仏教辞典『法寶義林』目録のデジタル化
── TEI と XML DB を利用した情報の共有化に向けて ──

松田 訓典
東京大学 東洋文化研究所

彌永 信美
フランス極東学院

永崎 研宣
人文情報学研究所

下田 正弘
東京大学 人文社会系研究科

現在,『法寶義林』(*Hōbōgirin, Dictionaire encyclopédique du bouddhisme d'après les sources chinoises et japonaises*) に収められた仏教の経論の著者・訳者の目録部のデジタル化を TEI の標準的なマークアップ方法にしたがって進め，XML DB を利用したインターフェースを構築中である．本報告では，その概要とそこに起こりうる問題点，それに対する今後の展望を含めたその解決へのアプローチについて論ずる．

The Digitization of Buddhist Dictionary *Hobōgirin* using the TEI Guidelines and XML Database

Kuninori Matsuda
Institute for Advanced Studies on Asia
University of Tokyo

Nobumi Iyanaga
École française d'Extême-Orient
Centre de Tokyo

Kiyonori Nagasaki
International Institute for Digital Humanities

Masahiro Shimoda
Graduate School of Humanities and Sociology
University of Tokyo

We are now working on the project of digitizing the French Dictionary of Buddhism, *Hōbōgirin, Dictionaire encyclopédique du bouddhisme d'après les sources chinoises et japonaises*. Our project follows the standard XML mark-up guideline of TEI and uses a native XML database. In this report, we will state the outline of the project, its problems, and some of the approaches being used to resolve them.

1. はじめに

『法寶義林』(*Hōbōgirin, Dictionaire encyclopédique du bouddhisme d'après les sources chinoises et japonaises*) とは，その副題が示す通り，中国・日本の資料に基づいた仏教に関する一種の百科事典である．本書は，1920 年代，仏教学研究の一大拠点であったフランスにおいて計画されたヨーロッパで最初の大規模な仏教辞典であり，フランス学士院 (Institut de France)，日本学士院，フランス極東学院 (École française d'Extrême-Orient)，日仏会館 (Maison Franco-Japonaise)，フランス国立科学研究センター (Centre national de la recherche scientifique) といった諸機関の後援を受けて進められた一大プロジェクトとして，1929 年に A から Bombai までの項目を収めた第一分冊が刊行されて以来，2003 年には Daishō Kongō から Den'e までを収めた第八分冊まで刊行されており，この間二冊の別冊が発刊されている．一つは大正新修大蔵経（以下，大正蔵）に収められた経論の目録，その著者・訳者の解説付き目録（1931, 改訂増補版 1978, [1]）であり，もう一つは第一分冊から第四分冊までの漢字索引（1984）である．

この『法寶義林』の性格としては，

> 第一には，印度学者には中国・朝鮮・日本に於ける仏教教義の発展について資料蒐集を可能にし，一方，東洋学者には仏教によって伝播された多くの概念の印度起源を見出すことを可能にすることであり，第二には，西洋人に日本の同時代的な仏教学の業績を紹介する

と紹介されており[1]，その成果は，既存の日本の仏教辞典や研究を踏まえつつ，新規情報をも盛り込んだものとなっている．そしてその '広い文化史的視野' によってもたらされた学術的成果のインド学・仏教学への貢献は，日本においても既に広く認められているところである ([4])．

[1] 文献 [2] p. 99．なお，H. Durt 氏は現在も同プロジェクトにおいて中核的な役割を果たしており，文献 [3] でも『法寶義林』について紹介している．

その中でも，先述の別冊に収録されている，大正蔵に収められた経論の著者・訳者の目録部分のオンライン化を，現在，大蔵経テキストデータベース研究会（SAT）とフランス極東学院の協力のもと進めている．本プロジェクトにおける TEI（Text Encoding Initiative)[2] のガイドライン（[5]）に沿った XML マークアップの方法の実例と，それに際して見出された問題点に関しては既に報告した（[6]）．本報告ではさらにそれを踏まえ補足した上で，実際のデータベースとして XML DB を利用した本プロジェクトの現状とその課題について報告する．

2. 目録の構成

まず，当該目録の構成を改めて間単に紹介しておこう．目録のタイトルは Table des noms d'auteurs et de traducteurs rangés par ordre des transcriptions sino-japonaises avec notices biographiques とされており，900 人あまりの人物（インド，中国，日本等）の情報が，日本語読みのアルファベット順で収録されたものである．

AIDŌ (*Ngai t'ong*) 愛同. Ch. spécialiste du Vinaya; trav. ca. 705 - 707. [Sk. IV] — 1424.

