

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——

松田 訓典^{†1} 彌永 信美^{†2}
永崎 研宣^{†3} 下田 正弘^{†4}

現在、『法寶義林』(*Hōbōgirin, Dictionnaire 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ている。本報告では、そのデジタル化の概要、デジタル化に際しての問題点、および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。

Digitization of French Dictionary of Buddhism *Hōbōgirin* using TEI Guideline

MATSUDA Kuninori,^{†1} IYANAGA Nobumi,^{†2}
NAGASAKI Kiyonori^{†3} and SHIMODA Masahiro^{†4}

We are working on the digitizing project of *Hōbōgirin, Dictionnaire encyclopédique du bouddhisme d'après les sources chinoises et japonaises*, following the standard mark-up guideline of TEI. In this report, we would like to state the outline of the project, its problems, and the approach to resolve them.

^{†1} 東京大学東洋文化研究所

Institute for Advanced Study on Asia, University of Tokyo

^{†2} フランス極東学院東京支部

École française d'Extrême-Orient, Centre de Tokyo

^{†3} 人文情報学研究所

International Institute for Digital Humanities

^{†4} 東京大学人文社会系研究科

1. はじめに

『法寶義林』(*Hōbōgirin, Dictionnaire encyclopédique du bouddhisme d'après les sources chinoises et japonaises*)とは、その副題が示す通り、仏教に関する一種の百科事典であり、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る。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が刊行されており、その間二冊の別冊も発刊されている。一つは大正新修大藏經 (以下、大正藏) 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』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

ことをその目的としており^{*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、日本においても広く認められているところである⁴⁾。

その中でも、先述の別冊に収録されている、大正藏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、大藏經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、TEI (Text Encoding Initiative)^{*2} の標準的なマークアップ方法にしたがって進めている。本報告では、そのデジタル化の概要、デジタル化に際して見出された問題点、および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じたい。

Graduate School of Humanities and Sociology, University of Tokyo

^{*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。

^{*2} <http://www.tei-c.org/>

2. 目録の構成とそのデジタル化

今回デジタル化の対象となっている経論の著者・訳者の目録 (Table des noms d'auteurs et de traducteurs rangés par ordre des transcriptions sino-japonaises avec notices biographiques) には 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ている。本節では, その中から冒頭の二人を例として, その内容, デジタル化の詳細を紹介しよう。

2.1 目録の内容

基本的な項目としては,

- [1] AIDŌ (*Ai tong*^{*1}) 愛同. Ch. spécialiste du Vinaya; trav. ca. 705 - 707. [Sk. IV] — 1424.
- [2] AJIKUTA (*A di qu duo*) 阿地瞿多. Td. MUGOKUKŌ (*Wu ji gao*) 無極高; sk. Atikūṭa, Atigupta (?). Orig. Inde centr.; arr. à *Chang an* 651, y trav. 653 - 654. [B.; Mn.; T. 901, 785a-b] — 901.

というようになっており, 各人物の人名の日本語読み, 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名前, それに続いて, 生まれや活動場所・時期, 事績, これらの情報の典拠^{*2} が掲げられ, そして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・改訂等) が列挙されている。なお, 先頭の [] 内の数字は今回のデジタル化に当たって付加した番号であり, 目録自体に付されているものではない。

2.2 マークアップ

2.2.1 基本方針

実際のマークアップ方法に移る前に, まず, 本計画にあたっての非常に大まかな方針として, 下記の三点を紹介しておこう。

- もとのテキストの情報 (誤植等を含む) を可能な限り保存しておくこと^{*3}
- 修正・追加情報の記述にも注意を払うこと
- TEI によるガイドライン⁵⁾ に可能な限りしたがうこと

*1 ただし, 文献 1) における中国語アルファベット表記は基本的に当時の EFEO 方式が用いられている。今回は方針としてすべて pinyin 表記に統一している。

