字形の自由記述という性格が強い。これは上述 (3) の観点から問題となる。

また、(2) に関して、漢字はしばしばある文字が標準化 (またはそれ以前) のプロセスで複数の符号化表示が割り当てられることがあるが、その履歴を記録するための機構がないことが問題となる。さらにこれに関連して、現在の TEI の意味論では、Unicode 規格にも存在するバージョン間で変更される可能性のある属性の記述にやや不便があることがわかった。

このような問題を解決するために、TEI ガイドラインに以下の拡張を提案しており、現在議論中である (属性名はすべて仮称)。

- 文字属性要素への非 Unicode ソース・バージョン属性の追加 (@otherSource, @otherVer)
- 文字属性要素への最小 (追加) ・最大 (廃止) バージョン番号属性の追加 (@minVer, @maxVer, @otherMinVer, @otherMaxVer)
- <mapping> 要素への日時記述用属性群 (att.datable) の導入

これらの追加によって、下図のような記述が可能となり、外字を含む TEI 文書の永続的な文字情報の対照・同定に寄与すると考えられる。

#### 4. 割注

『續一切経音義』の記述フォーマットは、全体にわたって見出し語の後に二行立ての割注で釈文を記載している。

現時点の『續一切経音義』マークアップでは、一時的に割注範囲を<seg type="wari">、割注内の改行を<lb type="wari"/>で示しているのみである (『續一切経音義』では概ね割注は釈文専用であるが、本文の意味的マークアップにはまだ立ち入っていないため、汎用要素で対応しているにすぎない)。

割注という形式自体は意味中立的な道具であり、内容のマークアップにおいて割注であるか否かはさほど重要ではないが、割注は一行を対等な

```
<char xml:id="myChar" source="http://gaiji.example.com/gaiji-0000">
  <unicodeProp name="Name" value="CJK UNIFIED IDEOGRAPH-NNNN" version="1X.0">
  <unihanProp name="kIRG_USource" value="U-012345"
    minVer="1X.0" maxVer="1Y.0" />
  <unihanProp name="kIRG_SSource" value="S-567890" minVer="1Z.0" />
  <localProp name="Name" value="MY GAIJI" />
  <localProp name="Reading" value="GAI"
    otherSource="http://gaiji.example.com" otherVer="2012.10" />
  <localProp name="Info" value="foobar" otherSource="http://gaiji.example.com"
    otherMinVer="2012.10" otherMaxVer="2015.01" />
  <mapping type="internal">0xABCD</mapping>
  <mapping type="PUA" from="2012-01-01" to="2018-03-31">U+FXXX</mapping>
  <mapping type="standard" from="2018-04-01" to="2019-10-15">U+YYYY</mapping>
  <mapping type="standard" from="2019-10-16">U+YYYY U+E0100</mapping>
</char>
```

※<mapping>の値は、実際の文字を記入することが想定されている。ここでは説明のためにコードポイント表記で示している。

複数行に割るという特殊な操作にもかかわらず、東アジア文献に普遍的に出現する形態であって形式上の均質性が高いことから、東アジア文献を記述した TEI 文書の可搬性・通用性を確保するためにも、何らかの規格上のサポートが必要ではないかと思われる。

[11] でも詳述したが、特に、割注において問題となるのは改行の標示である。TEI においては基本的に同一レベルの行が連続する様態を想定しているため、行（段落）を入れ子にしたり、改行がどの階層に属するかを示したりする標準的な機構が存在しない。割注は時折 2 階層以上深くなる場合があるため、汎用的に入れ子された注記内それぞれの階層の改行を区別可能なマークアップ手段が求められている。

ここで [11] での議論を整理すると、技術的な実装方針として、大きく 2 点の論点が現れている。

- 割注（の改行）に対して特別の要素を用意するか、もしくは小さな行として扱う（`<lb/>`と要素を共有する）か
- 割注（の改行）の階層の表現として、深さを示す属性を追加するか、もしくは特定の範囲を親（包含要素）として指定することでその内部の改行であることを示すか