AJIKUTA (*A ti k'iu to*) 阿地瞿多. Td. *MUGOKUKŌ* (*Wou ki kao*) 無極高; sk. Atikūṭa, Atigupta (?). Orig. Inde centr.; arr. à *Tch'ang ngan* 651, y trav. 653 - 654. [B.; Mn.; T. 901, 785a–b] — *901*.

安樂寺 (pr. Aichi), trav. 1328. [H.] — 2294.

AN GEN (*Ngan Hiuan*) 安玄. Orig. parthe. Arr. à *Lo yang* à la fin du règne de *Han Ling ti* 漢靈帝 (168 - 189). Trav. 181. [B.; Mn.] — *322, 1508*.

AN HŌKIN (*Ngan Fa k'in*) 安法欽. Orig.

図 1

具体的には図 1 のような形で二段組になっているが，各項目には，各人名の日本語読み，中国語読み，漢字，さらに必要に応じてその異名や原語による表記，そして生まれや活動地域・時期，事績等々といった内容とその典拠となる書誌情報が記され，最後にその人物が関わった経論の大正蔵番号のリスト（立体は著・編，斜体は訳・改定等）が列挙されている．

これが今回のデジタル化の主たる対象である．さらに，文献 [1] には大正蔵の経論目録，解説中に現れる年代・地名の一覧等が付されており，これも必要に応じて適宜デジタル化される．

3. デジタル化の目的と特徴

さて，本デジタル化プロジェクトの目的は，大まかに言えば以下の四点にまとめられる．

1. 既存の目録情報を正確にデジタル化し，利用者の便に供すること
2. 既存の関連データベースとの連携を実現すること
3. 既存の目録情報の修正をオンライン上で実現すること
4. 研究者からの新規情報をオンライン上で共有するシステムを提供すること

まず 1 に関しては目録のデジタル化を行う以上当然のことであるので，特に言及する必要はないであろう．2 に関しても同様であるが，関連データベースとして例えば「大正新脩大蔵經テキストデータベース」(SAT)[3] などがある．ここでは既に仏教学分野において展開されている SAT や INBUDS のコラボレーションシステムにおける知見（[7]）等を踏まえていくことのみを指摘しておきたい．

残りの 3，4（特に後者）は，今回のデジタル化にとって，ある意味最も重要な要素であるといえる．『法寶義林』利用者の中には，各自の研究によって得られた知見を自ら付け加え，いわば独自の『法寶義林』を作り上げていくケースも多々見受けられる[4]．こうした情報の共有の場を作り上げることもまた，本プロジェクトに求められているところである．

---

[2] http://www.tei-c.org/

[3] http://21dzk.l.u-tokyo.ac.jp/SAT/

[4] もちろんこうした事態は『法寶義林』に限ったことではない．

このような目的を達成することは，一方で既存の書物の正確なデジタル化の基底路線として，他方では既存のデータに加えて新たな情報を蓄積していくという試みとして，という両側面での意義を果たしうるものであろうと考えている．したがって，ひとつのテストケースとして，データに対する意味づけ（タグづけ）を利用者側から確認できるようにしておくこともまた重要になると思われる．