*2 上記の例における典拠は以下のようになっている。

[1] Sk.: 宋高僧伝; [2] B.: 佛教大辞彙; Mn.: 佛教大年表; T.: 大正蔵。

*3 ただし, 今回は中国語 pinyin 表記のみを例外として扱っている。脚注 *1 参照。

こうした方針は, 一方で既存の書物のデジタル化の基底路線として, 他方では既存のデータに新たな情報を蓄積していくという試みとして, 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。これらの方針に沿いつつ, 現在, 以下に述べるように二段階でのマークアップを行っている。

2.2.2 マークアップ (A)

まず, TEI に準拠したマークアップを行うにあたり, 原テキストの‘版面’を尊重したものが以下である。

```
<p xml:id="NOM001"><persName n="1" xml:lang="ja-Latn">AIDŌ
</persName> (<persName n="1" xml:lang="zh-Latn">Ai tong
</persName>) <persName n="1" xml:lang="zh">愛同</persName>.
<country key="CN">Ch.</country> <note type="occupation">
spécialiste du Vinaya</note>; <date type="trav">trav. ca. 705-707
</date>. <note type="witDetail">[Sk. IV]</note> - <ref type="wr">
1424</ref>.</p>
<p xml:id="NOM002"><persName n="1" xml:lang="ja-Latn">AJIKUTA
</persName> (<persName n="1" xml:lang="zh-Latn">A di qu duo
</persName>) <persName n="1" xml:lang="zh">阿地瞿多</persName>.
Td. <persName n="2" xml:lang="ja-Latn" type="tr">MUGOKUKŌ
</persName> (<persName n="2" xml:lang="zh-Latn" type="tr">Wu ji
gao</persName>) <persName n="2" xml:lang="zh" type="tr">無極高
</persName>; sk. <persName n="3" xml:lang="sa" type="supposed">
Atikūṭa</persName>, <persName n="3" xml:lang="sa" type="
supposed">Atigupta</persName> (?). <note type="birth">Orig.
<placeName>Inde centr.</placeName></note>; <note type="event">
arr. à <placeName>Chang an</placeName> <date>651</date>, y trav.
<date>653-654</date></note>. <note type="witDetail">[B.; Mn.; T.
901, 785a-b]</note> - <ref type="tr">901</ref>.</p>
```

このマークアップ方式は既存の文章構造をそのまま活かし, その間に適宜タグを挿入したものであるため, データとしては必ずしも見やすいものとはなっていないが, 大まかな内容としては一目瞭然であろう。

以下, ここで用いているタグについて若干補足しておこう。

段落 各項目は p つまり paragraph として扱っている。これは、後述の (B) のように ‘列挙されたリストの一項目’ としてマークアップすることももちろん可能であるが、ここでは ‘版面’ としての役割を重視し、また ‘版面’ としての情報を今後付加する可能性も考慮に入れて ‘段落’ としている。

人名および言語コード TEI のガイドラインでは人名を表記する際、用途に応じていくつかのマークアップ方法を紹介しているが、中でもここでは比較的汎用的な persName を用いてマークアップしている。また本目録では、前述のように、一人の人物に対して、日本語読み、中国語読み、漢字表記のセットがあり、さらに別名に対しても同じセットが可能であり、サンسكريットなどの原語表記もありうるため、これらのセットを分類するために @n (= number) によって識別している。^{*1}

さらに、表記言語および表記の仕方を @xml:lang によって区別している。この点については W3C のガイドライン^{*2} にしたがって、primary language subtag (言語) と必要に応じて script subtag (表記) を組み合わせて用いている。

大正蔵番号の参照 ここでは ref を用い、@type により原テキストの表記に基づいた ‘著・編’ ‘訳・改訂等’ の二分類を行っている。

その他 記述内容からみてもふさわしく、マークアップ (B) でも用いている occupation (職掌) や event (誕生、移動など) といった要素は、既存の TEI スキーマでは p 中におくことが認められていない。そのため、本マークアップにおいては note を暫定的な汎用タグとして用い、属性 @type でもって細分化することによって、以下の (B) への足がかりとしている。