割注それ自体の固定した意味は乏しく、一つの行を書き方向に対して平行に複数段に分割することが特徴であると考え、仮に親要素への参照を介して割注の階層を指定する場合、幅広い種類の親要素を取ることとなり、加えて、原理上個々の範囲ごとに異なる識別子を与えなければならない。これはレンダリング上もクエリ（割注だけを抽出するなど）上も負担が大きいと考えられ、また XML ソースの記述も繁雑になるきらいがある。割注はほとんどの場合範囲要素に囲まれた形で出現すると期待できるとしても、他要素に依存せず、割注改行要素自身の属性のみによって意味論を構成する方が堅牢かつ簡便であろうと考えられる。

次に、割注の改行要素と一般の改行要素`<lb/>`の区別を設けるかどうかについては、割注が親となる行への埋め込みであり、親行の改行は必ず割注の改行を兼ねるといえる（さらに一般化すれば、上の階層の行の改行は必ずそれが含む下の階層の要素すべてに波及する）。これを考えると、割注の改行そのものに特別な要素を割り当てる必要はなく、むしろ`<lb/>`に一本化することでルールが簡潔になると考えられる。

なお、割注全体の表示は、親要素への `@style` もしくは `@rendition` 属性と当該改行要素の拡張を組み合わせることで対応可能であると考えている。

したがって、割注への正しい対応のためには、`<lb/>`要素に以下のふるまいを持つ属性 1 つを付け加えるべきである。

- 数値を値として取り、指定がない場合は 1 とする（通常行の改行を意味する）
- 属性値はこの`<lb/>`要素の属する階層の深さを意味し、ある要素は同時に自身より大きい値の階層すべてにおける改行を意味する

このような方向で TEI ガイドラインへの提案を行う予定である。

## 5. まとめ

本稿では大正蔵所収『續一切経音義』を題材に、大規模な漢文文献に TEI を適用する際に生じる課題を紹介した。これらは大なり小なり漢文文献のエンコーディングにあたって立ち現れる基礎的な問題であると認識しており、今後のプロジェクトにも応用していきたい。また、引き続き検討しなければならない事項や、解決が不十分な点などがまだ多く残っていると思われるので、ぜひ読者の皆様からのご指導、ご批評を賜りたい。

## 参考文献

- [1] 徐時儀. 一切経音義三種校本合刊緒論. 徐時儀(校注). 一切経音義三種校本合刊. 上海古籍出版社, 2012, 修訂版, p.1-155.
- [2] 京都仏教各宗学校連合会. 大蔵経: 成立と変遷. 法蔵館, 2020, 新編.
- [3] 王一凡. 慧琳撰『一切経音義』の符号化をめぐって. 下田正弘, 永崎研宣(編). デジタル学術空間の作り方: 仏教学から提起する次世代人文学のモデル. 文学通信, 2019, p.275-296.
- [4] 永崎研宣. 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, 2016, p.265-280.
- [5] “History”. <https://tei-c.org/about/history/>, (参照 2021-11-01).
- [6] “Default Text Structure”. <https://tei-c.org/Vault/P5/4.3.0/doc/tei-p5-doc/en/html/DS.html>, (参照 2021-09-06).
- [7] “Best Practices for TEI in Libraries”. <https://tei-c.org/extra/teilibraries/>, (参照 2021-11-01).
- [8] “Characters, Glyphs, and Writing Modes”. <https://tei-c.org/Vault/P5/4.3.0/doc/tei-p5-doc/en/html/WD.html>, (参照 2021-09-06).
- [9] Medieval Unicode Font Initiative. <https://mufi.info/>, (参照 2021-11-01).
- [10] 漢字文献情報処理研究会(編). 電腦中国学. 好文出版, 1998.
- [11] 王一凡, 永崎研宣. 東アジア文献への TEI の適用をめぐって. 情報処理学会研究報告, 2018, 2018-CH-118, No. 4, p.1-4.