次にデータの実際の処理に関する特徴として，先にも触れた以下の二点を挙げることができる．

1. TEI（Text Encoding Initiative）のガイドラインに沿ったマークアップ
2. XML DB の利用

次にこの二つの点について取り上げよう．

4. TEI（Text Encoding Initiative）のガイドラインに沿ったマークアップ

本プロジェクトにおけるマークアップの実際の手順とその問題等の詳細に関しては既に文献 [6] において論じたので，詳細は譲ることとして，ここでは TEI のガイドラインを利用したマークアップの簡単な紹介とその利用の利点について確認しておきたい．

まず TEI とは，テキストのデジタル化に際してスタンダードを標榜するコンソーシアムであり，主として人文学・社会学・言語学分野における利用を目的とした XML のタグセットとともにマークアップのガイドラインを公開している団体である．1990 年代より欧米を中心として広まり始めたそのガイドラインも改訂が重ねられ，現在は P5 がリリースされるに至っており，内容的に見ても，多岐に渡って非常に充実したものとなっている．そのため，複数の用途に応じたモジュールという形で公開されているが，本プロジェクトで扱うデータは人物のリストであるので，`namesdates` モジュールを主として利用している．

4.1. マークアップの例 —— 既存目録

次に，ここで本プロジェクトにおける実際のマークアップ内容は以下のようなものである．

```
<listPerson>
  <person xml:id="NOM001">
    <persName n="1">
      <persName xml:lang="ja-Latn">AIDŌ</persName>
      <persName xml:lang="zh-Latn">Ai tong</persName>
      <persName xml:lang="zh">愛同</persName>
    </persName>
    <nationality key="CN">Ch.</nationality>
    <occupation>spécialiste du Vinaya</occupation>
    <event type="trav" from="0705" to="0707">
      <desc>trav. ca. 705-707</desc>
    </event>
    <witDetail target="#NOM001" wit="#Sk">Sk. IV</witDetail>
    <ref type="wr" target="#T1424">1424</ref>
  </person>
  ...
</listPerson>
```

これは最初の項目である AIDŌ（図 1 参照）のマークアップである．全体としては人名リスト `listPerson` の中に各人物 `person` 要素が列挙される形となっている．

さらに，例えば典拠に関する情報（上記 `witDetail` で参照される `@wit` に対応）に関しては以下のようなリストが別途用意される．

```
<listWit>
  ...
  <witness xml:id="Sk">
    <bibl>
      <title>宋高僧伝</title>
      ...
    </bibl>
  </witness>
  ...
</listWit>
```

地名等に関しても同様のリストが用意される．

4.2. TEI ガイドライン適用の利点

単純な目録のデジタル化という点ではいわゆるリレーショナルデータベース（RDB）を利用すればシンプルであり，かつ十分であると思われる．にもかかわらず，本プロジェクトで TEI ガイドラインに従った XML 形式のマークアップを進めている最大の理由は[5]，第 3 節で述べた目的の第三・第四があるからである．つまり，既存の資料に対して様々な視点からの修正や追記等の情報を柔軟に受け入れられる余地を残しておくことが必要である．さらに前述したように，単純にデータを蓄積するだけではなく，その蓄積のあり方も含めて，利用者側から見えるようにしておくという意味でも，XML 形式を採用する利点があろう．

また副次的な効果として，現在インド学仏教学の分野では，その他の人文情報学に関与している諸分野と同様，TEI を利用したテキストのデジタル化プロジェクト等が計画・実行されており，そうしたプロジェクトとの相互連携という面も期待される．

5. XML DB の利用

以上のように本プロジェクトでは TEI を採用することとしたが，その TEI（XML）利用によって得られる利点を最大限活用するために，XML Database を利用することとした．

文献 [6] でも述べたように，今回のデジタル化の対象である目録が，現存の TEI スキーマでは十分に表現しきれない構造をもっている部分がある．本プロジェクトでは，既に述べたように研究者からの情報を募り，それを共有することを計画しているが，それは通常のコンテンツ，つまり純粋なインド学仏教学研究に基づく内容のみにとどまらず，既存の TEI マークアップの適用の是非に関するものをも含むことを想定している．したがって，マークアップそのものも流動的になりえるため，あらかじめそこに一定の枠を定めておく必要のある RDB より，ネイティヴに XML を扱い柔軟性に富む Native XML Database の利用が有効であろうと考えられる．