2.2.3 マークアップ (B)

そこで第一段階としての (A) を TEI スキーマの特性をより効果的に生かせる形に変換・処理したものが以下である。

```
<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>
<person xml:id="NOM002">
  <persName n="1">
    <persName xml:lang="ja-Latn">AJIKUTA</persName>
    <persName xml:lang="zh-Latn">A di qu duo</persName>
    <persName xml:lang="zh">阿地瞿多</persName>
  </persName>
  <persName n="2">
    <persName xml:lang="ja-Latn" type="tr">MUGOKUKŪ</persName>
    <persName xml:lang="zh-Latn" type="tr">Wu ji gao</persName>
    <persName xml:lang="zh" type="tr">無極高</persName>
  </persName>
  <persName n="3">
    <persName xml:lang="sa" type="supposed">Atikūṭa</persName>
    <persName xml:lang="sa" type="supposed">Atigupta</persName>
  </persName>
  <event type="birth">
    <desc>Orig. <placeName>Inde centr.</placeName></desc>
  </event>
  <event type="arr" when="0651">
    <label>arr</label>
    <desc>arr. à <placeName key="Chang-an">Chang an</placeName>
```

*1 これに関しては後述の (B) のように階層化した方がすっきりするかもしれない。

*2 <http://www.w3.org/International/articles/language-tags/>

```

    <date>651</date></desc>
  </event>
  <event type="trav" from="0653" to="0654">
    <label>trav</label>
    <desc><placeName key="Chang-an">y</placeName>
      trav. <date>653-654</date></desc>
  </event>
  <witDetail target="#NOM002" wit="#witB">B.</witDetail>
  <witDetail target="#NOM002" wit="#witMn">Mn.</witDetail>
  <witDetail target="#NOM002" wit="#witT">T. 901,
    785a-b</witDetail>
  <ref type="tr" target="#T0901">901</ref>
</person>
...
</listPerson>

```

これは一見してわかるように、先の (A) に比べてよりシンプルなデータ構造となっており、機械処理による様々な処理を効率的に行えるようになってきている。先と同様に、ここで新たに使用しているタグについて補足しておこう。

リスト構造 ここでは、‘段落’として扱った (A) とは異なり、目録全体を人名リスト `listPerson` として、各項目を `person` として扱っている。

事績 今回の試作データでは、(A) で汎用タグによって大まかに区分されていたものに一定のデータ処理を加えている。たとえば時期を `@when` あるいは `@from`, `@to` といった属性を付して明示したり、‘y’ という指示代名詞にそれが Chang an (長安) を指すことを `placeName/@key` により明示している、というような点である。

典拠 これについても (A) で汎用タグによって大まかに区分されていたものに一定のデータ処理を加えている。`witDetail/@wit` は別に用意される典拠となる書誌データリスト^{*1}

*1 具体的には以下のような形で用意される予定である。

```

<witList>
  <witness xml:id="witB">(B. の書誌情報) </witness>
  <witness xml:id="witMn">(Mn. の書誌情報) </witness>
  ...
</witList>

```

に対する参照としている。

大正蔵番号の参照 ここでは (A) の記述に加えて、`@target` 属性を付加している。この参照先は、具体的には本目録の前に収録されている大正蔵経論目録のデジタル化を待つことになる。

これらのデータに加えて、文献 1) には、年代や地名の一覧 (読み情報と漢字の対応表を含む) も付されており、これらを総合した形での検索システムを構築中である。その結果として、たとえば「唐の時代に長安で活躍した人」等々といった詳細な検索を可能とする一方、SAT との連携等⁶⁾ により当該人物が関わった経論を直接参照することも可能とする予定である。

3. デジタル化にあたっての問題点とその解決のために

既にこれまでに他の様々な人文系デジタル化プロジェクトが実施されてきたが、それらが直面してきた多くの問題がある。本プロジェクトにおいてもそうした問題を避けることは不可能である。

3.1 原テキストの維持

その中でも、ここでひとつの問題となるのは、どの程度、原テキストを維持しておくか、という点である。

一般的に言って、写本等をデジタル化する場合には、そこでの表記やレイアウトは極めて重要な問題を構成する場合があります。また、活版印刷された書物をデジタル化する場合でも、それがどのページのどの行にあるかということがデータとして共有できなければ、利便性を大きく損ってしまうことがある。しかしながら、今回のデジタル化の対象は一種の辞典であり、元来、他の情報を参照するためのツールという性格が非常に強いものである。このような場合には、原テキストの維持の重要性は、たとえば、大正新脩大蔵経のデジタル化において要求されるものほど高いものとは言えず、むしろ、他の情報を参照するツールとしての性格をより高い次元で実現できるようなアーキテクチャを目指すことが重要となる。

それにもかかわらず、原テキストの維持にある程度の注意を払うことはなお有用であろうと思われる。この点に関しては、前節で既に実例が挙がっていたように、^{*2} 二段階のマークアップを行うこと、特に (A) の方式を介在させることによって、データ維持の問題の解消

*2 たとえば異読情報を表示する際、原テキストでは“安慧, var. 惠” “菩提偈 [var. 仙] 那” など、書式の統一が見られるが、これも同種の問題といえよう。

に努めている。

3.2 言外にある情報

一方で、たとえば先に例に挙げた一人目の‘愛同’は‘Vinaya (律) の専門家’と記述されているのに対し、二人目の‘阿地瞿多’には専門に関する記述は存在しない。逆に、『法寶義林』の広範な地域・時代・分野にわたる研究史を踏まえて作成された記述を詳細に検討すれば、単に字面に現れている情報をはるかに超えた情報を含みうる場合も少なくない。しかし、それはある意味で利用者自身の見識が問われる問題ともいえ、そこまで踏み込んで正確に記述することは現時点では非常に困難である。

今回のデジタル化にあたっては、ある種暫定的な扱いとならざるを得ない部分を残しているが、この種の問題は、各分野の専門家の意見を募って、よりよいあり方を模索していくよりほかに解決の道はないと考えられる。

したがって本計画では、一通りの作業を終えた段階で、試作システムの公開と専門家による情報交換の場をオンライン上に設け、そこで得られたさまざまな意見をデータおよびシステムに反映させる予定である。これにあたっては、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⁾が有益となると思われる。

しかしながら、本計画において特に重要となるのは、各データに対してどのように‘意味づけ’しているか、ということを示明することであり、その‘意味づけ’自体の是非も含めて、よりふさわしいあり方を追求する作業そのものを実現し得るアーキテクチャが必要である。

また『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が、こうした情報の共有もまた、求められているところである。この点についても、情報交換の場を中心として実現に向けて取り組みたいと考えている。^{*1}

4. 結 び

以上、『法寶義林』に含まれる著者・訳者目録のデジタル化の計画と、TEI を利用したマークアップの概要、そしてデジタル化に際してのいくつかの問題点とその解決への基本方針について報告した。

研究者の参加を促すことによりよりよい研究環境が構築されることを目指しつつ、実際の

運用とそれに関する問題点に関しては稿を改めて報告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)*, 2^{ème} 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『法寶義林』（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) Nagasaki, K., Muller, C.A. and Shimoda, M.: Aspects of the Interoperability in the Digital Humanities, *Digital Humanities 2009*, pp.375-377 (2009).
- 7) 永崎研宣，鈴木隆泰，下田正弘：大正新脩大藏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).

*1 おそらく純粋にマークアップの方式という面での問題というよりは、むしろ訂正情報・追加情報に対する妥当性の確保、検索との兼ね合いといった点に検討の余地があると思われる。