本プロジェクトでは数ある XML Database の中から，eXist-db Open Souce Native XML Database[6] を利用することとした．eXist は Wolfgang Meier 氏により 2000 年に創始され，現行バージョンは 1.4，XQuery 1.0, XPath 2.0, XSLT 2.0 等に対応し，XQuery 標準にも高い水準で準拠している．またオープンソースである点，Java ベースでありプラットフォームに依存しない点，様々なインターフェースに対応している点，ドキュメント類の充実，活発なコミュニティの存在等も採用の理由となった．

6. 情報の共有化に向けて

情報の共有化を考える際，先にも少し触れたように，データに対する意味づけ（マークアップ）を利用者側から確認できるようにしておくこともまた重要であると思われる．既に存在するデータがどのような形で意味づけられているのか，ということは，それに対する情報の付加と大いに関係があろう．例えば比較的単純な方法としては図 2 のように色分け等を施したツリー表示が考えられよう．ただし，一般的な利用者が必ずしも TEI あるいは XML といったものに通じている可能性は低いであろうし，そのまま XML のツリー構造のまま理解させなければならないといえるほどの必然性がある

```
⊟<person xml:id="NOM001">
  ⊟<persName n="1">
      <persName xml:lang="ja-Latn">AIDŌ</persName>
      <persName xml:lang="zh-Latn">Ai tong</persName>
      <persName xml:lang="zh">愛同</persName>
  </persName>
  <nationality key="CN">Ch.</nationality>
  <occupation>spécialiste du Vinaya</occupation>
  ⊟<event type="trav" from="0705" to="0707">
      <desc>trav. ca. 705-707</desc>
  </event>
  <witDetail target="#NOM001" wit="#Sk">Sk. IV</witDetail>
  <ref type="wr" target="#T1424">1424</ref>
</person>
```

図 2

[5] 他の理由としては，版面情報（行数・位置関係・誤植情報等）の保持も必要に応じて残しておくことを可能にすることなどが挙げられる．

[6] http://exist.sourceforge.net/

ともいえないので，この点に関してはさらなる工夫が必要となる．さしあたって重要なことは，いかなる区分になっているのか，どのような区分（タグ付け）がありえるのか，あるいはそれでは不十分なのかということを判断させる材料を提示することであると考えている．

では実際に修正・追記する場合，どのような形でマークアップされるのかという点を見ていこう．

オンライン上での一般利用者からの情報の付加に際して問題となるのは，大きく言えば，(1) いかなる形で情報を保持しておくか，という点と，(2) いかにその情報の信頼性を確保するか，という点とを挙げることができよう．

6.1. 情報の保持

この点に関して言えば，当面 TEI のガイドラインにしたがっておけば十分である．例えば，

DOMMUSEN (Tan wu [var. mo] chan) 曇無 [var. 謨] 讖 [var. 懺], …

という例を見てみよう．この例では "DOMMUSEN" という読み方とは別に "DOMMUSHIN" という読み方もあるが，目録には収録されていない．これを #FOO という @xml:id を有する研究者が情報提供したとすると，それを反映したマークアップは例えば以下のようにすればよいであろう．

```
<listPerson>
  <person xml:id="NOM155">
    <persName n="1">
      <persName xml:lang="ja-Latn">
        <choice>
          <seg>DOMMUSEN</seg>
          <seg xml:id="FOO-DOMMUSHIN">DOMMUSHIN</seg>
        </choice>
      </persName>
      ...
    </persName>
  ...
</listPerson>
<respons target ="#FOO-DOMMUSHIN" locus="value" resp="#FOO"/>
```

TEI において <respons> 要素は @target に指定される要素（ここでは <seg> 要素）の内容 @locus（ここでは value つまり "DOUMMUSHIN" という文字列）に対して @resp（ここでは "#FOO"，別途リストアップされる）が責任をもつということを表示している．つまり，誤植の場合，他の可能性を挙げる場合，等いくつかのパターンに対応できるインターフェースを用意し，また必要に応じて補充していくことによって TEI 形式に沿った一定の形式でのデータの蓄積は可能であることになる．原則的には，その度ごとに出力面での変更を加えなければいけないという必要性が生じることもない．

6.2. 信頼性の確保

データベースの運用が始まり，情報提供に関わる研究者が不特定になるほど，情報の信頼性の問題は大きくなってしまうのは自明であろう．

本プロジェクトにおいては当面，限られた研究者を選定し，その中で情報提供・共有作業を試行していくことによって，情報の信頼性を含めた様々な問題点を洗い出し，それに対する解決の方策を模索していく予定であるが，現時点では以下の点にのみ触れておくことにする．

付加される情報は大きく二つに分けられよう．一つは確たる情報（発表済み論文などを典拠として提示しうるもの）に基づく情報，他方は個々の研究者が個別に確認した情報である．前者については基本的に問題となることはないであろうし，それだけでもかなりの情報量が期待される．一方，後者に関しては，情報の信頼性の問題はもちろん，情報提供者側の視点では研究成果の帰属の問題等も生じえようが，これは個々の研究者の態度に依存するところが大きいかと思われる．いずれの場合も先に示したように，情報提供者を明確に示しておくことが解決への一助となると考えられる．

なお，実際のマークアップという点では両者とも <respons> 要素を使用することになると思われるが，比較的確度の高い前者とそうではない後者とはやはり区別しておくべきである．この点に関しては上記に加えて

<certainty> 要素を用いることで判別可能である．

前節の例を後者の例だとすると，例えば，

```
<certainty target="#FOO-DOMMUSHIN" locus="value" degree="0.5"/>
```

と表すことができる．ここで @degree は確度を表しており，情報提供者自身による申告を許すことによって暫定的なメモともなりうるし，一般利用者からの評価として利用することも可能であると思われる．また <certainty> 要素は値として <desc> 要素（description）をとることもできるため，そこで ISBN，DOI 等による情報の具体的な典拠等の情報を保持することもできる．

7. 結び

以上，『法寶義林』別冊に収められた大正新修大蔵経著者・訳者目録のデジタル化の概要と，研究者による情報の修正あるいは追加による共有のあり方，およびそれに伴って想定される問題点について論じた．今後は小規模な実際の運用を通して実用上の問題点を確認するとともに，関連プロジェクト等の成果も踏まえて，より精度の高いデータベースの構築を目指していきたいと考えている．

参考文献

[1] Demiéville, P., Durt, H. and Seidel, A.(eds.): *Fascicule annexe du Hōbōgirin : Répertoire du Canon bouddhique sino-japonais. édition de Taishō (Taishō Shinshū Daizōkyō)*, 2ème édition révisé et augmentée, Paris/Tokyo (1978).

[2] H・デュルト: フランス圏ヨーロッパの仏教学と『法宝義林』(仏教術語解説辞典)，日本学報， Vol. 2, pp. 97–101 (1983).

[3] Durt, H.: Pour Commémorer le Soixantième Anniversaire du Hobogirin, 日仏文化，Vol. 49, pp. 1–4 (1987).

[4] 櫻部　建: (書評) HÔBÔGIRIN (Cinquième Fascicule), Paris & Tokyo, 1979，仏教学セミナー， Vol. 32, pp. 81–83 (1980).

[5] Burnard, L. and Bauman, S.(eds.): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, 1.6.0, last updated on February 12th 2010, TEI Consortium.
http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf (July 3rd 2010).

[6] 松田訓典，彌永信美，永崎研宣，下田正弘: フランス語仏教辞典『法寶義林』目録のデジタル化とその課題 —— TEI ガイドラインの適用を通して ——，情報処理学会研究報告，2010-CH-87，pp. 69–78 (2010).

[7] 永崎研宣，鈴木隆泰，下田正弘: 大正新脩大蔵經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，情報処理学会研究報告，2006-CH-70，pp. 33–40 (2006).